Jordan Journal of Mathematics and Statistics (JJMS) 12(3), 2019, pp 329 - 350

ANALYSIS OF BIVARIATE SURVIVAL DATA USING SHARED ADDITIVE HAZARD GAMMA FRAILTY MODELS

ARVIND PANDEY [(1)], LALPAWIMAWHA [(2)], PRAVEEN KUMAR MISRA [(3)] AND R. LALAWMPUII [(4)]

Abstract. In this article, we propose additive hazard shared gamma frailty model with generalized Pareto, generalized Rayleigh and xgamma distributions as baseline distribution to analyze the bivariate survival data set of McGilchrist and Aisbett [16]. Assumption of the model is that frailty acts additively to hazard rate. The Bayesian approach of Markov Chain Monte Carlo technique was employed to estimate the parameters involved in the models. We present a simulation study to compare the true values of the parameters with the estimated values. Additive hazard shared gamma frailty model with generalized Pareto baseline distribution fits better than other propose models for kidney infection data.

1. Introduction

The bivariate survival data are said to be related if the person encounters two events or repeated events. This relationship may be due to a few other unnoticed covariates, which surreptitiously plays critical parts in investigation of survival data, which is shared by an individuals in a cluster of groups. A few illustration of bivariate survival data are - the survival times of pair of testis in the study of testicular cancer, which may be due to undescended testis, family history of the illness or past history of

2000 *Mathematics Subject Classification.* 62F15,62N01,62P10.

Key words and phrases. Additive hazard, Bayesian method, gamma frailty, generalized Pareto distribution, generalized Rayleigh distribution, xgamma distribution.

Received: April 13, 2018 Accepted: Aug. 1, 2018 .

the testicular cancer, the breakdown times of two motors of aeroplane, damages times of a pair of shoe soles, repeats times of specific cancer and so on. To examine such information, it is fundamental to present other random component, which account for within-subject reliance, inconspicuous random variable is term as frailty. Clayton [3] suggested random effect model to fit such sort of issue emerge in genuine life circumstance. The term "frailty" was first given by Vaupel et al. [21] in the study of the survival analysis.

The most used frailty distribution is gamma distribution. Gamma and inverse Gaussian distribution are more attractive because of the unconditional survival and hazard function can be expressed as simple closed form. Gamma distributions have been utilized for numerous a long times to create blends in exponential and Poisson models. From a computational point of view, they fit exceptionally well to survival models, since it is simple to determine the equations for any number of events. This is due to the effortlessness of the derivatives of the Laplace transform. This is moreover the reason why this distribution has been connected in most of the applications published until presently. The shared gamma frailty models were recommended by Clayton [3] for the examination of the relationship between clustered survival times in hereditary the study of disease transmission. An advantage is that without covariates its scientific properties are helpful for estimation [18]. Be that as it may, when adjusting for environmental risk components the examination of the clustering is more troublesome [19]. The gamma model has the advantage that choice is as it were alter of the scale of the frailty distribution.

The frailty approach of modeling has gained more attention from the past few years due to the unique features of the frailty parameters [15]. Keyfitz and Littman [13] showed that neglecting individual heterogeneity results in the wrong conclusions. Generally, a multiplicative effect of frailty on the baseline hazard function is assessed in the shared frailty models [10]. But sometimes the random effect acts additively

on the baseline hazard function has more reality in model fitting. Aalen [1,2] first suggested additive hazard model by adding covariate term in the baseline hazard functions for a lifetime of an individual t and is given as

$$m(t/X) = m_o(t) + X'\beta \tag{1.1}$$

Different way of expressing additive hazard model is given by

$$m(t/X) = m_o(t) + e^{X'\beta} \tag{1.2}$$

where m_o(t) is a baseline hazard function at time $t > 0$, X is the row vector of covariates, and β is column vector of regression coefficients. Assuming that the frailties are acting additively on the baseline hazard for a given frailty variable W=w at time $t > 0$ is

$$m(t/X) = m_o(t) + e^{X'\beta + W'\beta_w} \tag{1.3}$$

which can express as

$$m(t/X) = m_o(t) + ve^{X'\beta}, v > 0, -\infty < v < \infty \tag{1.4}$$

where v = $e^{W'\beta_w}$. Then the cumulative hazard function is

$$M(t/X) = M_o(t) + vte^{X'\beta} \tag{1.5}$$

where H_0(t) is the cumulative baseline hazard function at time $t > 0$. The conditional survival function for given frailty at time $t > 0$ is

$$S(t/v) = e^{-[M_0(t)+vte^{X'\beta}]} \tag{1.6}$$

The marginal survival function is obtained by integrating out V having the probability density f(v) and is given by

$$S(t) = S_0(t)L_z[te^{X'\beta}] \tag{1.7}$$

where $L_v(.)$ is the Laplace transformation of the distribution of V and $S_0(t)$ is the baseline survival function. Once we get the survival function at time $t > 0$, of life time random variable for an individual, we can obtain probability structure and make their inferences based on it.

In this manuscript, we consider right censored data with gamma distribution as the frailty distribution and we propose different baseline distributions such as generalized Pareto distribution, generalized Rayleigh and xgamma distributions as the baseline distribution to explore the salient features of the shared gamma frailty models based on additive hazard. Here the dependence between survival times is due to gamma distribution. When frailty distribution has zero variance, it is said to have degenerate distribution and when the distribution of frailty variable is not degenerate, positive dependence occurred. The heterogeneity of the population is determined by the value of the estimated frailty distribution parameter. The three distributions are so chosen as baseline distribution to compare due to in the univariate case, the p-values of KS-test are very large to say that the data are from generalized Pareto, generalized Rayleigh and xgamma distributions, which is also applicable in the bivariate case and all have increasing hazard rate, which is common in real life distribution.

The two common methods for estimation of parameters are maximum likelihood estimation and Bayesian method of estimation. Bayesian method have advantages in computational and analytical point of view. Thus, we employed Bayesian approach of Markov Chain Monte Carlo technique to estimate the parameters involved in the models. MCMC method can derive different features of the posterior distributions by combining information obtained from prior distribution and likelihood function. Model choice criteria can also be formulated according to posterior predictive loss [6]. Further a simulation study also presented to check the performance of the models. All the estimation procedure and models are illustrated with bivariate survival data of McGilchist and Aisbett [16] related to kidney infection data. Comparison of the

proposed models was done by the use of Bayesian comparison technique such as AIC, DIC, BIC and Bayes factor. The least values of AIC, BIC, DIC indicates better the model for consider data.

The remaining sections are as follows - in section 2, the introduction of general shared frailty model was provided and in section 3, an gamma shared frailty model based on additive hazard was also discussed. In section 4, we introduce baseline distributions. Different proposed models are given in section 5. An outline of model fitting, using Bayesian approach is presented in section 6. Section 7 is devoted to simulation study and analysis of kidney infection data respectively. Finally, section 8 consists of the discussion of the results.

2. General Shared Frailty Model

In the study it is assumed that there are n individuals, let (t_{1j}, t_{2j}) be the first and second failure times for a person, X_{kj} (k = 1,2,...,a) be the found covariate for the j^{th} person. Here it is accepted that the two failure times share the same sort of covariates. Let V_j be the shared frailty for the j^{th} person, accepting that the frailties are acting additively on the baseline hazard function. The two survival times of a person are conditionally independent for given shared frailty. Under these conditions, the conditional hazard function and conditional survival function for the j^{th} person at i^{th} (i=1,2) survival times t_{ij} for given frailty gets to be

$$m(t_{ij}/v_j, X) = m_0(t_{ij}) + v_j\eta_j \tag{2.1}$$

$$S(t_{ij}/v_j, X) = e^{-[M_0(t_{ij})+v_j t_{ij}\eta_j]} \tag{2.2}$$

where $m_0(t_{ij})$ and $M_0(t_{ij})$ are respectively hazard function and cumulative hazard function at time $t_{ij} > 0$, $\eta_j = e^{X_j\beta}$ and β is a vector of order a, of regression coefficients.

Under the assumption of independence, the bivariate survival function for the given frailty $V_j = v_j$ at time $t_{1j} > 0$ and $t_{2j} > 0$ is

$$S(t_{1j}, t_{2j}/v_j, X_j) = e^{-[(M_{01}(t_{1j})+M_{02}(t_{2j}))+v_j(t_{1j}+t_{2j})\eta_j]} \tag{2.3}$$

The unconditional survival function is obtained by integrating the conditional survival function with respect to frailty variable V_j having the probability density function $f(v_j)$, for the j^{th} individual

$$\begin{aligned} S(t_{1j}, t_{2j} \mid X_j) &= \int_{V_j} e^{-[(M_{01}(t_{1j})+M_{02}(t_{2j})+v_j(t_{1j}+t_{2j})\eta_j)]} f_v(v_j)dv_j \\ &= e^{-(M_{01}(t_{1j})+M_{02}(t_{2j}))} L_{V_j}[(t_{1j}+t_{2j})\eta_j] \end{aligned} \tag{2.4}$$

where$L_{V_j}(.)$ is the Laplace transform of the frailty variable of V_j for j^{th} individual. Here onwards $S(t_{1j}, t_{2j}/X_j)$ expressed as $S(t_{1j}, t_{2j})$.

3. Shared Gamma Frailty

A continuous random variable V is said to follow gamma distribution with parameters ζ and ξ, if its probability density function is

$$f(v) = \begin{cases} \frac{\frac{1}{\xi}^{\frac{1}{\xi}}}{\Gamma\frac{1}{\xi}} v^{\frac{1}{\xi}-1} e^{\frac{-z}{\xi}} & ;\ v > 0, \xi > 0 \\ 0 & ;\ otherwise, \end{cases} \tag{3.1}$$

For the identifiability of the distribution, the expected value of the distribution is assumed to be one and having finite variance. By using Laplace transformation, the unconditional bivariate survival function for the j^{th} individual becomes

$$S\left(t_{1j}, t_{2j}\right) = e^{-(M_{01}(t_{1j})+M_{02}(t_{2j}))}[1 + \eta_j\xi(t_{1j}+t_{2j})]^{-1/\xi} \tag{3.2}$$

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

4. Baseline distributions

Here, generalized Pareto distribution is considered as the first baseline distribution; Haktanir [9] utilized Pareto distribution to analyse the yearly optimum series for the unregulated stream in Anatolia. Davison and Smith [5] mentioned that the generalized Pareto might frame the premise of a wide modeling approach to high-level exceedances. Davison [4] modeled defilement due to the long-range are transport of radionuclides. Van Monfort and Otten [20] connected the generalized Pareto distribution to show the crests over an edge stream flow and downpour sequence. Smith [23] connected it to analyse inundation frequencies and wave statures.

If a continuous random variable T follows the three-parameter generalized Pareto distribution, then the cumulative distribution function, hazard function, and cumulative hazard function are, respectively,

$$S(t) = e^{-\gamma t}\left(1+\tfrac{t}{\lambda}\right)^{-\alpha}, t>0, \lambda>0, \gamma>0, \alpha \geq -\lambda\gamma \tag{4.1}$$

$$m(t) = \frac{f(t)}{S(t)} = \gamma + \tfrac{\alpha}{t+\lambda}, t>0$$

$$M(t) = -lnS(t) = \gamma t + \alpha ln\left(1+\tfrac{t}{\lambda}\right), t>0 \tag{4.2}$$

Where λ, α and γ are the parameters of the generalized Pareto distribution. The failure rate of the generalized Pareto distribution is increasing when $\alpha > 0$, decreasing if $\alpha < 0$ and constant for $\alpha = 0$.

A generalized Rayleigh distribution is considered as the second baseline distribution. Surles and Padgett [24] presented two-parameter Burr type X distribution and called it as generalized Rayleigh distribution. It is moreover an uncommon case of the generalized Weibull distribution, initially proposed by Mudholkar and Srivastava [17]. Kundu and Raqab [14] mentioned that the probability density function of the

generalized Pareto distribution is increasing if the shape ≤ 0.5 and decreasing if the shape parameter > 0.5. The two-parameter generalized Rayleigh distribution can be utilized viably in modeling data and moreover in modeling general lifetime data.

A continuous random variable T is said to follow the generalized Rayleigh distribution if its survival function is

$$S(t) = 1 - \left(1 - e^{-(\lambda t)^2}\right)^{\alpha}; t > 0, \alpha > 0, \lambda > 0 \tag{4.3}$$

And the hazard function and cumulative hazard function are respectively

$$m(t) = \frac{2\alpha\lambda_2 t e^{-(\lambda t)^2}(1 - e^{-(\lambda t)^2})^{\alpha-1}}{1 - (1 - e^{-(\lambda)^2})^{\alpha}}; t > 0, \alpha > 0, \lambda > 0$$

$$M(t) = -lnS(t) = -log\left[1 - (1 - e^{-(\lambda t)^2})^{\alpha}\right] \tag{4.4}$$

where α and λ are the shape and scale parameters of the distribution. The hazard function is bathtub shape if the parameter $\alpha \leq 1/2$ and increasing if the parameter $\alpha > 1/2$

The third baseline distribution considered here is xgamma distribution. Xgamma distribution is determined from the blend of exponential and gamma distributions. It is moreover utilized to analyse the alleviation times to understand the need for pain-relieving treatment.

A continuous random variable T is said to follow the xgamma distribution if its survival function is

$$S(t) = \frac{\left(1 + \alpha + \alpha t + \frac{\alpha^2 t^2}{2}\right)}{1 + \alpha} e^{-\alpha t}; t > 0; \alpha > 0 \tag{4.5}$$

And the hazard and cumulative hazard functions are respectively

$$\begin{aligned} m(t) &= \frac{\alpha^2\left(1+\frac{\alpha t^2}{2}\right)}{\left(1+\alpha+\alpha t+\frac{\alpha^2 t^2}{2}\right)} \\ M(t) &= -lnS(t) = \alpha t - log\left(\frac{1+\alpha+\alpha t+\frac{\alpha^2 t^2}{2}}{1+\alpha}\right) \end{aligned} \tag{4.6}$$

The hazard function of the xgamma distribution is increasing in t and α with $\frac{\alpha^2}{1+\alpha} < m(t) < \alpha$.

5. Proposed Models

The unconditional survival function is obtained by replacing the cumulative hazard function of generalized Pareto distribution, generalized Rayleigh distribution and xgamma distribution in equation (3.2). Then,

$$\begin{aligned} S(t_{1j},t_{2j}) &= exp[-\{(\gamma_1 t_{1j}+\alpha_1 ln(1+\tfrac{t_{1j}}{\lambda_1})+\gamma_2 t_{2j}+\alpha_2 ln(1+\tfrac{t_{2j}}{\lambda_2}))\}] \\ &\quad [1+\xi\eta_j(t_{1j}+t_{1j})]^{-1/\xi} \end{aligned} \tag{5.1}$$

$$\begin{aligned} S(t_{1j},t_{2j}) &= exp[-\{-log(1-(1-e^{-(\lambda_1 t_{1j})^2})^{\alpha_1})-log(1-(1-e^{-(\lambda_2 t_{2j})^2})^{\alpha_2})\}] \\ &\quad [1+\xi\eta_j(t_{1j}+t_{1j})]^{-1/\xi} \end{aligned} \tag{5.2}$$

$$\begin{aligned} S(t_{1j},t_{2j}) &= exp[-\{\alpha_1 t_{1j}-log(\tfrac{1+\alpha_1+\alpha_1 t_{1j}+\frac{\alpha_1^2 t_{1j}^2}{2}}{1+\alpha_1})+\alpha_2 t_{2j} \\ &\quad -log(\frac{1+\alpha_2+\alpha_2 t_{2j}+\frac{\alpha_2^2 t_{2j}}{2}}{1+\alpha_2})\}][1+\xi\eta_j(t_{1j}+t_{1j})]^{-1/\xi} \end{aligned} \tag{5.3}$$

The equations (5.1), (5.2), and (5.3) are additive hazard shared gamma frailty models with generalized Pareto distribution, generalized Rayleigh distribution, and xgamma distribution as the baseline distributions and called as model-I, model-II, and model-III respectively.

6. Bayesian Approach to parameters estimation and Model Fitting

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

$$L(\underline{\Psi},\underline{\beta},\xi) = \prod_{j=1}^{n_1} f_1(t_{1j},t_{2j}) \prod_{j=1}^{n_2} f_2(t_{1j},d_{2j}) \prod_{j=1}^{n_3} f_3(d_{1j},t_{2j}) \prod_{j=1}^{n_4} f_4(d_{1j},d_{2j}) \tag{6.1}$$

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

$$L(\underline{\Psi},\underline{\beta}) = \prod_{j=1}^{n_1} f_1(t_{1j},t_{2j}) \prod_{j=1}^{n_2} f_2(t_{1j},d_{2j}) \prod_{j=1}^{n_3} f_3(d_{1j},t_{2j}) \prod_{j=1}^{n_4} f_4(d_{1j},d_{2j}) \tag{6.2}$$

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

$$\begin{aligned} f_1(t_{1j},t_{2j}) &= \frac{\partial^2 S(t_{1j},t_{2j})}{\partial t_{1j}\partial t_{2j}} \\ f_2(t_{1j},d_{2j}) &= -\frac{\partial S(t_{1j},d_{2j})}{\partial t_{1j}} \\ f_3(d_{1j},t_{2j}) &= -\frac{\partial S(d_{1j},t_{2j})}{\partial t_{2j}} \\ f_4(d_{1j},d_{2j}) &= S(d_{1j},d_{2j}) \end{aligned} \tag{6.3}$$

Substituting the unconditional survival function in equation (6.3) for different proposed baseline distributions and differentiating, we get the likelihood function given in equation (6.1). Similarly, we can obtain the likelihood function for without frailty.

The expression of the likelihood function in equation (6.1) is not easy to solve by using Newton Raphson method. MLEs fail to converge as it involves the large dimensional optimization problem. Therefore, Bayesian approach was utilized to

estimate the parameters involved in the models, which does not endure any such kind of troubles.

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

$$\pi(\alpha_1, \lambda_1, \gamma_1, \alpha_2, \lambda_2, \gamma_2, \xi, \underline{\beta}) \propto L(\alpha_1, \lambda_1, \gamma_1, \alpha_2, \lambda_2, \gamma_2, \xi, \underline{\beta})$$
$$\times g_1(\alpha_1) g_2(\lambda_1) g_3(\gamma_1) g_4(\alpha_2) g_5(\lambda_2) g_6(\gamma_2) g_7(\xi) \prod_{i=1}^{5} p_i(\underline{\beta}_i)$$

where $g_i(.)$ $(i = 1, 2, \cdots, 7)$ indicates the prior density function with known hyper parameters of corresponding arguments for baseline parameters and frailty variance; $p_i(.)$ is prior density function for regression coefficient β_i; $\underline{\beta}_i$ represents a vector of regression coefficients except β_i, $i = 1, 2, \ldots, a$ and likelihood function $L(.)$ is given by equation (6.1) or (6.2). Here it is assumed that all the parameters are independently distributed.

Prior distributions are used as follows - gamma distribution with mean one and large variance, say $\Gamma(\Psi, \Psi)$ is used as prior distribution for frailty parameter with a small choice of Ψ, where Ψ is the parameter of the gamma distribution. 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. [11] and Sahu et al. [22] and non-informative prior assumed as the baseline parameters since we do not have any information about the baseline parameters. The two non-informative prior distributions considered are $\Gamma(a_1, b_1)$ and $U(a_2, b_2)$. All the hyper-parameters $\Psi, \varphi, a_1, a_2, b_1$ and b_2 are assumed to be known. Here $\Gamma(a_1, b_1)$ represent 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 $\Psi = 0.0001, \varphi^2 = 1,000, a_1 = 1, b_1 = 0.0001, a_2 = 0$, and $b_2 = 100$. To estimate the parameters in the models fitted with the above prior density function and likelihood

equation, 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 [8] and Gelman and Rubin [7]. To check the behavior of the chain, to decide burn-in period and autocorrelation lag, we used trace plots, coupling from the past plots and sample autocorrelation plots respectively. Trace plots describe whether or not the chain is compounding well. Nevertheless, if the chain does not converge to a stationary distribution there would be a resultant longer burn-in period. Burn-in period removes the beginning portion of the Markov chain sample to minimize the effect of the initial values of the posterior illation and chain converges to distinctive stationary distribution. Running mean plots were also used to observe the convergence of the parameter values to a posterior mean of the parameters. Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC), and Deviance Information Criteria (DIC) are utilized to compare the proposed models. Bayes factor also employed for comparison of Model M_r against Model M_v. Markov Chain Monte Carlo approach is considered to compute Bayes factor as given by Kass and Raftery [12].

7. Simulation Study and Data Analysis

To evaluate the performance of the Bayesian estimation procedure we carried out a simulation study, considering it as one covariate X_1 for the simulation purpose. X_1 was assumed to take normal distribution. As the Bayesian strategies are time expending, fifty sets of lifetimes were generated utilizing inverse transform procedure. Both the chains were iterated for 100000 times. Trace plots exhibited zigzag design indicating that parameters are moving freely and fittingly. GelmanRubin scale reduction factor values are very close to one and Geweke test values are quite small and corresponding p-values are large enough to say that the chain attains stationary distribution. Further the convergence rate was not enormously diverse. There was no

impact of prior distribution on posterior summaries because estimates of parameters were about the same. For both the chains the results were to some degree comparative so the analysis was displayed as one chain with $\Gamma(a_1, b_1)$ as prior to baseline distribution for all the models. From the above conditions, we can said that the models satisfy necessary conditions to apply the propose models in the data. Table 1, 2, and 3 present the posterior summaries of generalized Pareto, generalized Rayleigh and xgamma distributions as baseline distribution. It provides estimates (posterior means), standard error and upper and lower credible limits.

Parameter (value)	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 5000; autocorrelation lag = 240							
α_1 (0.0020)	0.0020	0.0003	0.0014	0.0026	-0.0043	0.4982	1.0002
α_2 (0.0019)	0.0020	0.0005	0.0010	0.0029	0.0173	0.5069	1.0025
λ_1 (57.120)	57.118	0.5784	56.151	58.035	0.0035	0.5014	1.0095
λ_2 (55.100)	55.084	0.5763	54.150	56.049	0.0029	0.5011	1.0004
γ_1 (0.0059)	0.0060	0.0005	0.0050	0.0069	0.0064	0.5025	1.0019
γ_2 (0.0070)	0.0070	0.0005	0.0060	0.0079	0.0147	0.5058	1.0015
ξ (2.6290)	2.6302	0.0059	2.6203	2.6394	0.0005	0.5002	2.6394
β (-0.0608)	-0.0606	0.0007	-0.0619	-0.0595	0.0159	0.5063	1.0089

Table 1: Gamma frailty with generalized Pareto distribution as baseline (Simulation for model-I)

Parameter (value)	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 7900; autocorrelation lag = 163							
α_1 (6.2997)	6.2998	0.0324	6.2348	6.3631	-0.0045	0.4981	1.0001
α_2 (1.0040)	1.0040	0.0522	0.9055	1.0922	-0.0037	0.4984	1.0017
λ_1 (0.0031)	0.0031	0.0004	0.0022	0.0039	0.0013	0.5005	1.0003
λ_2 (0.0028)	0.0028	0.0020	0.0037	0.0046	-0.0106	0.4957	1.0000

ξ (1.9080)	1.9102	0.0044	1.9014	1.9189	-0.0062	0.4975	1.0001
β (-0.0516)	-0.0516	0.0030	-0.0581	-0.0464	0.0033	0.5013	1.0014

Table 2: Gamma frailty with generalized Rayleigh distribution as baseline (Simulation for model-II)

Parameter (value)	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 8000; autocorrelation lag = 155							
α_1 (0.0102)	0.0102	0.0018	0.0066	0.0140	-0.0021	0.4991	0.9999
α_2 (0.0110)	0.0109	0.0027	0.0063	0.0165	0.0007	0.5002	1.0000
ξ (2.9920)	2.9931	0.0516	2.8987	3.0820	-0.0045	0.4981	1.0000
β (-0.0730)	-0.0735	0.0044	-0.0798	-0.0631	-0.0002	0.4998	1.0045

Table 3: Gamma frailty with xgamma distribution as baseline (Simulation for model-III)

The applicability of the models was also checked by applying them to the kidney infection data. The urinary organ infection knowledge has appeared in McGilchrist and Aisbett [16]. It is associated with return time to infection during the course of insertion of the tube for thirty-eight urinary organ patients due to mistreatment with portable dialysis instrument. For every patient, initial and second return times (in days) of infection attributable to infection from the time of insertion of the tube till it is to be removed are recorded. The tube ought to be removed for reasons apart from urinary organ infection, and this will be regarded as censoring. Therefore survival times for a patient given in the study is also first or second infection time or censoring time. The value zero is employed for censoring and one is employed for the incidence of infection. Once the incidence or censoring of the primary infection occurred, decent time (10 weeks interval) was allowed for the infection to be cured before the tube was inserted for the second time. So, the primary and second return times will be thought of as independent except the common frailty element. The information comprises 3

risk variables - age, sex, and disease type- GN, AN, and PKD, where GN, AN, and PKD are brief forms of Glomerulo Nephritis's, Acute Nephritis's, and Polycyatic Kidney Disease. The infection times of every patient share an equivalent value of the covariates. Let T_1 and T_2 be representing first and second recurrences of infection. Five covariates age, sex, and presence or absence of disease type GN, AN and PKD are portrayed by X_1, X_2, X_3, X_4, and X_5.

	Recurrence time	
Distribution	first	second
Generalized Pareto	0.1398	0.1230
Generalized Rayleigh	0.9024	0.7912
Xgamma	0.2628	0.2221

Table 4: p-values of K-S Statistics for goodness of fit test for Kidney Infection data set

First, we check the goodness of fit for the kidney infection data by considering Kolmogorov Smirnov test. The p-values obtained for the first and second recurrences are large enough to say that there is no reason to reject the hypothesis that the first and second recurrence time to follow one of the distributions with the survival function as given in equations (4.1), (4.3) and (4.5). The corresponding p-values are given in Table 1. Trace plot (Figure 1(a)) shows zigzag design, it indicates that the parameters are more move freely. Coupling from the past plot and running mean plot also shows that the two chains are mixing well.

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6800; autocorrelation lag = 270							
α_1	0.0020	0.0003	0.0013	0.0027	0.0126	0.5050	1.0001

λ_1	57.238	0.5871	56.220	58.148	0.0070	0.5028	0.9999
γ_1	0.0060	0.0005	0.0050	0.0069	0.0010	0.5004	1.0032
α_2	0.0020	0.0005	0.0010	0.0029	-0.0043	0.4982	1.0018
λ_2	55.031	0.5832	54.075	56.000	-0.0019	0.4992	1.0045
γ_2	0.0068	0.0005	0.0060	0.0078	0.0085	0.5034	1.0000
ξ	2.9686	0.0505	2.8808	3.0588	0.0014	0.5005	1.0022
β_1	-0.1264	0.0198	-0.1673	-0.0891	0.0053	0.5021	0.9999
β_2	-4.6359	0.5679	-5.4667	-3.3278	-0.0019	0.5021	1.0072
β_3	3.3682	0.5222	2.5480	4.4146	-0.0037	0.4985	1.0021
β_4	2.6786	0.5181	1.7705	3.6208	0.0032	0.5013	1.0009
β_5	0.0021	0.0005	0.0013	0.0031	-0.0046	0.4981	1.0005

Table 5: Posterior results for the Kidney infection data set(model-I)

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6100; autocorrelation lag = 345							
α_1	6.3201	0.0348	6.2402	6.3919	-0.0020	0.4991	1.0001
λ_1	0.0032	0.0003	0.0024	0.0039	-0.0071	0.4971	1.0001
α_2	0.9998	0.0556	0.9053	1.0935	0.0006	0.5002	1.0000
λ_2	0.0029	0.0004	0.0020	0.0038	0.0009	0.5003	1.0000
ξ	1.8738	0.0517	1.7821	1.9624	0.0003	0.5001	1.0071
β_1	-0.0228	0.0051	-0.0297	-0.0104	0.0029	0.5011	1.0038
β_2	-4.1923	0.3997	-5.0179	-3.4107	-0.0005	0.5011	1.0009
β_3	-0.0485	0.0061	-0.0592	-0.0368	0.0078	0.5031	1.0000

β_4	0.8769	0.0510	0.7877	0.9674	-0.0078	0.4968	0.9999
β_5	-0.0492	0.0057	-0.0584	-0.0377	-0.0030	0.4987	1.0027

Table 6: Posterior results for the Kidney infection data set(model-II)

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 8200; autocorrelation lag = 180							
α_1	0.0117	0.0019	0.0081	0.0153	-0.0163	0.4934	1.0001
α_2	0.0127	0.0030	0.0075	0.0189	-0.0083	0.4966	1.0002
ξ	2.9690	0.0511	2.8774	3.0592	-0.0075	0.4970	1.0008
β_1	-0.0317	0.0053	-0.0394	-0.0211	0.0034	0.5013	1.0000
β_2	-5.0794	0.4604	-5.9323	-4.1939	0.0127	0.4940	1.0008
β_3	-0.1037	0.0503	-0.1932	-0.0061	0.0019	0.5007	1.0041
β_4	2.7863	0.0500	2.6961	2.8812	-0.0156	0.4937	0.4937
β_5	0.4986	0.0495	0.4135	0.5886	-0.0120	0.4951	1.0014

Table 7: Posterior results for the Kidney infection data set(model-III)

The comparison between the proposed models is done by utilizing AIC, BIC, and DIC. Though the values of model-I is less than other model-II and model-III, the distinction between AIC, BIC, and DIC values for the proposed models are exceptionally little, so AIC, BIC, and DIC values are not commendable to take a choice between the models. Presently we consider Bayes factor D_{rv} for comparing the models r and v. For two models of substantive interest, M_r and M_v, twice the log of the Bayes factor is approximately equal to the difference in their BIC approximations.

Model No.	AIC	BIC	DIC	log-likelihood
Model I	692.0086	711.6596	671.5694	-334.0043
Model II	705.3641	721.7400	689.7909	-342.6821
Model III	701.0399	714.1406	689.9846	-342.5200

Table 8: AIC, BIC and DIC values for six models

numerator model against denominator model	$B_{rv} = 2log_e(D_{rv})$	range	Evidence against model in denominator
M_I against M_{II}	16.7886	> 10	Very strong positive
M_I against M_{III}	18.8892	> 10	Very strong positive
M_{II} against M_{III}	2.1005	≥ 2 and ≤ 6	Positive

Table 9: Bayes factor values and decision for test of significance for frailty fitted to Kidney Infection Data Set

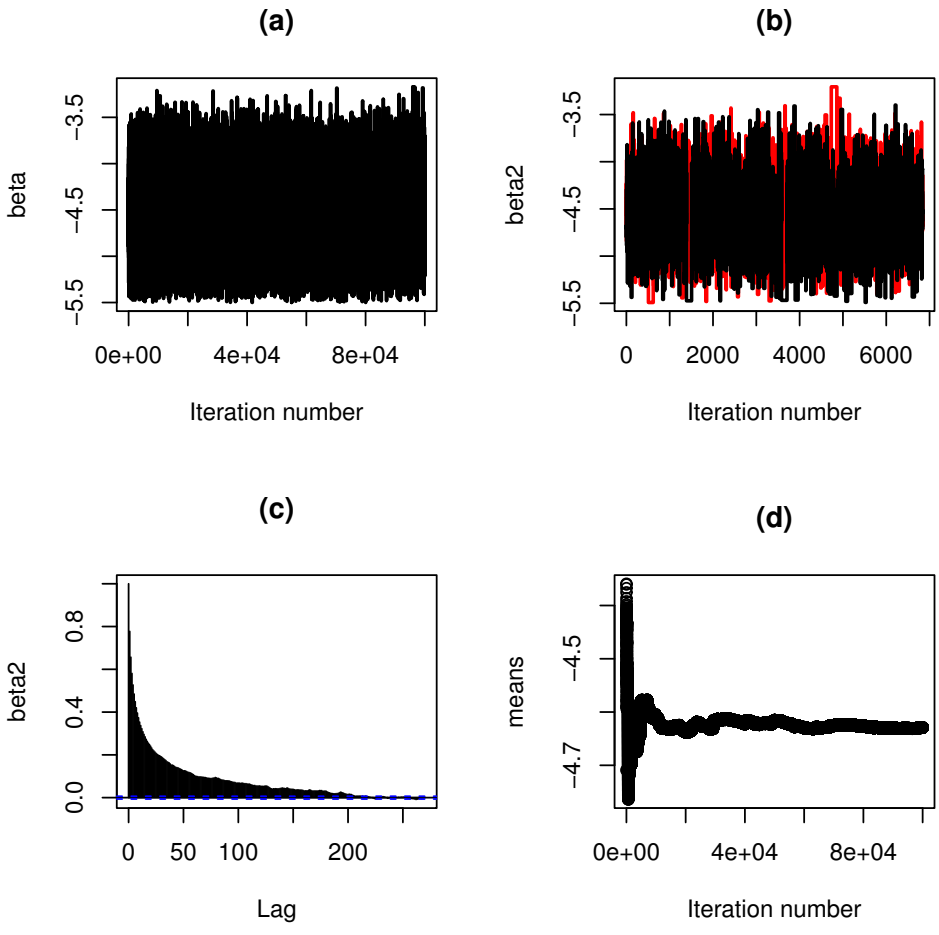

FIGURE 1. Figure-Graph of (a)Trace plot, (b) Coupling from past plot, (c) ACF plot and (d) Running mean plot for model-I

B_{rv} for the model-I against model-II is 16.7886; for model-I against model-III is 18.8892; for model-II against model-III is 2.1005. In comparison to model-I, model-II, model-III, for model-I, B_{12}, and B_{13} both are much higher than 10 which recommend that model-II and model-III are not better than model-I which affirm our earlier result given in Table 5. Hence from all the demonstrate comparison criteria we can say model-I is better than model-II, model-III for modeling kidney infection data.

8. Discussion

In this study, we examine the additive hazard shared gamma frailty model with three baseline distributions such as generalized Pareto, generalized Rayleigh and xgamma distributions and without frailty models based on the same baseline distributions.

The Metropolis-Hastings and Gibbs sampler was utilized to fit all the proposed models. Kidney infection data was analyzed using the proposed models and the finest model is suggested. We have utilized self-composed programs in R statistical software have been utilized to perform the analysis.

All the demonstrated comparison criteria exhibits that additive hazard shared gamma frailty demonstrated with generalized Pareto baseline is better for modeling of kidney infection data rather than generalized Rayleigh and xgamma baselines. The estimates of frailty parameters are high in all three models which are 2.9686, 1.8738 and 2.9690 for generalized Pareto, generalized Rayleigh and xgamma baseline models respectively. This demonstrates that there is a strong evidence of high degree of heterogeneity in the population of patients. A few patients are anticipated to be exceptionally inclined to infection compared to others with the same covariate values. We can further establish that there is a strong positive relationship between the two infection times for the same patient. Now, we are in a position to say that we have

a new additive hazard shared gamma frailty model with generalized Pareto baseline distribution for analysis of the kidney infection data.

Acknowledgement

We thank the editor and the referees for the valuable suggestions and comments which improved the earlier version of the manuscript..

References

[1] O.O. Aalen, *A model for non-parametric regression analysis of counting processes.* In Lecture notes on statistics 2: Mathematical statistics and probability theory (eds Klonecki, W., Kozek, A. & Rosinski, J.), Springer, New York, 1980.

[2] O.O. Aalen, A linear regression model for the analysis of lifetimes,*Stat. Med.* **8**(1989), 907–925.

[3] 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.* **65**(1978), 141–151.

[4] A.C. Davison, *Modelling excesses over high threshold, with an application.* In: Statistical extremes and Applications, ed.J. Tiago de Oliveira, **461–481**. Reidel, Dordrecht, The Netherlands, 1994.

[5] A.C. Davison, R.L. Smith, Models for exceedances over high thresholds,*J. Roy.Statist. Soc B.* **52(3)**(1990), 393–442.

[6] A.E. Gelfand, S.L. Ghosh, Model Choice: A Minimum Posterior Predictive Loss Approach,*Biometrika.* **85**(1998), 1–11.

[7] 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, 1992.

[8] 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,1992.

[9] T. Haktanir, Comparison of various flood frequency distributions using annual peaks data of rivers in Anatolia,*J. Hydrol.* **136**(1992), 1–31.

[10] D.D. Hanagal, A. Pandey, Inverse Gaussian shared frailty for modeling kidney infection data,*Advances in Reliability.* **1**(2014), 1–14.

[11] J.G. Ibrahim, Ming-Hui Chen and D. Sinha, *Bayesian Survival Analysis.* Springer, Verlag, 2001.

[12] K.E. Kass, A.E. Raftery, Bayes Factors,*Journal of the American Statistical Association.* 90(430)(1995), 773–95.

[13] N. Keyfitz, G. Littman, Mortality in a Heterogeneous Population, *Population Studies.* **33**(1979), 333–342.

[14] D. Kundu, M.Z. Raqab, Burr Type X distribution,*Journal of Probability and Statistical Sciences.* **4(2)**(2006), 179–193.

[15] T. Lancaster, S. Nickell, The Analysis of Re-employment Probabilities for the Unemployed (with Discussion),*Journal of the Royal Statistical Society A.* **143**(1980), 141–165.

[16] C.A. McGilchrist, C.W. Aisbett, Regression with frailty in survival analysis,*Biometrics.* **47**(1991),461–466.

[17] G.S. Mudholkar, D.K. Srivastava, Exponentiated Weibull family for analyzing bathtub failure data,*IEEE Transactions on Reliability.* **42**(1993), 299–302.

[18] D. Oakes, A model for association in bivariate survival data, *J. R. Statist. Soc. B.* **44**(1982), 414–22.

[19] E. Parner, Asymptotic theory for correlated gamma-frailty model,*Annals of Statistics.* **26**(1998), 183–214.

[20] M.A.J. van Monfort, A. Otten, The first and the second effect of the extreme value distribution, EV1, Stochastic Hydrol,*Hudraul.* **5**(1991),69–76.

[21] J.W. Vaupel, K.G. Manton and E. Stallaed, The impact of heterogeneity in individual frailty on the dynamics of mortality,*Demography.* **16**(1979), 439–454.

[22] S.K. Sahu, D.K. Dey, H. Aslanidou and D. Sinha, A Weibull regression model with gamma frailties for multivariate survival data*Life time data analysis.* **3**(1997), 123–137.

[23] R.L. Smith, *Threshold methods for sample extremes, In: Statistical Extremes and Applications* ed.J. Trago de Olivera, 621-638. Reidel, Dordrecht, The Nehterlands, 1984.

[24] J.G. Surles, W.J. Padgett, (2001). Inference for reliability and stress-strength for a scaled Burr Type X distribution , *Lifetime Data Analysis.* **7**(2001), 187–200.

(1) Department of Statistics,Central University of Rajasthan, India

E-mail address: arvindmzu@gmail.com

(2) Department of Statistics,Pachhunga University College, Mizoram - 796001

E-mail address: raltelalpawimawha08@gmail.com

(3) Department of Mathematics and Statistics,Dr Shakuntala Misra National Rehabilitation University, India

E-mail address: PraveenKumarmisra@gmail.com

(4) Department of Mathematics,Govt Zirtiri Residential Science College, India

E-mail address: rlalawmpuii@gmail.com