Pak. J. Statist.
2018 Vol. 34(4), 311-330

ADDITIVE SHARED INVERSE GAUSSIAN FRAILTY MODEL

Arvind Pandey[1], Lalpawimawha[2§] and **Shashi Bhushan[3]**
[1] Department of Statistics, Central University of Rajasthan, India
[2] Pachhunga University College, Aizawl, Mizoram-796001, India
[3] Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, Uttar Pradesh, India
[§] Corresponding author Email: raltelalpawimawha08@gmail.com

ABSTRACT

The study proposes additive hazard shared inverse Gaussian frailty model with generalized Pareto, generalized Rayleigh and xgamma distributions as baseline distribution to analyze the bivariate data set of McGilchrist and Aisbett (1991). The estimation of the parameters involved in the models was done by Bayesian approach of Markov Chain Monte Carlo technique. The true values and the estimated values of the parameters are compared by using simulation study. The proposed models are fitted to the real life data set and the best model suggested for the data.

KEYWORDS

Bayesian method, generalized Pareto distribution, generalized Rayleigh distribution, inverse Gaussian frailty, xgamma distribution.

1. INTRODUCTION

The bivariate survival data are said to be correlated if the individual experiences two events or recurrence events and so on. This correlation may be due to some other hidden covariates, which are unobserved and plays important roles in analysis of survival data. The unobservable random variable is termed as “frailty", which is shared by individuals in a group. Some examples of bivariate survival data are - the survival time of a pair of testis in the study of testicular cancer, which may be due to undescended testis or previous history of testicular cancer in medical research; the failure time of two engines of aero plane in engineering research, damage time of a pair of shoe soles, recurrences of a particular cancer and so on. To analyze such kind of data, it is necessary to introduce other random components, which accounts for within-subject dependency. Clayton (1978) first suggested random effect model for such type of problem arising in real-life situation, which was later given the term “frailty" by Vaupel et al. (1979) in the study of mortality.

The most used frailty distribution is gamma distribution due to mathematical convenience, but it has some demerits (Kheri et al., 2007). The alternative frailty distribution is inverse Gaussian distribution. The gamma and inverse Gaussian distributions have distinct properties. Hougaard (1984) mentioned that the homogeneity of the population with time is due to inverse Gaussian; inverse Gaussian distribution is also more flexible than gamma for modeling of the survival data. When there are more

failures at the beginning of life time distribution and non-monotonic failures rate is expected, the inverse Gaussian model is more appropriate for the life time model. Gamma and inverse Gaussian distribution are more attractive because the unconditional survival function and hazard function can be expressed as simple closed form.

The frailty approach of modeling has gained more attention in some recent reference due to the unique features of the frailty parameters (Lancaster and Nickell, 1980). Keyfitz and Littman (1979) 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 (Hanagal and Pandey, 2014). But sometimes the random effect acts additively on the baseline hazard function accounting for more realistic nature. Aalen (1980, 1989) first suggested additive hazard model by adding covariate term in the baseline hazard function for the lifetime of an individual t and is given as

$$r(t/X) = r_0(t) + X'\beta$$

Different way of expressing additive hazard model is given by

$$r(t/X) = r_0(t) + e^{X'\beta}$$

where $m_0(t)$ 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

$$r(t/X) = r_0(t) + e^{X'\beta + W'\beta_w}$$

which can be expressed as

$$r(t/X) = r_0(t) + ve^{X'\beta} \ , v > 0, -\infty < w < \infty$$

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

$$R(t/X) = R_0(t) + vte^{X'\beta}$$

where $R_0(t)$ is the cumulative baseline hazard function at time $t > 0$. The conditional survival function for a given frailty at time $t > 0$ is

$$S(t/v) = e^{-[R_0(t) + vte^{X'\beta}]}$$

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

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

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 inferences based on it.

Hanagal and Pandey (2016, 2017) also studied the additive frailty by using frailty distribution as gamma and inverse Gaussian distributions and exponential power distribution, generalized log-logistic distribution and generalized Weibull distribution as baseline distributions.

In this manuscript, we consider right censored data with inverse Gaussian distribution as frailty distribution and generalized Pareto, generalized Rayleigh and xgamma distributions as the baseline distribution to explore the salient features of the additive hazard shared inverse Gaussian frailty models. Here the dependence between survival times is due to inverse Gaussian distributed common frailty variable. When the frailty distribution has zero variance, it is said to have degenerate distribution and when the distribution of frailty variable is not degenerate, positive dependence occurs. The heterogeneity of the population is determined by the value of the estimated frailty parameter. The three distributions are chosen as baseline distributions for comparison since there are few differences in the property of the hazard functions for each proposed baseline distribution.

The two common approaches for estimation of parameters are maximum likelihood estimation method and Bayesian method of estimation. Bayesian method has advantages from 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 distribution by combining information obtained from prior distribution and likelihood function. Model choice criteria can also be formulated according to posterior predictive loss (Gelfand and Ghosh, 1998). Further, a simulation study is presented to check the performance of the models. All the estimation procedures and models are illustrated with bivariate survival data of Aisbett and McGilchrist (1991) related to kidney infection data. Comparison of the proposed models is done by using Bayesian comparison technique such as AIC, BIC, DIC and Bayes factor.

The remaining sections are categorized as- in section 2, the introduction of general shared frailty model is provided and in section 3, an inverse Gaussian shared frailty model based on additive hazard is discussed. Section 4 introduces the baseline distributions. Different proposed shared frailty models are given in section 5. An outline of model fitting, using Bayesian approach is presented in section 6. Sections 7 and 8 are devoted to simulation study and analysis of kidney infection data respectively. Finally, section 9 consists of the discussion of the results.

2. GENERAL SHARED FRAILTY MODEL

In the study it is assumed that there are n individuals, let $\left(t_{1q}, t_{2q}\right)$ be the first and second failure times for a person, $X_{kq}\left(k=1,2,\ldots,a\right)$ be the found covariate for the q^{th} person. Here it is assumed that the two failure times share the same sort of covariates. Let V_q be the shared frailty for the q^{th} person, assuming 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 q^{th} person at p^{th} $(p=1,2)$ survival times t_{pq} for given frailty is

$$r(t_{pq}/v_q, X) = r_0(t_{pq}) + v_q\eta_q \tag{1}$$

$$S(t_{pq}/v_q, X) = e^{-[R_0(t_{pq})+v_q t_{pq}\eta_q]} \tag{2}$$

where $r_0(t_{pq})$ and $R_0(t_{pq})$ are respectively hazard function and cumulative hazard function at time $t_{pq}>0$, $\eta_q = e^{X_q\beta}$ and β is the regression coefficient of order a.

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

$$S(t_{1q}, t_{2q}) = e^{-[(R_{01}(t_{1q})+R_{02}(t_{2q})+v_q(t_{1q}+t_{2q})\eta_q]} \tag{3}$$

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

$$\begin{aligned} S(t_{1q}, t_{2q}) &= \int_{V_q} e^{-[(R_{01}(t_{1q})+M_{02}(t_{2q})+v_q(t_{1q}+t_{2q})\eta_q]} f_v(v_q)dv_q \\ &= e^{-[(R_{01}(t_{1q})+R_{02}(t_{2q})]} L_{V_q}[(t_{1q}+t_{2q})\eta_q] \end{aligned} \tag{4}$$

where $L_{Vq}(.)$ is the Laplace transformation of the frailty variable of V_q for q^{th} individual. Here onwards $S(t_{1q},\ t_{2q}/X_q)$ is expressed as $S(t_{1q},\ t_{2q})$.

3. SHARED INVERSE GAUSSIAN FRAILTY

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

$$f(v) = \begin{cases} \left[\frac{1}{2\pi\zeta}\right]^{\frac{1}{2}} v^{\frac{-3}{2}} e^{\frac{(v-\xi)^2}{2v\zeta\xi^2}} & ;v>0,\zeta>0,\xi>0 \\ 0 & ;Otherwise \end{cases} \tag{5}$$

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

$$S(t_{1q}, t_{2q}) = e^{-(R_{01}(t_{1q})+R_{02}(t_{2q}))} \exp\left[\frac{1-(1+2\xi\eta_q(t_{1q}+t_{2q}))^{1/2}}{\xi}\right] \tag{6}$$

where $R_{01}\left(t_{1q}\right)$ and $R_{02}\left(t_{2q}\right)$ are the cumulative baseline hazard functions of the lifetime T_{1q} and T_{2q}.

The bivariate survival distribution function in the absence of frailty (without frailty) is given by

$$S(t_{1q}, t_{2q}) = e^{-(R_{01}(t_{1q})+R_{02}(t_{2q})+(t_{1q}+t_{2q})\eta_q)}. \tag{7}$$

4. BASELINE HAZARD ASSUMPTIONS

Here, generalized Pareto distribution is considered as the first baseline distribution; Haktanir (1992) utilized Pareto distribution to analyze the yearly optimum series for the unregulated stream in Anatolia. Davison and Smith (1990) mentioned that the generalized Pareto might frame the premise of a wide modeling approach to high-level exceedances. Davison (1994) modeled defilement due to the long-range are transport of radionuclides. Van Monfort and Otten (1991) connected the generalized Pareto distribution to show the crests over an edge stream flow and downpour sequence. Smith (1984) connected it to analyze inundation frequencies and wave statures. Davison and Smith (1990) displayed a comprehensive examination of the extremes of information by utilizing the generalized Pareto distribution for modeling the sizes and events of exceedances over a tall limit.

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,

$$S(t) = e^{-\gamma t}\left(1+\frac{t}{\lambda}\right)^{-\alpha} ; t>0, \lambda>0, \gamma>0, \alpha \geq -\lambda\gamma \tag{8}$$

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

$$R(t) = -\ln S(t) = \gamma t + \alpha \ln\left(1+\frac{t}{\lambda}\right), t>0 \tag{9}$$

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 (2001) presented a new 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 (1993). Kundu and Raqab (2006) mentioned that the probability density function of the

generalized Pareto distribution is increasing if the shape parameter ≤ 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 have generalized Rayleigh distribution if its survival function is

$$S(t) = 1 - \left(1 - e^{-(\lambda t)^2}\right)^{\alpha} ; t > 0, \alpha > 0, \lambda > 0 \tag{10}$$

And the hazard function and cumulative hazard function are respectively

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

$$R(t) = -\ln S(t) = -\log\left[1 - \left(1 - e^{-(\lambda t)^2}\right)^{\alpha}\right] \tag{11}$$

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 (Sen et al., 2016). It is moreover utilized to analyze the alleviation times to understand the need for pain relieving treatment.

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

$$S(t) = \frac{\left(1 + \alpha + \alpha t + \frac{\alpha^2 t^2}{2}\right)}{1 + \alpha} e^{-\alpha t}; t > 0, \alpha > 0 \tag{12}$$

And the hazard and cumulative hazard functions are respectively

$$r(t) = \frac{\alpha^2\left(1 + \frac{\alpha^2 t^2}{2}\right)}{\left(1 + \alpha + \alpha t + \frac{\alpha^2 t^2}{2}\right)}$$

$$R(t) = -\log S(t) = \alpha t - \log\left(\frac{\left(1 + \alpha + \alpha t + \frac{\alpha^2 t^2}{2}\right)}{1 + \alpha}\right) \tag{13}$$

The hazard function of the xgamma distribution is increasing in $t>0$ and α with $\frac{\alpha^2}{1+\alpha} < r(t) < \alpha$. Sen et al. (2017) also proposed the weighted xgamma distribution as a generalization of xgamma distribution, which is a useful tool for describing time-to-event data sets.

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 (6). Then,

$$S(t_{1q},t_{2q}) = \exp\left[-\left\{\left(\gamma_1 t_{1q} - \alpha_1 \ln\left(1+\frac{t_{1q}}{\lambda_1}\right)+\gamma_2 t_{2q} + \alpha_2 \ln\left(1+\frac{t_{2q}}{\lambda_2}\right)\right)\right\}\right] \exp\left[\left\{\frac{1-\left(1+2\xi\eta_q(t_{1q}+t_{2q})\right)^{1/2}}{\xi}\right\}\right] \quad (14)$$

$$S(t_{1q},t_{2q}) = \exp\left[-\left\{-\log\left(1-\left(1-e^{-\left(\lambda_1 t_{1q}\right)^2}\right)^{\alpha_1}\right)-\log\left(1-\left(1-e^{-\left(\lambda_2 t_{2q}\right)^2}\right)^{\alpha_2}\right)\right\}\right] \exp\left[\left\{\frac{1-\left(1+2\xi\eta_q(t_{1q}+t_{2q})\right)^{1/2}}{\xi}\right\}\right] \quad (15)$$

$$S(t_{1q},t_{2q}) = \exp\left[-\left\{\alpha_1 t_{1q} - \log\left(\frac{1+\alpha_1+\alpha_1 t_{1q}+\frac{{\alpha_1}^2 {t_{1q}}^2}{2}}{1+\alpha_1}\right) +\alpha_2 t_{2q} - \log\left(\frac{1+\alpha_2+\alpha_2 t_{2q}+\frac{{\alpha_2}^2 {t_{2q}}^2}{2}}{1+\alpha_2}\right)\right\}\right] \exp\left[\left\{\frac{1-\left(1+2\xi\eta_q(t_{1j}+t_{2q})\right)^{1/2}}{\xi}\right\}\right] \quad (16)$$

$$S(t_{1q},t_{2q})=\exp\left[-\left\{\left(\gamma_1 t_{1q}-\alpha_1\ln\left(1+\frac{t_{1q}}{\lambda_1}\right)+\gamma_2 t_{2q}+\alpha_2\ln\left(1+\frac{t_{2q}}{\lambda_2}\right)\right)+\eta_q\left(t_{1q}+t_{2q}\right)\right\}\right] \tag{17}$$

$$S(t_{1q},t_{2q})=\exp\left[-\left\{-\log\left(1-\left(1-e^{-\left(\lambda_1 t_{1q}\right)^2}\right)^{\alpha_1}\right)-\log\left(1-\left(1-e^{-\left(\lambda_2 t_{2q}\right)^2}\right)^{\alpha_2}\right)+\eta_q\left(t_{1q}+t_{2q}\right)\right\}\right] \tag{18}$$

$$S(t_{1q},t_{2q})=\exp\left[-\left\{\begin{array}{l}\alpha_1 t_{1q}-\log\left(\dfrac{1+\alpha_1+\alpha_1 t_{1q}+\dfrac{{\alpha_1}^2{t_{1q}}^2}{2}}{1+\alpha_1}\right)\\ +\alpha_2 t_{2q}-\log\left(\dfrac{1+\alpha_2+\alpha_2 t_{2q}+\dfrac{{\alpha_2}^2{t_{2q}}^2}{2}}{1+\alpha_2}\right)+\eta_q\left(t_{1q}+t_{2q}\right)\end{array}\right\}\right] \tag{19}$$

The equations (14), (15), and (16) are additive hazard shared inverse Gaussian frailty models with generalized Pareto distribution, generalized Rayleigh distribution, and xgamma distribution as the baseline distributions and are named as model-I, model-II, and model- III. Equations (17), (18), and (19) are without frailty models with generalized Pareto distribution, generalized Rayleigh distribution, and xgamma distribution as the baseline distributions and called as model-IV, model-V and model-VI.

6. BAYESIAN APPROACH TO PARAMETERS ESTIMATION AND LIKELIHOOD FUNCTIONS

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

$$L(\underline{\Psi},\underline{\beta},\xi)=\prod_{q=1}^{n_1} f_1(t_{1q},t_{2q})\prod_{q=1}^{n_2} f_2(t_{1q},d_{2q})\prod_{q=1}^{n_3} f_3(d_{1q},t_{2q})\prod_{q=1}^{n_4} f_4(d_{1q},d_{2q}) \tag{20}$$

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

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

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

$$f_1(t_{1q}, t_{2q}) = \frac{\partial^2 S(t_{1q}, t_{2q})}{\partial t_{1q} \partial t_{2q}}$$

$$f_2(t_{1q}, d_{2q}) = -\frac{\partial S(t_{1q}, d_{2q})}{\partial t_{1q}}$$

$$f_3(d_{1q}, t_{2q}) = -\frac{\partial S(d_{1q}, t_{2q})}{\partial t_{2q}}$$

and

$$f_4(d_{1q}, d_{2q}) = S(d_{1q}, d_{2q}) \tag{22}$$

Substituting the unconditional survival function in equation (22) for different proposed baseline distributions and differentiating, we get the likelihood function given in equation (20). Similarly, we can obtain the likelihood function for without frailty.

The expression of the likelihood function in equation (20) is not easy to solve. So, Newton Raphson method will be utilized to estimate the parameters, but 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

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

where $g_p(.)(p = 1,2,\ldots,7)$ indicates the prior density function with known hyper parameters of corresponding arguments for baseline parameters and frailty variance; $C_p(.)$ is prior density function for regression coefficient β_p; $\underline{\beta}_p$ represents a vector of regression coefficients except β_p, $p = 1,2,\ldots,$ a and likelihood function $L(.)$ is given by equation (20) or (21). 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 $\Gamma(\Psi, \Psi)$ is used as prior distribution for frailty parameter. 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) and Sahu et al. (1997) 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 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 Ψ = 0.0001; φ^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 equation (20), 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) and Gelman and Rubin (1992). 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. Burn-in period is the practical minimization of the initial values effect on the posterior inference and chain converges to stationary distribution by discarding the initial portion of Markov chain sample. 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 (1995).

7. SIMULATION STUDY

To assess the execution of the Bayesian estimation method a simulation study was carried out, 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. Gelman-Rubin scale reduction factor values were very close to one and p-values for Geweke test were huge, which sufficiently demonstrates that the chains achieve stationary distribution for both the prior sets. Further the convergence rate was not enormously diverse. There is no impact of prior distribution on posterior summaries because estimates of parameters were about the same for both sets of priors. For both the chains the results are 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. Table 2, 3, and 4 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.

8. DATA ANALYSIS

The applicability of the models was checked by applying them to the kidney infection data. The urinary organ infection knowledge has appeared in McGilchrist and Aisbett (1991). 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.

First, we check the goodness of fit for the kidney infection data by considering Kolmogorov Smirnov test and 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 (8), (10) and (12). The corresponding p-values are given in Table 1.

The outlines of the posterior results are displayed in Table 5, 6, 7, 8, 9 and 10. These tables consist of an estimate of parameters (posterior mean), which is highlighted in the 1st column, standard error in the 2nd column, credible limits in 3rd and 4th columns, Gelman Rubin values, Geweke values and p-values are in 5th, 6th and 7th columns. For all the frailty models, the value zero was not a credible value for all the credible intervals of the regression coefficients. So, all the covariates are significant in all frailty models, whereas, the value zero in the credible interval of the regression coefficient of X_4 in without frailty model, shows that Acute Nephritis is not significant covariates for a kidney infection. The negative value of regression coefficient β_2 implies that females have a lower risk of infection than males, while positive values of regression coefficients β_3, β_4 and β_5, shows the presence of disease types like Glomerulo Nephritis, Intense Nephritis, and Polycystic Kidney disease can increase kidney infection.

The comparison between the proposed models is done by utilizing AIC, BIC and DIC values given in Table 11. Table 11 and Table 12 shows that frailty models are better than without frailty models. The AIC, BIC and DIC values for model-I and model-III are comparatively much lesser than that for other models, so model-I or model-III may be better than model-II, model-IV, model-V and model-VI. The distinction between AIC,

BIC and DIC values for model-I and model-III are exceptionally little, so AIC, BIC and DIC values are not commendable to take a choice between the model-I and model-III.

Under the present study Bayes factor Df_{rv} is considered for comparing the models r and v. $Df_{rv} = 2\log(B_{rv})$ for the model-I against model-II is 21.7808; for model-I against model-III is 12.0108; for model-II against model-III is -9.7700. On comparison of model-I, model-II and model-III, Df_{12} and Df_{32} both are about 10 which manifest that model-II is not better than model-I and model-III affirming our earlier result. For model-I and model-III, D_{13} is 12.010 as given in Table 11, therefore model-I is better than model-III.

Hence, from all the demonstrated comparison criteria we can say model-I is better than model-II, model-III, model-IV, model-V and model-VI for modeling kidney infection data.

9. DISCUSSION

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

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. Self-composed programs in R statistical software have been utilized to perform the analysis.

All the demonstrated comparison criteria exhibits that additive hazard shared inverse Gaussian frailty demonstrated with generalized Pareto baseline is better for modeling of kidney infection data rather than generalized Rayleigh and xgamma as baseline. The estimates of frailty parameters ξ are high in all three models which are 3.7217, 34.923 and 7.4710 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 solid positive relationship between the two infection times for the same patient. Now, we have another additive shared frailty model with generalized Pareto distribution as baseline distribution for the analysis of kidney infection data.

REFERENCES

1. Aalen, O.O. (1980). 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. and Rosinski, J.), Springer, New York.
2. Aalen, O.O. (1989). A linear regression model for the analysis of lifetimes. *Stat. Med.*, 8, 907-925.

3. Clayton, D.G. (1978). A model for association in bivariate life tables and its applications to epidemiological studies of familial tendency in chronic disease incidence. *Biometrica*, 65, 141-151.
4. Davison, A.C. (1994). Modeling excesses over high threshold, with an application. In: *Statistical Extremes and Applications*, Ed. J. Tiago de Oliveira, 461-481. Reidel, Dordrecht, The Netherlands.
5. Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. B*, 52(3), 393-442.
6. Gelfand, A.E. and Ghosh, S.K. (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika*, 85, 1-11.
7. Gelman, A. and Rubin, D.B. (1992). A single series from the Gibbs sampler provides a false sense of security. *Bayesian Statistics*, 4, 625-631.
8. Geweke, J. (1992). 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, 169-193.
9. Haktanir, T. (1992). Comparison of various flood frequency distributions using annual peaks data of rivers in Anatolia. *J. Hydrol.*, 136, 1-31.
10. Hanagal, D.D. and Pandey, A. (2014). Inverse Gaussian shared frailty for modeling kidney infection data. *Advances in Reliability*, 1, 1-14.
11. Hanagal, D.D. and Pandey, A. (2016). Shared Gamma Frailty Models Based on Additive Hazards. *Journal of the Indian Society for Probability and Statistics*, 17(2), 161-184.
12. Hanagal, D.D. and Pandey, A. (2017). Shared inverse Gaussian frailty models based on additive hazards. *Communications in Statistics-Theory and Methods*, 46(22), 11143-11162.
13. Hougaard, P. (1984). Life Table Methods for Heterogeneous Populations. *Biometrika*, 71, 75-83.
14. Ibrahim, J.G., Chen, M.H. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, Verlag.
15. Kass, K.E. and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773-95.
16. Keyfitz, N. and Littman, G. (1979). Mortality in a Heterogeneous Population. *Population Studies*, 33, 333-342.
17. Kheiri, S., Kimber, A. and Meshkani, M.R. (2007). Bayesian Analysis of an Inverse Gaussian Correlated Frailty Model. *Computational Statistics and Data Analysis*, 51, 5317-5326.
18. Kundu, D. and Raqab, M.Z. (2006). Burr Type X distribution. *Journal of Probability and Statistical Sciences*, 4(2), 179-193.
19. Lancaster, T. and Nickell, S. (1980). The Analysis of Re-employment Probabilities for the Unemployed (with Discussion). *Journal of the Royal Statistical Society*, A, 143, 141-165.
20. McGilchrist, C.A. and Aisbett, C.W. (1991). Regression with frailty in survival analysis. *Biometrics*, 47, 461-466.
21. Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure data. *IEEE Transactions on Reliability*, 42, 299-302.

22. Sahu, S.K., Dey, D.K., Aslanidou, H. and Sinha, D.(1997). A Weibull regression model with gamma frailties for multivariate survival data. *Life time Data Analysis*, 3, 123-137.
23. Sen, S., Maiti, S.S. and Chandra, N. (2016). The xgamma Distribution: Statistical Properties and Application. *Journal of Modern Applied Statistical Methods*, 15(1), 774-788.
24. Sen, S., Chandra, N. and Maiti, S.S. (2017). The weighted xgamma distribution: Properties and application. *Journal of Reliability and Statistical Studies*, 10(1), 43-58.
25. Smith, R.L. (1984). Threshold methods for sample extremes, In: *Statistical Extremes and Applications*, Ed. J. Trago de Olivera. Reidel, Dordrecht, The Netherlands, 621-638.
26. Surles, J.G. and Padgett, W.J. (2001). Inference for reliability and stress-strength for a scaled Burr Type X distribution. *Lifetime Data Analysis*, 7, 187-200.
27. Van Monfort, M.A.J. and Otten, A. (1991). The first and the second effect of the extreme value distribution, EV1. *Stochastic Hydrol. Hudraul*., 5, 69-76.
28. Vaupel, J.W., Manton, K.G. and Stallaed, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439-454.

APPENDIX: TABLES AND FIGURES

Table 1
p-values of K-S Statistics for Goodness of Fit Test for Kidney Infection Data Set

Distribution	Recurrences Time	
	First	Second
Generalized Pareto	0.1398	0.1229
Generalized Rayleigh	0.2628	0.5722
Xgamma	0.2630	0.5721

Table 2
Inverse Gaussian Frailty with Generalized Pareto Distribution as Baseline (Simulation for Model-I)

Parameter (values)	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR Values	Geweke Values	p-values
Burn in period = 7800 ; lag=190							
α_1 (0.0031)	0.0032	0.0003	0.0026	0.0038	0.9999	0.0035	0.5014
α_2 (0.0031)	0.0030	0.0005	0.0020	0.0039	1.0001	0.0025	0.5010
λ_1 (21.40)	21.428	0.5939	20.446	22.368	1.0001	-0.0158	0.4936
λ_2 (20.71)	20.674	20.674	19.749	21.617	1.0017	0.0039	0.5015
γ_1 (0.006)	0.0059	0.0006	0.0050	0.0069	1.0001	-0.0004	0.4998
γ_2 0.0060	0.0059	0.0005	0.0050	0.0069	1.0011	0.0064	0.5025
ξ (3.72)	3.7183	3.7183	3.6298	3.8114	1.0006	0.0018	0.5007
β_1 (-0.108)	-0.1086	0.0232	-0.1555	-0.0656	1.0009	0.0019	0.5047

Table 3
Inverse Gaussian Frailty with Generalized Rayleigh Distribution as Baseline (Simulation for Model II)

Parameter (values)	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR values	Geweke values	p-values
Burn in period = 7100 ; lag=205							
α_1 (6.299)	6.2992	0.0319	6.2390	6.3607	1.0003	-0.0097	0.4961
α_2 (4.61)	4.6154	0.0573	4.5200	4.7046	1.0014	0.0172	0.5068
λ_1 (0.0027)	0.0027	0.0004	0.0020	0.0036	1.0054	0.0088	0.5035
λ_2 (0.0036)	0.0036	0.0004	0.0030	0.0046	1.0011	-0.0047	0.4981
ξ (35.12)	35.118	0.0555	35.025	35.215	0.9999	0.0087	0.5034
β_1 (0.0016)	0.0016	0.0004	0.0010	0.0025	1.0098	0.0061	0.5024

Table 4
Inverse Gaussian Frailty with xgamma Distribution as Baseline
(Simulation for Model III)

Parameter (values)	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR Values	Geweke values	p-values
Burn in period = 6600 ; lag=240							
α_1 (0.008)	0.0080	0.0003	0.0074	0.0087	0.9999	-0.0021	0.4991
α_2 (0.0091)	0.0090	0.0005	0.0080	0.0099	1.0003	-0.0030	0.4987
ξ (7.49)	7.4955	0.0585	7.4047	7.5955	1.0010	-0.0093	0.4962
β_1 (-0.057)	-0.0568	0.0071	-0.0693	-0.0457	1.0089	0.0031	0.5012

Table 5
Posterior Results for the Kidney Infection Data Set for Model I

Parameter	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR values	Geweke values	p-values
Burn in period = 7800 ; lag=160							
α_1	0.0032	0.0003	0.0025	0.0038	1.0001	-0.0022	0.4990
α_2	0.0030	0.0005	0.0020	0.0039	1.0004	-0.0058	0.4976
λ_1	21.391	0.5780	20.451	22.338	1.0025	-0.0044	0.4982
λ_2	20.702	0.5647	19.735	21.655	1.0042	0.0052	0.5021
γ_1	0.0059	0.0005	0.0050	0.0069	1.0025	-0.0012	0.4994
γ_2	0.0059	0.0005	0.0050	0.0069	0.9999	-0.0050	0.4979
ξ	3.7217	0.0529	3.6274	3.8135	1.0058	-0.0199	0.4920
β_1	-0.1127	0.0154	-0.1469	-0.0842	1.0000	-0.0047	0.4981
β_2	-6.6615	1.0207	-8.2949	-4.4569	1.0071	0.0057	0.4981
β_3	2.5510	0.0521	2.4581	2.6410	1.0016	-0.0030	0.4987
β_4	2.8489	0.0531	2.7558	2.9455	1.0041	-0.0066	0.4973
β_5	0.2979	0.0529	0.2068	0.3879	0.9999	0.0024	0.5009

Table 6
Posterior Results for the Kidney Infection Data Set for Model II

Parameter	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR Values	Geweke Values	p-values
Burn in period = 15000 ; lag=110							
α_1	6.5199	0.0315	6.4587	6.5814	1.0000	0.0004	0.5001
α_2	4.6087	0.0549	4.5177	4.7034	1.0003	-0.0034	0.4986
λ_1	0.0031	0.0004	0.0022	0.0039	1.0002	-0.0030	0.4987
λ_2	0.0038	0.0038	0.0030	0.0048	1.0001	-0.0042	0.4983
ξ	34.923	0.5151	33.986	35.830	1.0029	-0.0076	0.4969
β_1	0.0019	0.0005	0.0010	0.0029	1.0010	0.0024	0.5009
β_2	-2.7290	0.3649	-3.4217	-2.0477	1.0000	0.0053	0.5009
β_3	0.0301	0.0053	0.0207	0.0393	1.0008	-0.0092	0.4963
β_4	0.6691	0.0530	0.5773	0.7615	1.0004	0.0008	0.5003
β_5	0.0399	0.0052	0.0309	0.0491	1.0028	-0.0041	0.4983

Table 7
Posterior Results for the Kidney Infection Data Set for Model III

Parameter	Estimate	SE	Lower Credible Limit	Upper Credible Limit	GR values	Geweke values	p-values
Burn in period = 7600 ; lag=180							
α_1	0.0080	0.0003	0.0073	0.0086	1.0006	0.0057	0.5022
α_2	0.0090	0.0005	0.0080	0.0099	1.0004	0.0025	0.5103
ξ	7.4710	0.5515	6.5426	8.4171	1.0020	-0.0002	0.4999
β_1	-0.0593	0.0065	-0.0692	-0.0464	1.0034	-0.0150	0.4940
β_2	-2.8775	0.4455	-3.6445	-1.9899	1.0028	0.0059	0.4940
β_3	1.1416	0.0514	1.0501	1.2294	1.0014	0.0056	0.5022
β_4	1.9744	0.4898	1.0812	2.8979	1.0034	0.0020	0.5081
β_5	0.3873	0.0558	0.2923	0.4835	1.0000	-0.0053	0.4978

Table 8
Posterior Results for the Kidney Infection Data Set for Model IV

Parameters	Estimates	SE	Lower Credible Limit	Upper Credible Limit	GR Values	Geweke Values	p-values
Burn in period = 8200 ; lag=145							
α_1	0.5397	0.0336	0.4737	0.6093	1.0001	-0.0054	0.4978
α_2	0.5818	0.0572	0.4952	0.6830	1.0000	0.0026	0.5010
λ_1	12.424	0.5991	11.470	13.387	1.0000	-0.0113	0.4954
λ_2	10.065	0.5746	9.0509	10.947	1.0000	-0.0028	0.4988
γ_1	0.0069	0.0005	0.0058	0.0077	1.0012	-0.0023	0.4990
γ_2	0.0077	0.0005	0.0068	0.0087	1.0003	-0.0108	0.4956
β_1	-1.2943	0.4881	-2.1274	-0.5050	1.0005	0.0065	0.5026
β_2	-0.0974	0.0578	-0.1947	-0.0035	1.0020	0.0150	0.5026
β_3	0.5371	0.0560	0.4450	0.6324	1.0009	6.3e-05	0.5000
β_4	2.7692	0.0542	2.6757	-0.5050	1.0000	-0.0040	0.4983
β_5	1.3210	0.0585	1.2293	1.4205	1.0014	0.0021	0.5008

Table 9
Posterior Results for the Kidney Infection Data Set for Model V

Parameters	Estimates	SE	Lower Credible Limit	Upper Credible Limit	GR Values	Geweke Values	p-values
Burn in period = 6200 ; lag=280							
α_1	0.3200	0.0033	0.3136	0.3263	1.0000	0.0013	0.5005
α_2	0.5095	0.0054	0.5003	0.5196	1.0012	-0.0011	0.4995
λ_1	0.0027	0.0003	0.0020	0.0035	1.0001	-0.0114	0.4954
λ_2	0.0033	0.0004	0.0023	0.0039	1.0000	-0.0006	0.4997
β_1	-0.1076	0.0043	-0.1136	-0.0982	1.0091	-0.0010	0.4995
β_2	-4.4828	0.0635	-4.5918	-4.3689	0.9999	0.0065	0.4995
β_3	0.0300	0.0049	0.0209	0.0389	1.0004	0.0011	0.5046
β_4	0.3049	0.0518	0.2087	0.3947	1.0000	0.0016	0.5066
β_5	0.0307	0.0052	0.0214	0.0214	0.9999	-0.0004	0.4998

Table 10
Posterior Results for the Kidney Infection Data Set for Model VI

Parameters	Estimates	SE	Lower Credible Limit	Upper Credible Limit	GR Values	Geweke Values	p-values
Burn in period = 6500 ; lag=270							
α_1	0.0163	0.0018	0.0128	0.0202	1.0003	0.0013	0.5005
α_2	0.0181	0.0026	0.0130	0.0232	1.0003	0.0032	0.5012
β_1	-0.0836	0.0050	-0.0928	-0.0725	1.0030	-0.0085	0.4965
β_2	-4.5946	0.5188	-5.5136	-3.4842	1.0006	-0.0007	0.4965
β_3	0.0600	0.0050	0.0515	0.0693	1.0022	-0.0099	0.4960
β_4	0.0503	0.0050	0.0415	0.0599	1.0001	0.0102	0.4959
β_5	-3.8465	0.6197	-4.8477	-2.4814	1.0033	-0.0228	0.4908

Table 11: AICAIC, BIC and DIC Values for six Models

Baseline Distribution	Model	AIC	BIC	DIC	-logLikelihod
Generalized Pareto	With Frailty	687.9486	707.5996	667.2012	-331.9743
	Without Frailty	713.0356	731.0491	731.0491	-345.5178
Generalized Rayleigh	With Frailty	705.9384	722.3142	689.9339	-342.9692
	Without Frailty	707.7898	722.528	692.9076	-344.8949
Xgamma	With Frailty	692.8048	705.9055	680.339	-338.4024
	Without Frailty	728.9751	740.4382	720.553	-357.4875

Table 12: Bayes Factor Values and Decision for Shared Frailty Models Fitted to the Kidney Data Set

Numerator Model against Denominator Model	$B_{rv}= 2\log_e(B_{rv})$	range	Evidence against Model in Denominator
I against II	21.7808	≥ 10	Very strong positive
I against III	12.0108	≥ 10	Very strong positive
I against IV	24.2951	≥ 10	Very strong positive
I against V	26.0028	≥ 10	Very strong positive
I against VI	53.9072	≥ 10	Very strong positive
II against III	-9.7700	< 0	Strong negative
II against IV	2.5143	≥ 2 and ≤ 6	Positive
II against V	4.2220	≥ 2 and ≤ 6	Positive
II against VI	32.1264	≥ 10	Very strong positive
III against IV	12.2843	≥ 10	Very strong positive
III against V	13.9920	≥ 10	Very strong positive
III against VI	41.8964	≥ 10	Very strong positive
IV against V	1.7076	≥ 0 and ≤ 2	Not worth more than a bare mention
IV against VI	29.6121	≥ 10	Very strong positive
V against VI	27.9044	≥ 10	Very strong positive

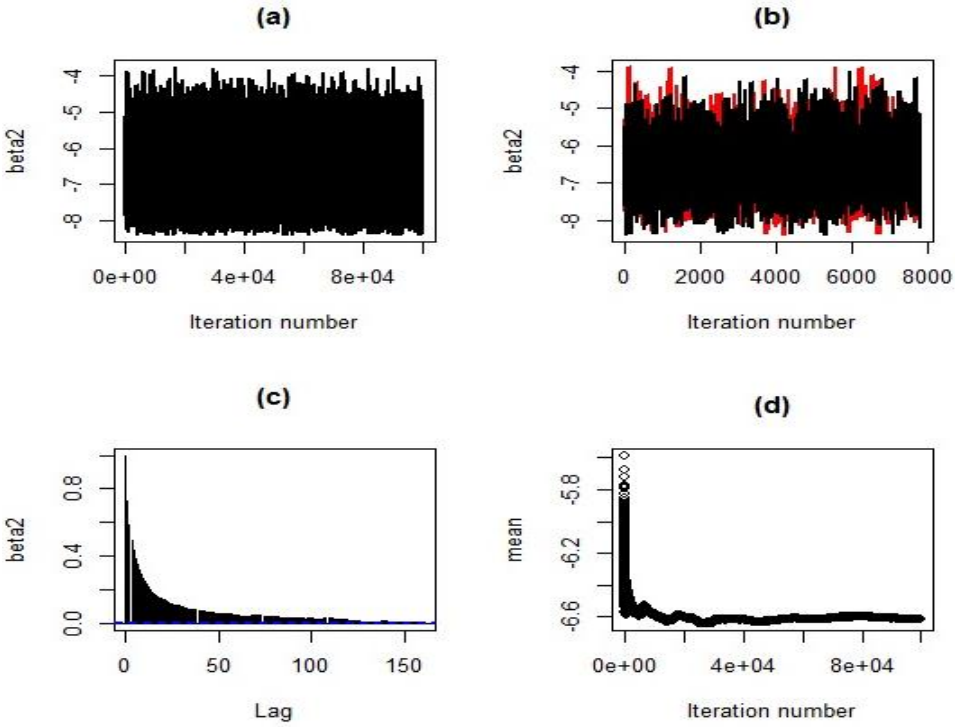

Figure 1: (a) Trace Plot
(b) Coupling from the Past Plot
(c) ACF Plot
(d) Running Mean Plot