

# Inverse Gaussian Shared Frailty Models with Generalized Exponential and Generalized Inverted Exponential as Baseline Distributions

David D. Hanagal and Arvind Pandey\*

<sup>1</sup> *Department of Statistics, University of Pune, Pune-411007, India.*

*Abstract:* The unknown or unobservable risk factors in the survival analysis cause heterogeneity between individuals. Frailty models are used in the survival analysis to account for the unobserved heterogeneity in individual risks to disease and death. To analyze the bivariate data on related survival times, the shared frailty models were suggested. The most common shared frailty model is a model in which frailty act multiplicatively on the hazard function. In this paper, we introduce the shared inverse Gaussian frailty model with the reversed hazard rate and the generalized inverted exponential distribution and the generalized exponential distribution as baseline distributions. We introduce the Bayesian estimation procedure using Markov Chain Monte Carlo(MCMC) technique 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. Also we apply the proposed models to the Australian twin data set and a better model is suggested.

*Key words:* Bayesian estimation, Generalized Exponential distribution, Generalized Inverted Exponential distribution, Inverse Gaussian frailty, MCMC, Reversed hazard rate.

## 1. Introduction

The statistical analysis of lifetime data plays an important role in medicine, epidemiology, biology, demography, economics, engineering and other fields. In the analysis of survival data, unknown or unobservable risk factors cause heterogeneity between individuals. This unobserved heterogeneity is known as frailty. Frailty is a common random effect that acts multiplicatively on the hazard rates of all subgroup members. The notion of frailty, introduced by Vaupel et al. (1979). The model assumes that the hazard function for lifetime  $T$  given an unobservable random variable,  $Z = z$  is  $h(t | z) = zh_0(t)$ ; where  $h_0(t)$  is the baseline hazard function.

Clayton (1978) introduced the notion of shared relative-risk for the analysis of survival data. Clayton and Cuzick (1985) introduced a bivariate model with shared relative-risk and observed

---

\* Corresponding author.

covariates. A shared frailty model with a positive stable distribution of frailty was suggested by Hougaard (1987).

Frailty model with gamma distribution has been studied in Vaupel et al. (1979) , Oakes (1982), Clayton and Cuzick (1985), and Andersen et al. (1993). The gamma distribution is the most commonly used frailty distribution, largely because of its mathematical convenience; see, for example, Hanagal (2006, 2007 & 2013). Another choice is the inverse Gaussian distribution. The inverse Gaussian makes the population homogeneous with time, whereas for gamma the relative heterogeneity is constant[ see Hougaard (1984)]. Duchateau and Janssen (2008) fit the inverse Gaussian (IG) frailty model with Weibull hazard to the udder quarter infection data. The IG distribution has a unimodal density and is a member of the exponential family. While its shape resembles that of other skewed density functions, such as lognormal and gamma, it provides much flexibility in modeling. Furthermore, there are many striking similarities between the statistics derived from this distribution and those of the normal; see Chhikara and Folks (1986). These properties make it potentially attractive for modeling purposes with survival data. The models derived above are bases on the assumption that a common random effect acts multiplicatively on the hazard rate function.

In many practical situations reversed hazard rate (RHR) is more appropriate to analyze the survival data. Reversed hazard rate was proposed as a dual to the hazard rate by Barlow et al. (1963). Shaked and Shantikumar (1994) and Block et al. (1998) provided a general definition of reversed hazard rate (RHR) as,

$$m(t) = \lim_{\Delta t} P(t - 4t < T \leq t | T \leq t) / \Delta t, \quad t > 0. \quad (1.1)$$

The reversed hazard rate specifies the instantaneous rate of death or failure at time t, given that it failed before time t. 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  $(t - \Delta t, t]$ . In lifetime data analysis, the concepts of reversed hazard rate has potential application when the time elapsed since failure is a quantity of interest in order to predict the actual time of failure. The reversed hazard rate is more useful in estimating reliability function when the data are left censored or right truncated. Reversed hazard rate plays a vital role in the analysis of parallel systems, in reliability and survival analysis. For example, in certain systems or situations, sometimes the failure is prevented through numerous safety measures [see Gleeja, 2008]. Andersen et al. (1993), Lawless (2003) have discussed the use of reversed hazard rate for the analysis of left censored or right truncated data.

Duffy et al. (1990) considered Australian twin data which consist of information about the age at appendectomy of monozygotic (MZ) and dizygotic (DZ) twins. There were some pairs with missing age at onset and those are the left censored observations. Duffy et al. (1990) excluded these left censored observations in the analysis. It is therefore, appropriate to model common random effect by including those left censored observations, which can be done by

developing frailty models using RHR. Accordingly, Sankaran and Gleeja (2011) introduced frailty as a common random effect that acts multiplicatively on reversed hazard rates, which is useful for the analysis of left censored data. In the present paper we introduce three parametric shared frailty model with inverse Gaussian frailty using reversed hazard rate. We use generalized exponential distributions and generalized inverted exponential distribution as baseline distribution and we compare these models for Australian twin data.

The remainder of the paper is organized as follows. In Section 2, we introduce the general shared frailty model with inverse Gaussian frailty. In Section 3 in we discuss the baseline distribution and proposed models in Section 4. In Section 5, we discuss how the MCMC technique is used to estimate the parameters of the proposed models and different model selection criteria. Section 6 provides simulation study. In Section 7, we present analysis of Australian twin data set. Section 8 contains the major conclusions of the study.

## 2. General Shared Frailty Model

The shared frailty model is relevant to event time of related individuals, similar organs and repeated measurements. For example, the failure time of paired organs like kidneys, lungs, eyes, ears, dental implants, etc. are considered as event times. In this model individuals from a group share common risks. For the shared frailty model it is assumed that survival times are conditionally independent, for a given shared frailty. The shared frailty means the dependence between the survival times is only due to unobservable covariates or frailty. When there is no variability in the distribution of frailty variable  $Z$ , then  $Z$  has a degenerate distribution and when the distribution of  $Z$  is not degenerate, there is a positive dependence.

Suppose  $n$  individuals are observed for the study and let the bivariate random vector  $(T_{1j}, T_{2j})$  represent the first and the second lifetimes of the  $j$ th individual ( $j = 1, 2, 3, \dots, n$ ). Also suppose that  $X_0, X_1$  and  $X_2$  are the observed covariates. Where  $X_0$  is the common covariate,  $X_1$  and  $X_2$  are the covariates corresponding to  $T_{1j}$  and  $T_{2j}$  respectively. Let a vector  $X_{1j} = (X_{11j}, \dots, X_{k1j})$ , ( $i = 0, 1, 2$ ) for the  $j$ th individual where  $X_{aj}$  ( $a = 1, 2, 3, \dots, k1$ ) represents the value of the  $a$ th observed covariate for the  $j$ th individual. We assume that the first and the second survival times for each individual share the same value of the covariates. Let  $Z_j$  be shared frailty for the  $j$ th individual. Assuming that the frailties are acting multiplicatively on the baseline reversed hazard rate and both the survival times of individuals are conditionally independent for a given frailty, the conditional reversed hazard rate for the  $j$ th individual at the  $i$ th ( $i = 1, 2$ ) survival time  $t_{ij}$  for a given frailty  $Z_j = z_j$  has the form,

$$m_i(t_{ij} | z_j, \underline{X}_j) = z_j m_{0i}(t_{ij}) e^{-X_{0j}\beta_0 - X_{ij}\beta_i}, i = 1, 2. \quad (2.1)$$

where  $m_{0i}(t_{ij})$  is the baseline reversed hazard at time  $t_{ij}$  and  $\beta$  is a vector of order  $k$ , of regression coefficients. The conditional cumulative reversed hazard rate for the  $j^{th}$  individual at  $i^{th}$  lifetime  $t_{ij}$  for a given frailty  $Z_j = z_j$  is,

$$M_i(t_{ij} | z_j, \underline{X}_j) = z_j M_{0i}(t_{ij}) \eta_{0j} \eta_{ij} \tag{2.2}$$

where  $\eta_{0j} = e^{X_{0j}\beta_0}$ ,  $\eta_{ij} = e^{X_{ij}\beta_i}$ ,  $i = 1, 2$  and  $M_{0i}(t_{ij})$  is the cumulative baseline reversed hazard rate at time  $t_{ij}$ . The conditional distribution function for the  $j$ th individual at the  $i$ th lifetime  $t_{ij}$  for a given frailty  $Z_j = z_j$  is,

$$\begin{aligned} F(t_{ij} | z_j, \underline{X}_j) &= e^{-M_i(t_{ij}|z_j, \underline{X}_j)} \\ &= e^{-z_j M_{0i}(t_{ij}) \eta_{0j} \eta_{ij}} \end{aligned} \tag{2.3}$$

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

$$\begin{aligned} F(t_{1j}, t_{2j} | \underline{X}_j) &= \int_{Z_j} F(t_{1j}, t_{2j} | z_j) f_Z(z_j) dz_j \\ &= \int_{Z_j} e^{-z_j (M_{01}(t_{1j}) \eta_{1j} + M_{02}(t_{2j}) \eta_{2j}) \eta_{0j}} f_Z(z_j) dz_j \\ &= L_{Z_j}[(M_{01}(t_{1j}) \eta_{1j} + M_{02}(t_{2j}) \eta_{2j}) \eta_{0j}] \end{aligned} \tag{2.5}$$

where  $L_{Z_j}(\cdot)$  is the Laplace transform of the frailty variable of  $Z_j$  for the  $j^{th}$  individual.

Here onwards we represent  $F(t_{1j}, t_{2j} | \underline{X}_j)$  as  $F(t_{1j}, t_{2j})$ .

### 2.1. Shared Inverse Gaussian Frailty Model

The gamma distribution is most commonly used frailty distribution because of its mathematical convenience. The inverse Gaussian makes the population homogeneous with time, whereas for gamma the relative heterogeneity is constant Hougaard (1984). Alternative to the gamma distribution Hougaard (1984) introduced the inverse Gaussian as a frailty distribution. The inverse Gaussian distribution have many similarities to standard Gaussian distribution (see Chikkara and Folks (1986)). Furthermore, it provides much flexibility in modeling, when early occurrences of failures are dominant in a life time distribution and its failure rate is expected to be non monotonic. In such situations the inverse Gaussian distribution might provide a suitable choice for the lifetime model.

Also inverse Gaussian is almost an increasing failure rate distribution when it is slightly skewed and hence is also applicable to describe lifetime distribution which is not dominated by early failures. Secondly, for the inverse Gaussian distribution the surviving population becomes more homogeneous with respect to time, where as for gamma distribution the relative heterogeneity is constant. The inverse Gaussian distribution has unimodal density and is the member of exponential family. While its shape resembles the other skewed density functions, such as log normal and gamma. These properties of inverse Gaussian distribution motivate us to use inverse Gaussian as frailty distribution. The inverse Gaussian distribution has a history dating back to 1915 when Schrodinger and Smoluchowski presented independent derivations of the density of the first passage time distribution of Brownian motion with positive drift. Villman et al., (1990) have studied the histomorphometrical analysis of the influence of soft diet on masticatory muscle development in the muscular dystrophic mouse. The muscle fibre size distributions were fitted by an inverse Gaussian law. Barndorff-Nielsen (1994) considers a finite tree whose edges are endowed with random resistances, and shows that, subject to suitable restrictions on the parameters, if the resistances are either inverse Gaussian or reciprocal inverse Gaussian random variables, then the overall resistance of the tree follows a reciprocal inverse Gaussian law. Gacula and Kubala (1975) have analyzed shelf life of several products using the IG law and found to be a good fit. For more real life applications see Seshadri(1999).

Let a continuous random variable  $Z$  follows the inverse Gaussian distribution with parameters  $\mu$  and  $\theta$  then the density function of  $Z$  is

$$f_Z(z) = \begin{cases} \left[ \frac{1}{2\pi\theta} \right]^{\frac{1}{2}} z^{-\frac{3}{2}} e^{-\frac{(z-\mu)^2}{2z\theta\mu^2}} & ; z > 0, \mu > 0, \theta > 0 \\ 0 & ; otherwise, \end{cases} \quad (2.6)$$

and the Laplace transform is

$$L_Z(s) = \exp \left[ \frac{1}{\mu\theta} - \left( \frac{1}{\theta^2\mu^2} + \frac{2s}{\theta} \right)^{\frac{1}{2}} \right]. \quad (2.7)$$

The mean and variance of frailty variable are  $E(Z) = \mu$  and  $V(Z) = \mu^3\theta$ . For identifiability, we assume  $Z$  has expected value equal to one i.e.  $\mu = 1$ . Under this restriction, the density function and the Laplace transformation of the inverse Gaussian distribution reduces to

$$f_Z(z) = \begin{cases} \left[ \frac{1}{2\pi\theta} \right]^{\frac{1}{2}} z^{-\frac{3}{2}} e^{-\frac{(z-1)^2}{2z\theta}} & ; z > 0, \theta > 0 \\ 0 & ; otherwise, \end{cases} \quad (2.8)$$

and the Laplace transform is

$$L_Z(s) = \exp \left[ \frac{1 - (1 + 2\theta s)^{\frac{1}{2}}}{\theta} \right]. \quad (2.9)$$

with the variance of  $Z$  as  $\theta$ . The frailty variable  $Z$  is degenerate at  $Z = 1$  when  $\theta$  tends to zero. Replacing the Laplace transform in equation (2.5), we get the unconditional bivariate distribution function for the  $j^{th}$  individual as

$$F(t_{1j}, t_{2j}) = \exp \left[ \frac{1 - (1 + 2\theta\eta_{0j}(M_{01}(t_{1j})\eta_{1j} + M_{02}(t_{2j})\eta_{2j}))^{\frac{1}{2}}}{\theta} \right]. \quad (2.10)$$

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

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

$$F(t_{1j}, t_{2j}) = e^{-\eta_{0j}((M_{01}(t_{1j})\eta_{1j} + M_{02}(t_{2j})\eta_{2j}))}. \quad (2.11)$$

### 3. Baseline Distributions

We present below three baseline distributions with the interesting properties. The goodness of fit based on Kolmogorov-Smirnov (K-S) statistic in Table 4 shows that these four baseline models fit well to the Australian data set.

#### 3.1. Generalized Exponential distribution

The one parameter exponential distribution is one of the most widely used lifetime models in reliability and survival analysis because of its simple mathematical form and some interesting properties. Gupta and Kundu (1999) generalized this model by introducing a shape parameter, which

is known as the generalized exponential distribution. We have used the generalized exponential distribution as the baseline distribution. A continuous random variable  $T$  is said to follow the generalized exponential distribution if its distribution function and probability density function is,

$$F_0(t) = \begin{cases} (1 - e^{-\lambda t})^\alpha & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; otherwise. \end{cases} \quad (3.1)$$

and

$$f_0(t) = \begin{cases} \alpha \lambda e^{-\lambda t} (1 - e^{-\lambda t})^{\alpha-1} & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; otherwise. \end{cases} \quad (3.2)$$

where  $\lambda$  and  $\alpha$  are respectively scale and shape parameters of the distribution. The reversed hazard rate and the cumulative reversed hazard rate are respectively,

$$m_0(t) = \begin{cases} \frac{\alpha \lambda e^{-\lambda t} (1 - e^{-\lambda t})^{\alpha-1}}{(1 - e^{-\lambda t})^\alpha} & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; otherwise. \end{cases} \quad (3.3)$$

$$M_0(t) = \begin{cases} -\alpha \ln(1 - e^{-\lambda t}) & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; otherwise. \end{cases} \quad (3.4)$$

For  $\alpha = 1$ , the distribution reduces to one parameter exponential distribution. If  $\alpha < 1$ ,  $f(x)$  decreases monotonically with  $x$ . If  $\alpha > 1$ ,  $f(x)$  attains a mode at  $x = (\log \alpha) / \lambda$ . When  $\alpha > 1$ , the hazard function is an increasing function of time and for  $\alpha < 1$ , hazard function is a decreasing function of time. Generalized exponential distribution can be used effectively in analyzing many life time data sets particularly in place of gamma and Weibull. For generalized exponential distribution, hazard rate increases from zero to a finite constant, when shape parameter  $\alpha$  increases and hazard rate decreases from  $\infty$  to a finite number when is less than one. A nearly constant rate after a certain time period implies that the occurrence of failure is purely random and is independent of past life; this is a property of the failure rate of an exponential distribution which has been extensively used in reliability studies. So we thought generalized exponential distribution as baseline distribution may gives better results comparable to Weibull distribution

### 3.2. Generalized inverted Exponential distribution

Another modification to exponential distribution has been done by using its inverted version, known as the inverted exponential distribution (IED) and was studied by Lin et al. (1989). They obtained the maximum likelihood estimator, confidence limits and UMVUE for the parameter and the reliability function using complete samples. They also compared this model with that of inverted Gaussian and log-normal distributions based on a maintenance data set. The cumulative distribution function (cdf) of the generalized inverted exponential distribution (GIED) with shape parameter  $\alpha$  and scale parameter  $\lambda$  is given by

$$F_0(t) = \begin{cases} 1 - (1 - e^{-\frac{\lambda}{t}})^\alpha & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; otherwise. \end{cases} \quad (3.5)$$

$$f_0(t) = \begin{cases} \frac{\alpha\lambda}{t^2} e^{-\frac{\lambda}{t}} (1 - e^{-\frac{\lambda}{t}})^{\alpha-1} & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; otherwise. \end{cases} \quad (3.6)$$

where  $\lambda$  and  $\alpha$  are respectively scale and shape parameters of the distribution. The reversed hazard rate and the cumulative reversed hazard rate are respectively,

$$m_0(t) = \begin{cases} \frac{\frac{\alpha\lambda}{t^2} e^{-\frac{\lambda}{t}} (1 - e^{-\frac{\lambda}{t}})^{\alpha-1}}{1 - (1 - e^{-\frac{\lambda}{t}})^\alpha} & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; otherwise. \end{cases} \quad (3.7)$$

$$M_0(t) = \begin{cases} -\ln(1 - (1 - e^{-\frac{\lambda}{t}})^\alpha) & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; otherwise. \end{cases} \quad (3.8)$$

Several interesting properties of GIED have been studied in detail by Abouammoh and Alshingiti (2009) and Nadarajah and Kotz (2003). The hazard rate functions of GIED can be increasing, or decreasing but not constant depending on the value of the shape parameter. The hazard rate functions of GIED can be increasing, or decreasing but not constant depending on the value of the shape parameter. The GIED has a unimodal and right skewed density function for the shape parameter greater than 4. Moreover, they observed that in many situations, the GIED provides a better fit than

gamma, Weibull, generalized exponential and inverted exponential distributions (see Abouammoh and Alshingiti (2009)). Recently, Krishna and Kumar (2013) studied reliability estimation in the context of this distribution under progressively type II censored sample and Deya and Pradhanb (2014) discuss Bayes estimates are evaluated by applying Lindleys approximation method, the importance sampling procedure and MetropolisHastings algorithm. The importance sampling technique is used to compute the highest posterior density credible intervals. Finally, they discuss a method of obtaining the optimum hybrid censoring scheme.

## 4. Proposed Models

### 4.1. Parametric Models

Substituting the cumulative reversed hazard function for the generalized exponential and the generalized inverted exponential as baseline distributions in equation (2.10) and equation (2.11), we get the unconditional bivariate distribution functions at time  $t_{1j} > 0$  and  $t_{2j} > 0$  as,

$$F(t_{1j}, t_{2j}) = \exp \left[ \frac{1 - (1 - 2\theta\eta_{0j}(\alpha_1 \ln(1 - e^{-\lambda_1 t_{1j}})\eta_{1j} + \alpha_2 \ln(1 - e^{-\lambda_2 t_{2j}})\eta_{2j}))^{\frac{1}{2}}}{\theta} \right]. \quad (4.1)$$

$$F(t_{1j}, t_{2j}) = \exp \left( \eta_{0j}(\alpha_1 \ln(1 - e^{-\lambda_1 t_{1j}})\eta_{1j} + \alpha_2 \ln(1 - e^{-\lambda_2 t_{2j}})\eta_{2j}) \right). \quad (4.2)$$

$$F(t_{1j}, t_{2j}) = \exp \left[ \frac{1 - (1 - 2\theta\eta_{0j}(\eta_{1j} \ln(1 - (1 - e^{-\frac{\lambda_1}{t_{1j}}})^{\alpha_1}) + \eta_{2j} \ln(1 - (1 - e^{-\frac{\lambda_2}{t_{2j}}})^{\alpha_2})))^{\frac{1}{2}}}{\theta} \right]. \quad (4.3)$$

$$F(t_{1j}, t_{2j}) = \exp \left( \eta_{0j} \left( \eta_{1j} \ln(1 - (1 - e^{-\frac{\lambda_1}{t_{1j}}})^{\alpha_1}) + \eta_{2j} \ln(1 - (1 - e^{-\frac{\lambda_2}{t_{2j}}})^{\alpha_2}) \right) \right). \quad (4.4)$$

Here onwards we call equation (4.1), (4.2), (4.3) and (4.4) as Model I, Model II, Model III and Model IV respectively. Model I and Model III are inverse Gaussian frailty models with baseline distributions as generalized exponential and generalized inverted exponential respectively and Model II and Model IV are the corresponding baseline distributions without frailty models.

## 4.2. Semiparametric Models

The semi parametric model with frailty is given by

$$F(t_{1j}, t_{2j}) = \exp \{ -z\eta_{0j} (\eta_{1j}M_{01}(t_{1j}) + \eta_{2j}M_{02}(t_{2j})) \} \quad (4.7)$$

and the semi parametric model without frailty is given by

$$F(t_{1j}, t_{2j}) = \exp \{ -\eta_{0j} (\eta_{1j}M_{01}(t_{1j}) + \eta_{2j}M_{02}(t_{2j})) \} \quad (4.8)$$

for more details see Hanagal(2011). Here onwards we call Eqns (4.5) and (4.6) as Model V and Model VI which correspond to the inverse Gaussian frailty with semiparametric model and semiparametric model without frailty respectively.

## 5. Likelihood Specification and Bayesian Estimation of Parameters

Suppose there are n individuals under study, whose first and second observed failure times are represented by  $(t_{1j}, t_{2j})$ . Let  $c_{1j}$  and  $c_{2j}$  be the observed censoring times for the  $j$ th individual ( $j = 1, 2, 3, \dots, n$ ) for the first and the second recurrence times respectively. We use the left censoring scheme. Also we assume independence between the censoring scheme and the lifetimes of individuals.

The contribution of the bivariate lifetime random variable of the  $j$ th individual in likelihood function is given by,

$$L_j(t_{1j}, t_{2j}) = \begin{cases} f_1(t_{1j}, t_{2j}), & t_{1j} > c_{1j}, t_{2j} > c_{2j}, \\ f_2(t_{1j}, c_{2j}), & t_{1j} > c_{1j}, t_{2j} < c_{2j}, \\ f_3(c_{1j}, t_{2j}), & t_{1j} < c_{1j}, t_{2j} > c_{2j}, \\ f_4(c_{1j}, c_{2j}), & t_{1j} < c_{1j}, t_{2j} < c_{2j}. \end{cases}$$

and likelihood function is,

$$L(\underline{\psi}, \underline{\beta}, \theta) = \prod_{j=1}^{n_1} f_1(t_{1j}, t_{2j}) \prod_{j=1}^{n_2} f_2(t_{1j}, c_{2j}) \prod_{j=1}^{n_3} f_3(c_{1j}, t_{2j}) \prod_{j=1}^{n_4} f_4(c_{1j}, c_{2j}) \quad (5.1)$$

where  $\theta$ ,  $\psi$  and  $\beta$  are respectively the frailty parameter, the vector of baseline parameters and the vector of regression coefficients. The counts  $n_1, n_2, n_3$  and  $n_4$  be the numbers of individuals for which first and second failure times  $(t_{1j}, t_{2j})$  lie in the ranges  $t_{1j} > c_{1j}, t_{2j} > c_{2j}; t_{1j} > c_{1j}, t_{2j} < c_{2j}; t_{1j} < c_{1j}, t_{2j} > c_{2j}$  and  $t_{1j} < c_{1j}, t_{2j} < c_{2j}$  respectively and let for inverse Gaussian frailty,

$$\begin{aligned} f_1(t_{1j}, t_{2j}) &= \frac{\partial^2 F(t_{1j}, t_{2j})}{\partial t_{1j} \partial t_{2j}} = \frac{m_{01}(t_{1j})m_{02}(t_{2j})F(t_{1j}, t_{2j})\phi_1(t_{1j}, t_{2j})\eta_0^2\eta_1\eta_2}{[\phi_2(t_{1j}, t_{2j})]^{\frac{3}{2}}} \\ f_2(t_{1j}, c_{2j}) &= \frac{\partial F(t_{1j}, c_{2j})}{\partial t_{1j}} = \frac{m_{01}(t_{1j})F(t_{1j}, c_{2j})\phi_1(t_{1j}, c_{2j})\eta_0\eta_1}{[\phi_2(t_{1j}, c_{2j})]^{\frac{1}{2}}} \\ f_3(c_{1j}, t_{2j}) &= \frac{\partial F(c_{1j}, t_{2j})}{\partial t_{2j}} = \frac{m_{02}(t_{2j})F(c_{1j}, t_{2j})\phi_1(c_{1j}, t_{2j})\eta_0\eta_2}{[\phi_2(c_{1j}, t_{2j})]^{\frac{1}{2}}} \\ \text{and } f_4(c_{1j}, c_{2j}) &= F(c_{1j}, c_{2j}) \end{aligned} \quad (5.2)$$

where  $\phi_1(a_j, b_j) = 1 + \theta[1 - \ln(F(a_j, b_j))]$  and  $\phi_2(a_j, b_j) = 1 + 2\theta\eta_0(M_{01}(a_j)\eta_1 + M_{02}(b_j)\eta_2)$ . Substituting the reversed hazard functions  $m_{01}(t_{1j})$ ,  $m_{02}(t_{2j})$ , the distribution function  $F(t_{1j}, t_{2j})$  and the cumulative reversed hazard functions  $M_{01}(t_{1j})$  and  $M_{02}(t_{2j})$  for the baseline distributions, we get the likelihood function given by Eq. (5.1). Similarly we get the likelihood function for without frailty model.

Unfortunately computing the maximum likelihood estimators (MLEs) involves solving a 7 to 8 dimensional optimization problem for all models. There is problem of convergence of the estimates in the Newton-Raphson iterative procedure because we are estimating parameters simultaneously. As the method of maximum likelihood fails to estimate the parameters due to convergence problem, we use Bayesian approach. For all six models we use the Bayesian approach. The traditional maximum likelihood approach to estimation is commonly used in survival analysis, but it can encounter difficulties with frailty models. Moreover, standard maximum likelihood based inference methods may not be suitable for small sample sizes or situations in which there is heavy censoring (see Kheiri et al. (2007)). Thus, in our problem a Bayesian approach, which does not suffer from these difficulties, is a natural one, even though it is relatively computationally intensive.

Several authors have discussed the Bayesian approach for the estimation of parameters of the frailty models. Some of them are, Ibrahim et al. (2001) and references therein. Santos and Achcar (2010) considered the parametric models with Weibull and the generalized gamma distribution as the baseline distribution and gamma and log-normal as the frailty distributions. Ibrahim et al. (2001) and references therein considered the Weibull model and the piecewise exponential model with the gamma frailty. They also considered the positive stable frailty models.

The joint posterior density function of the parameters for given failure times in the proposed frailty models is obtained as,

$$\pi(\lambda_1, \alpha_1, \lambda_2, \alpha_2, \theta, \beta) \propto L(\lambda_1, \alpha_1, \lambda_2, \alpha_2, \theta, \beta) g_1(\lambda_1) g_2(\alpha_1) \times g_3(\lambda_2) g_4(\alpha_2) g_5(\theta) \prod_{i=1}^3 p_i(\beta_i)$$

where  $g_i(\cdot)$  ( $i = 1, 2, \dots, 5$ ) indicates the prior density function with known hyper parameters of corresponding argument for the baseline parameters and the frailty variance;  $p_i(\cdot)$  is the prior density function for the regression coefficient  $\beta_i$  and the likelihood function  $L(\cdot)$  is given by equation (5.1). Here we assume that all the parameters are independently distributed. In the same way one can write the joint posterior density function of the parameters in the models without frailty.

To estimate the parameters of the model, we used the Metropolis-Hastings algorithm and Gibbs sampler. We monitored the convergence of a Markov chain to a stationary distribution by the Gelman-Rubin convergence statistic and the Geweke test. The trace plots, the coupling from the past plots and the sample autocorrelation plots are used to check the behaviour of the chain, to decide the burn-in period and the autocorrelation lag respectively.

The algorithm consists in successively obtaining a sample from the conditional distribution of each of the parameter given all other parameters of the model. These distributions are known as full conditional distributions. In our case full conditional distributions are not easy to integrate out. So full conditional distributions are obtained by considering that they are proportional to the joint distribution of the parameters of the model.

We have full conditional distribution of the parameter  $\lambda_1$  as,

$$\pi_1(\lambda_1 | \alpha_1, \lambda_2, \alpha_2, \theta, \beta) \propto L(\lambda_1, \alpha_1, \lambda_2, \alpha_2, \theta, \beta) \cdot g_1(\lambda_1) \tag{5.3}$$

Similarly full conditional distributions for other parameters can be obtained even if parameters are less.

In order to compare the proposed models, we use the Bayesian Information Criteria (BIC), the Akaike Information Criteria (AIC), the Deviance Information Criteria (DIC) and the Bayes factor. The BIC was introduced by Schwarz(1978) and the BIC is defined as,

$$BIC = D(\hat{\Theta}) + p \cdot \log(n) \tag{5.4}$$

where  $p$  represents the number of parameters of the model and  $n$  represents the number of data points.  $D(\hat{\Theta})$  represents an estimate of the deviance evaluated at the posterior mean

$\hat{\Theta} = E(\Theta | \text{data})$ . The deviance is defined by,  $D(\Theta) = -2 \cdot \log L(\Theta)$ , where  $\Theta$  is a vector of the unknown parameters of the model and  $L(\Theta)$  is the likelihood function of the model. AIC was introduced by Akaike(1973) and AIC is defined as,

$$AIC = D(\hat{\Theta}) + 2p. \tag{5.5}$$

DIC, a generalization of AIC was introduced by Spiegelhalter et al.(2002) and is defined as,

$$DIC = D(\hat{\theta}) + 2 \cdot p_D \quad (5.6)$$

where  $p_D$  is the difference between the posterior mean of the deviance and the deviance of the posterior mean of the parameters of interest, that is,  $p_D = D - D(\hat{\theta})$ , where  $D = E(D(\theta) | data)$ .

The Bayesian model examination for adequacy and model comparison can be proceeds by the predictive distribution. Let  $y = \{y_1, y_2, \dots, y_n\}$  be a set of observations, where  $n$  is total number of observations and  $y_{obs}$  denotes realization of  $y$ . The posterior predictive density  $\pi(y | y_{obs})$  is the predictive density of a new independent set of observations under the model, given the actual set of observations. By marginalizing  $\pi(y | y_{obs})$ , we obtain the posterior predictive density of a single observation  $y_r, r = 1, 2, \dots, n$  as follows,

$$\pi(y_r | y_{obs}) = \int \pi(y_r | \theta) \pi(\theta | y_{obs}) d\theta \quad (5.7)$$

A simple checking for assessment of model is predictive interval. Suppose we generate a sample  $yr_1, yr_2, \dots, yr_n$  from the predictive density (5.7) for the  $r$ th observation and create the  $100(1 - \alpha)\%$  equal tailed credible interval also known as the predictive interval, then the model under consideration would be an adequate model for data if  $100(1 - \alpha)\%$  of the  $yr, obs$  to fall in their respective interval.

To draw a random sample from the predictive density (5.7), suppose we have  $\theta_j^*$ ; ( $j = 1, 2, \dots, n$ )  $n$  samples from the posterior density  $\pi(\theta | y)$  possibly using one of the MCMC methods. Then a random sample  $y_r^j$  drawn from  $\pi(y_r | \theta_j^*)$  is a sample from the predictive density (5.7), since for given parameters  $\theta$ , if observations are conditionally independent then  $\pi(y_r | y, \theta) = \pi(y_r | \theta)$ .

Another approach for model selection is based on cross-validation predictive density. The cross-validation predictive densities are the set  $\{f(y_r | \underline{Y}_{(r)}); r = 1, 2, \dots, n\}$  where  $\underline{Y}_{(r)}$  denotes all elements of data set  $y$  except observation  $y_r$  and

$$f(y_r | \underline{Y}_{(r)}) = \int f(y_r | \theta, \underline{Y}_{(r)}) f(\theta | \underline{Y}_{(r)}) d\theta \quad (5.8)$$

$f(y_{r,obs} | \underline{Y}_{(r),obs})$  is popularly known as the conditional predictive ordinate (CPO). The smaller values of CPO does not support the model, so we prefer a model for which CPO values are higher than others. We can compare more than one models using CPO's. We plot CPO's versus  $r$  for different models in a single graph and compare the models visually. If we plot difference of CPO values for the models A and B, i.e.  $CPO_A - CPO_B$  then negative differences favour model B where positive difference favour model A. The larger the positive or negative difference better the model A or B.

Sometimes CPO values are quite close to each other so that difference may be clustered at zero and plot can not be distinguishable. To overcome this situation one can use log of CPO values to plot. For the cross-validation predictive density, in general we can write,

$$\begin{aligned} f(y_r | \underline{Y}_{(r)}) &= \frac{f(\underline{y})}{f(\underline{Y}_{(r)})} \\ &= \frac{1}{\int \frac{f(\underline{Y}_{(r)} | \underline{\theta})}{f(\underline{y}, \underline{\theta})} f(\underline{\theta} | \underline{y}) d\underline{\theta}} \\ &= \frac{1}{\int \frac{1}{f(y_r | \underline{Y}_{(r)}, \underline{\theta})} f(\underline{\theta} | \underline{y}) d\underline{\theta}} \end{aligned}$$

So an immediate Monte Carlo estimate of CPO, is given by,

$$\frac{n}{\sum_{j=1}^n \frac{1}{f(y_{r,obs} | \theta_j^*)}} \tag{5.9}$$

which is the harmonic mean of the conditional density function of  $y_r$  evaluated at the posterior sample values. See Gelfand (1996) for more details.

Gelfand and Ghosh (1998) have proposed a model choice criterion by studying utility functions. They consider loss functions which reward an action for its closeness to the predictive value and penalizes the action if it is too far from the observed value. The criterion is then obtained by minimizing this posterior predictive loss. As they claim, the criterion emerges approximately as a form, partitioned into a goodness-of-fit term and a penalty term for a wide range of models. With squared error loss the criterion is:

$$D'_\omega = \sum_{r=1}^n \sigma_r^2 + \frac{\omega}{\omega + 1} \sum_{r=1}^n (\mu_r - y_{r,obs})^2 \tag{5.10}$$

where  $\mu_r$  and  $\sigma_r^2$  are the Monte Carlo estimates of the posterior predictive mean and the variance of  $y_r$  under the density (5.7) based on the sample  $y_r^j, j = 1, 2, \dots, n$  and  $\omega > 0$  is constant. The first term is a penalty term which penalizes both under-fitted and overfitted models, since the predictive variances in such cases will tend to be larger. The second term without the factor involving  $\omega$  is a goodness-of-fit measure. Model selection using  $D'_\omega$  is usually not sensitive to  $\omega$ . For censored data, the criterion must be modified because  $y_{r,obs}$  is not available for censored cases. The modified criterion is,

$$D_{\omega} = \sum_{r=1}^n \sigma_r^2 + \frac{\omega}{\omega + 1} \sum_{r=1}^n (\mu_r - v_r)^2 \quad (5.11)$$

where  $v_r = y_{r,obs}$  if the  $r^{th}$  observation is a failure time and  $v_r = \max(\mu_r, c_r)$  if the  $r^{th}$  observation is censored at  $c_r$ . A model with minimum value of  $D_{\omega}$  is selected as the best model among all the models considered.

The Bayes factor  $B_{jk}$  for a model  $M_j$  against  $M_k$  for a given data  $D = (t_{1j}, t_{2j})$ ; ( $j = 1, 2, 3, \dots, n$ ) is

$$B_{jk} = \frac{P(D|M_j)}{P(D|M_k)} \quad (5.12)$$

where  $P(D|M_k) = \int_S P(D|M_k)\pi(\theta_k|M_k)d\theta_k$ ; ( $k = 1, 2, 3, \dots, m$ ) where  $\theta_k$  is the vector of unknown parameters of model  $M_k$ ,  $\pi(\theta_k|M_k)$  is the prior density and  $S$  is the support of the parameter  $\theta_k$ . Here  $m$  represents the model number. Raftery (1994), following Jeffreys (1961), proposes the rules of thumb for interpreting twice the logarithm of the Bayes factor. For two models of substantive interest,  $M_j$  and  $M_k$ , twice the log of the Bayes factor is approximately equal to the difference in their BIC approximations.

To compute Bayes factor we need to obtain  $I_k = P(D|M_k)$ , we consider one of the approach given in Kass and Raftery (1995), a MCMC estimate of  $I_k$  is given by,

$$\hat{I}_k = \left\{ \frac{\sum_{i=1}^N P(D|\theta^{(i)})^{-1}}{N} \right\}^{-1} \quad (5.13)$$

which is harmonic mean of the likelihood values. Here  $N$  represents the posterior sample size and  $\theta^{(i)}$  ( $i = 1, 2, \dots, N$ ) is the sample from the prior distribution.

## 6. Simulation Study

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 binomial distribution for Model I and Model III. The frailty variable  $Z$  is assumed to have inverse Gaussian distribution for all models with known variance. Lifetimes  $(T_{1j}, T_{2j})$  for

the  $j$ th individual are conditionally independent for a given frailty  $Z_j = z_j$ . We assume that  $T_{ij}(i = 1, 2, \dots, j = 1, 2, \dots, n)$  follows one of the baseline distributions given above.

As the Bayesian methods are time consuming, we generate only fifty pairs of lifetimes. According to the assumption, for a given frailty ( $Z$ ), lifetimes of individuals are independent. Thus the conditional distribution function for an individual for given a frailty,  $Z = z$  and a covariate  $X_0$  at time  $t > 0$  is,

$$F(t | z, X_0) = e^{-z M_0(t) \eta}$$

where  $\eta = e^{X_0 \beta_0}$ .

For Model-I, equating  $F(t) = e^{-z \eta \alpha \ln(1 - e^{-\lambda t})} = (1 - e^{-\lambda t})^{z \eta \alpha}$  to a random number say  $u$  ( $0 < u < 1$ ), we get,

$$t = -\frac{1}{\lambda} \ln \left\{ 1 - u^{\frac{1}{z \eta \alpha}} \right\} \quad (6.1)$$

where  $z$  is the inverse Gaussian frailty.

For Model-III, equating  $F(t) = e^{z \eta \ln(1 - (1 - e^{-\frac{\lambda}{t}})^{\alpha})} = (1 - (1 - e^{-\frac{\lambda}{t}})^{\alpha})^{z \eta}$  to a random number say  $u$  ( $0 < u < 1$ ), we get,

$$t = \frac{-\lambda}{\ln(1 - (1 - r^{\frac{1}{z \eta}})^{\frac{1}{\alpha}})} \quad (6.2)$$

where  $z$  is the inverse Gaussian frailty.

Samples are generated using the following procedure;

1. Generate 50 covariate values for  $X_0$  from the binomial distribution for all models.
2. Compute  $\eta = e^{X_0 \beta_0}$  with the regression coefficient ( known ).
3. Generate a sample of size 50 for the frailty variable  $Z$  from the inverse Gaussian distribution with  $\theta = 2.0$ .
4. Generate 50 pairs of lifetimes  $(t_{1j}, t_{2j})$  for given frailty  $Z$  using the following generators, For Model-I,

$$t_{1j} = -\frac{1}{\lambda_1} \ln \left\{ 1 - u_{1j}^{\frac{1}{z_j \eta_j \alpha_1}} \right\}$$

$$t_{2j} = -\frac{1}{\lambda_2} \ln \left\{ 1 - u_{2j}^{\frac{1}{z_j \eta_j \alpha_2}} \right\}$$

For Model-III,

$$t_{1j} = \frac{-\lambda_1}{\ln(1 - (1 - r_1^{\frac{1}{z_j \eta_j}})^{\frac{1}{\alpha_1}})}$$

$$t_{2j} = \frac{-\lambda_2}{\ln(1 - (1 - r_2^{\frac{1}{z_j \eta_j}})^{\frac{1}{\alpha_2}})}$$

where  $z_j$  is the inverse Gaussian frailty. For Model-I and Model-III,  $u_1$  and  $u_2$  are the random variables having  $U(0,1)$  distribution and  $\alpha_1$  and  $\lambda_1$  are parameters of the baseline distribution for the first survival time and  $\alpha_2$  and  $\lambda_2$  are that of the second survival time.

5. Generate the censoring times ( $c_{1j}$  and  $c_{2j}$ ). For Model I, from the exponential distribution with the parameter 0.1 each and for Model III, from the exponential distribution with the parameter 0.2 each.
6. Observe the  $i^{\text{th}}$  survival time  $t_{ij}^* = \max(t_{ij}, c_{ij})$  and the censoring indicator  $\delta_{ij}$  for the  $j^{\text{th}}$  individual ( $i = 1, 2$  and  $j = 1, 2, \dots, 50$ ), where

$$\delta_{ij} = \begin{cases} 1, & ; t_{ij} > c_{ij} \\ 0, & ; t_{ij} < c_{ij} \end{cases}$$

Thus we have data consisting of 50 pairs of survival times ( $t_{1j}^*, t_{2j}^*$ ) and the censoring indicators  $\delta_{ij}$ .

A widely used prior for the frailty parameter  $\theta$  is the  $G(0.0001, 0.0001)$ . In addition, we assume that the prior for the regression coefficient is  $N(0, 1000)$ . 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  $\theta$  and the regression coefficients  $\beta$ . Since we do not have any prior information about the baseline parameters,  $\lambda_1, \alpha_1, \lambda_2$  and  $\alpha_2$ , the prior distributions are assumed to be flat. We consider two different non-informative prior distributions for the baseline parameters, one is  $G(a_1, a_2)$  and another is  $U(b_1, b_2)$ . All the hyper-parameters  $a_1, a_2, b_1$  and  $b_2$  are known. Here  $G(a, b)$  is a gamma distribution with the shape parameter  $a$  and the scale parameter  $b$  and  $U(b_1, b_2)$  represents the uniform distribution over the interval  $b_1$  to  $b_2$ . We only compare the comparison of this model with proposed frailty models. For Model I, we set  $\alpha_1 = 20.0, \lambda_1 = 0.05, \gamma_1 = 1.20, \alpha_2 = 20.0, \lambda_2 = 0.05, \gamma_2 = 1.20, \theta = 2.0$  and  $\beta = 0.80$  and  $X \sim \text{Binomial}(1, .5)$ , censoring distributions as the exponential distributions with the parameter 1.0 each. For Model III, we set  $\lambda_1 = 50, \alpha_1 = 20, \lambda_2 = 50, \alpha_2 = 20, \theta = 2.0$  and  $\beta = 0.50$  and  $X \sim \text{Binomial}(1, .45)$ , censoring distributions as the exponential distributions with the parameter 2.9 each. For Model V, We assume the value of the hyper-parameters as  $a_1 = 1, a_2 = 0.0001, b_1 = 0$  and  $b_2 = 100$ .

We run two parallel chains for all models 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 100000 times. There is no effect of the prior distribution on the posterior summaries because the estimates of the 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 running mean plots and the sample autocorrelation plots for the simulation study are not provided due to lack of space. Tables 1 and 2 present the estimates, the credible intervals, the Geweke test (Geweke 1992) and the Gelman-Rubin Statistics (Gelman and Rubin, 1992) for all the parameters of the Model I and Model III respectively based on the simulation study.

Estimated values of the parameters are close the true values and the bias for the estimated values are small. Standard errors for the estimated values are quit 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. Simulated values of the parameters have the autocorrelation of lag  $k$ , so every  $k$ th iteration is selected as a sample from the posterior distribution.

## 7. Australian Twin Data

Now we apply the all eight models to the Australian twin data given in Duffy et al. (1990). The data consists of six zygote categories. We consider the subset of the data with zygote category 2. The data consists of males gender only and consist of 567 pair of twins with 23 and 17 censored in twin 1 and twin 2 respectively. An individuals 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 factor or environmental factor involved in the risk of appendectomy is the frailty variable. Here there is a common covariate age for both T1 and T2 and one covariate each for T1 and T2, i.e., presence or absence of appendectomy. To check goodness of fit of Australian twin data set, We obtain Kolmogorov-Smirnov (K-S) statistics and their p values for T1 and T2 separately for two frailty models Model I and Model III. For all the models p values of K-S statistics are provided in Table 3. These p values of K-S test are quite high. We can say that there is no statistical evidence to the reject the hypothesis that data are from these models for marginal distributions and assume that they also fit for bivariate case. Figure 1 shows the parametric plot with semi parametric plot for all parametric models with frailty.

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 100000 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  $\theta$ . Due to lack of space we are presenting only for Model I ( the trace plots, the coupling from the past plots and the autocorrelation plots after thinning ) for the parameters. The trace plots for all the parameters shows zigzag pattern which indicates that parameters move and mix more freely, as shown in Figure 2. Thus, it seems that the Markov chain has reached the stationary state. Burn in period is decided by using coupling from the past plot (see Figure 3). However, a sequence of draws after burn-in period may have autocorrelation. Because of autocorrelation consecutive draws may not be random, but values at widely separated time points are approximately independent. So, a pseudo random sample from the posterior distribution can be found by taking values from a single run of the Markov chain at widely spaced time points (autocorrelation lag) after burn-in period. The autocorrelation of the parameters become almost negligible after the certain lag. ACF plot after thinning show that observations are independent, as shown in Figure 4. Thus, our diagnostic plots suggest that the MCMC chains are mixing very well. The Gelman-Rubin 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 attains 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 49. The posterior summary of the Model I to Model VI are given in Tables 4 to 9. Tables 4 to 9 present the estimates, the credible intervals, the Geweke test and the Gelman-Rubin statistics for all the parameters of

the Model I to Model VI respectively. To compare six models, we first use AIC, BIC and DIC values which are given in Table 10 and Bayes factor in Table 11. For Model I and Model III the estimates of the shared frailty parameter,  $\theta$  are respectively 318.04 and 3.251. This shows that there is a heterogeneity between the pairs of twins. From Table 10 and Table 11, we observe that parametric models are fit better than semi-parametric models based on AIC, BIC, DIC and Baye's factor. We observe that parametric Bayes factor for Model I with Model II is 561.55 , for Model III and Model IV is 130.238. This shows that the parametric frailty models are better than parametric without frailty models. Model I is better than Model III. This is also a Bayesian test based on Bayes factor for testing  $\theta = 0$  against  $\theta > 0$  and which supports the alternative hypothesis, i.e., models with frailty fit better. The credible interval of the regression coefficient  $\beta_0$  does not contains zero for Model II and Model III. The credible interval of the regression coefficient  $\beta_1$  contains zero for all models. The credible interval of the regression coefficient  $\beta_2$  contains zero for all six models. Hence age is the significant covariate for Model II and Model III. The convergence rate of the Gibbs sampling algorithm does not depend on these choices of the prior distributions in our proposed model for Australian twin data. The Geweke test values are near to zero and the corresponding p-values are quite high and the Gelman-Rubin Statistics for all the parameters of all models based on data are very close to one.

The AIC, BIC and DIC values for Model I is least among all six models. On the basis of AIC, BIC and DIC values, the parametric models are better than semi-parametric models. On the basis of AIC, BIC and DIC values, Model I is the best among all six models. Similarly the Bayes factors show that models with frailty (Model I and Model III) are better than the models without frailty (Model II and Model IV ) and Model I is the best and the frailty is significant. To check the adequacy of the Model I and Model III, we have constructed 99%,95%,90%,75% and 50% equal tailed predictive intervals of the generated random sample from the predictive distribution and counted the total number of intervals in which rth observation falls in their respective intervals. Details are given in Table 12. Table 12 shows that these two models are adequate for the Australian Twin Data.

Another diagnostic we now use is the CPO plot. We have plotted difference of log of CPO values for pair of models. Figure 5 represents the plot. In Figure 5 consider Model I with Model III which shows that Model I is better than Model III.

Finally, the values relating to the model choice criteria  $D_\omega$  (Eq. 5.7) for Model I and Model III are shown in Table 13. The first and second columns of the Table 13 give the penalty term and the goodness of fit term, next four columns give  $D_\omega$  values for different values of  $\omega = 1, 5, 10$  and  $\infty$ . Penalty term and goodness of fit term are minimum for Model I, also for all the values of  $\omega$   $D_\omega$  is minimum for Model I, so this criteria also suggest Model I, the inverse Gaussian frailty with the generalized exponential is best model.

## 8. Conclusion

Our main aim of the study is to examine the role of the bivariate frailty model based on the reversed hazard rate in survival studies. For this we used the inverse Gaussian frailty with two baseline distribution (the generalized exponential and the generalized inverted exponential) and these models are compared with their baseline models based on reversed hazard rate. 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 models on the basis of AIC, BIC and DIC values for Australian twin data set. Bayes factor also supports the shared frailty models.

Initially we thought to use the method of maximum likelihood to estimate the parameters but likelihood equations do not converge and the method of maximum likelihood fails to estimate the parameters so we used the Bayesian approach. In this study, the model is specified in a Bayesian framework and estimated with the MCMC algorithms. We have discussed the Bayesian estimation procedure including Gibbs sampling for computing the estimation of the unknown parameters by simulating samples of size 50. We have clearly written the steps involved in the iteration procedure. The entire estimation procedure using the Bayesian approach took large amount of computational time. The estimates of the parameters are not dependent on the different prior distributions.

Two different chains were run for the proposed models from different starting points using the Metropolis-Hastings algorithm within Gibbs sampler. We have provided 100,000 iterations to perform the simulation study. Estimates were calculated after discarding a burn-in interval for each chain. Trace plots for all the parameters shows zigzag pattern which indicates that parameters move freely. The quality of convergence was checked by Gelman-Rubin statistics (see Brooks and Gelman, 1998). The values of the Gelman-Rubin statistics in this case are quite close to one and also the Geweke test values are small with large p-values. Thus the sample can be considered to have arisen from stationary distribution and descriptive statistics can be seen as valid estimates of unknown parameters. The simulation results indicate that the performance of the Bayesian estimation method is quite satisfactory. Bayes factor is used to test the frailty parameter  $\theta = 0$  and it is observed that the frailty parameter is highly significant in all frailty models. From Table 10 it is clear that the models with frailty fit better than without frailty models and Model

I is best among the all six models. The covariate age is the only significant covariate for Model II and Model III.

The choice of the best model for Australian twin data is based on AIC, BIC, DIC and Bayes factor values. We found that Model I is a better Model on the basis of AIC, BIC,

DIC and Bayes factor values. The age is the only significant covariate for Model I and Model III. Shared frailty models (Model I and Model III) are better than their baseline model. Also Model I is better than Model III. On the basis of AIC, BIC, DIC and Bayes factor Model I is the best as compared other proposed models considered in this paper. On the basis of CPO plot and Eq (5.7) support the model I. Also shows that Model I is better than Model III. Also parametric models

are better than semi parametric models. By referring all the above analysis now we are in a position to say that, shared inverse Gaussian frailty model based on reversed hazard rate with the generalized exponential distribution as baseline is more suitable model for Australian twin data set, with left censored observations. The methods discussed in this paper may be extended into other frailty models and correlated frailty models with different base line distributions, using the Bayesian approach, provided the models fit to the data.

## References

- [1] Abouammoh, A.M. and Alshingiti, A.M. (2009). Reliability of generalized inverted exponential distribution, *Journal of Statistical Computation and Simulation* (79) , 1301-1315.
- [2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267-281). Budapest: Akademiai Kiado.
- [3] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Methods Based on Counting Processes*, Springer Verlag, New York.
- [4] Barlow, R.E., Marshall, A.W. and Proschan, F. (1963). Properties of the probability distribution with monotone hazard rate. *Ann Math Statist*, 34, 375-389.
- [5] Barndorff-Nielsen, O. E. (1994). A note on electrical networks. *Advances in Applied Probability*, 26, 63-67.
- [6] Block, H.W., Savits, T.H. and Singh, H. (1998). "On the reversed hazard rate function", *Probability in the Engineering and Informational Sciences*, 12, 69- 90.
- [7] Chhikara, R.S. and Folks, J.L. (1986). *The inverse Gaussian distribution*. Marcel Dekker, New York.
- [8] 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. *Biometrika*, 65, 141-151.
- [9] Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of Royal Statistical Society, Ser., A*, 148 , 82-117.
- [10] Deya, S. and Pradhan, B. (2014). Generalized inverted exponential distribution under hybrid censoring. *Statistical Methodology* (18), 1011-114.
- [11] Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. New York: Springer.
- [12] Duffy, D. L., Martin, N. G. and Mathews, J. D. (1990). Appendectomy in Australian twins. *Australian Journal of Human Genetics*, 47(3), 590-92.
- [13] Gacula, M. C. Jr. and Kubala, J. J. (1975). Statistical models for shelf life failures. *Journal of Food Science*, 40, 404-9.
- [14] Gelfand, A. E. and Ghosh, S. K. (1998). Model choice : A minimum posterior predictive loss approach. *Biometrika*, 85, 1-11.

- Gelman, A. and Rubin, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) . Oxford Univ. Press. pp 625-632.
- [15]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, pp 169-193.
- [16]Gleeja, V. L.(2008). Department Modelling and Analysis of Bivariate Lifetime Data using Reversed Hazard Rates. Ph. D. Thesis, Cochin University of Science and Technology, India.
- [17]Hanagal, D. D. (2006). A gamma frailty regression model in bivariate survival data.
- [18]IAPQR Transactions, 31, 73-83.
- [19]Hanagal, D. D. (2007). Gamma frailty regression models in mixture distributions. *Economic Quality Control*, 22 (2), 295-302.
- [20]Hanagal, D. D. (2011). *Modeling Survival Data Using Frailty Models*. Chapman & Hall/CRC.
- [21]Hanagal, D. D. and Dabade, A. D. (2013). Bayesian estimation of parameters and comparison of shared gamma frailty models. *Communications in Statistics, Simulation & Computation*, 42(4), 910-931.
- [22]Hougaard, P. (1984). Life table methods for heterogeneous populations. *Biometrika*, 71(1), 75-83.
- [23]Hougaard, P. (1987). Modelling Multivariate Survival. *Scandinavian Journal of Statistics*, 14(4), 291-304.[27]Hougaard, P. (2000). *Analysis of Multivariate survival Data*. Springer-Verlag, New York.
- [24]Ibrahim, J.G., Ming-Hui C. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, Verlag.
- [25]Jeffreys, H. (1961). *Theory of Probability*. 3rd edn., Oxford: Oxford University Press. [27] Kass, R.E. and Raftery, A.E. (1995): Bayes Factor. *Journal of the American Statistical Association*, 90, 430, 773-795.
- [26]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.
- [27]Krishna, H. and Kumar, K. (2013). Reliability estimation in generalized inverted exponential distribution with progressively type II censored sample, *Journal of Statistical Computation and Simulation*, 83(6), 10071019.
- [28]Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edition. John Wiley and Sons.
- [29]Lin, C.T., Duran, B.S. and Lewis, T.O.(1989). Inverted gamma as a life distribution, *Microelectronics and Reliability* (29) 619626.
- [30]Nadarajah, S. and Kotz, S. (2003). The exponentiated Frechet distribution, *Interstat*.
- [31]Oakes, D. (1982). A model for association in bivariate survival data. *Journal of Royal Statistical Society, B*, 44, 414-422.
- [32]Raftery, A. E. (1994). Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models. Technical Report, 255, Department of Statistics, University of Washington.

[33]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.

[34]Sankaran P.G. and Gleeja V.L. (2011). On proportional reversed hazards frailty models, Metron, 69, 151-173.

[35]Santos, C.A. and Achcar, J.A. (2010). A Bayesian analysis for multivariate survival data in the presence of covariates. Journal of Statistical Theory and Applications, 9 , 233-253.

[36]Schwarz. G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461-464.

[37]Seshadri, V. (1999).The Inverse Gaussian Distribution: Statistical Theory and Applications. Springer Science, New York

[38]Shaked, M. and Shantikumar, J.G. (1994). Stochastic Orders and Their Applications. Academic Press, New York.

[39]Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A. (2002): Bayesian measure of model complexity and fit (with discussion). Journal of the Royal Statistical Society, B. 64, 583-639.

[40]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.

[41]Vilmann, H., Kirkeby, S., and Kronberg, D. (1990). Histomorphometrical analysis of the influence of soft diet on masticatory muscle development in the muscular dystrophic mouse. Archives of oral Biology, 35(1), 37-42.

Received April 15, 2015; accepted June 19, 2015

David D. Hanagal , Arvind Pandey  
 Department of Statistics,  
 University of Pune, Pune-411007, India.

**Appendix : Summary Tables and Figures.**

Table 1: Baseline Distribution Generalized Exponential Distribution with Inverse Gaussian Frailty (Simulation for Model I)

Parameter (True value)	Estimate	Standard error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 2650; autocorrelation lag = 70							
$\alpha_1$ (20.0)	19.7978	1.1095	18.1058	21.8266	0.01864	0.5074	0.999
$\alpha_2$ (20.0)	20.2036	1.0508	18.2302	21.8838	-0.00537	0.4978	1.002
$\lambda_1$ (0.05)	0.0479	0.0031	0.0417	0.0539	-0.00716	0.4971	1.001
$\lambda_2$ (0.05)	0.0498	0.0032	0.0435	0.0565	-0.01660	0.4933	1.006
$\theta$ (2.00)	1.9914	0.1022	1.8136	2.1805	0.00592	0.5023	1.004
$\beta$ (0.50)	0.4955	0.0501	0.4081	0.5906	0.00886	0.5035	1.001

Table 2: Baseline Distribution Inverted Generalized Exponential Distribution with Inverse Gaussian Frailty (Simulation for Model III)

Parameter (True value)	Estimate	Standard error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 2700;		autocorrelation lag = 220					
$\alpha_1$ (20.0)	20.0501	1.1235	18.0924	21.8050	-0.00718	0.4971	1.000
$\alpha_2$ (20.0)	19.0847	0.8832	18.0447	21.2582	0.00876	0.5034	1.004
$\lambda_1$ (50.0)	51.0562	2.3318	46.6004	54.7153	0.00283	0.5011	1.004
$\lambda_2$ (50.0)	52.8335	1.6706	48.7157	54.8576	0.00727	0.5029	1.000
$\theta$ (2.00)	1.9988	0.1045	1.8235	2.1942	-0.00058	0.4997	1.000
$\beta$ (0.50)	0.4990	0.0540	0.4100	0.5939	-0.00591	0.4976	0.999

Table 3: p-values of K-S statistics for goodness of fit test for Australian twin data set

<i>Model</i>	$T_1$ <i>p - value</i>	$T_2$ <i>p - value</i>
Model I	0.55274	0.72409
Model III	0.41554	0.45502

Table 4: Posterior summary for Australian twin data set Model I

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 2500;		autocorrelation lag = 200					
$\alpha_1$	12698.25	390.8	12045.58	13359.01	-2.530e-05	0.4999	1.000
$\lambda_1$	0.1855	0.0045	0.1769	0.1945	0.00427	0.5017	1.008
$\alpha_2$	13681.94	366.4	13064.6	14342.98	0.00656	0.5026	1.000
$\lambda_2$	0.1864	0.0046	0.1772	0.1957	0.00423	0.5016	1.005
$\theta$	318.04	59.961	214.99	455.39	0.01118	0.5044	1.006
$\beta_0$	0.00037	0.0047	-0.0086	0.0089	0.01539	0.5061	1.000
$\beta_1$	0.00040	0.0246	-0.0436	0.0431	7.7e-05	0.5000	1.001
$\beta_2$	-0.0886	0.0787	-0.2304	0.0717	-0.00899	0.4964	1.010

Table 5: Posterior summary for Australian twin data set Model II

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6800; autocorrelation lag = 300							
$\alpha_1$	6.93902	0.26666	6.49089	7.4364	0.001298	0.50051	1.002965
$\lambda_1$	0.08417	0.00239	0.07922	0.0889	-0.011876	0.49526	1.008942
$\alpha_2$	7.65518	0.48385	6.74666	8.6348	-0.001748	0.49930	1.001512
$\lambda_2$	0.08689	0.00275	0.08158	0.0915	-0.009608	0.49617	0.999964
$\beta_0$	0.18469	0.08149	0.03381	0.3362	0.003245	0.50129	1.001766
$\beta_1$	0.00999	0.01436	-0.01836	0.0377	-0.000741	0.49970	1.002486
$\beta_2$	-0.00110	0.00133	-0.00363	0.0016	-0.012236	0.49512	1.000593

Table 6: Posterior summary for Australian twin data set Model III

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 4800; autocorrelation lag = 400							
$\alpha_1$	16.273	1.0495	14.4111	18.1231	-0.00929	0.4962	1.007
$\lambda_1$	86.271	3.6867	79.7175	94.1369	-0.00546	0.4978	1.014
$\alpha_2$	16.342	0.9251	14.6021	18.0726	0.00848	0.5033	1.002
$\lambda_2$	86.996	3.6117	80.5161	94.1552	0.00125	0.5004	1.000
$\theta$	3.251	0.4601	2.3816	4.1647	0.00168	0.5006	1.002
$\beta_0$	0.0287	0.0037	0.0202	0.0356	0.00246	0.5009	1.009
$\beta_1$	0.0032	0.0242	-0.0443	0.0477	-0.01077	0.4957	1.000
$\beta_2$	-0.0672	0.0751	-0.2241	0.0766	0.00591	0.5023	1.007

Table 7: Posterior summary for Australian twin data set Model IV

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6800; autocorrelation lag = 220							
$\alpha_1$	7.9165	0.7444	6.5666	9.3499	-0.01195	0.4952	1.000
$\lambda_1$	70.218	3.7680	63.3697	77.7286	-0.01095	0.4956	1.001
$\alpha_2$	8.4689	0.7100	7.0999	9.8922	-0.00588	0.4976	1.000
$\lambda_2$	73.228	3.5597	66.1471	80.4128	-0.00751	0.4970	1.000
$\beta_0$	-0.0027	0.0017	-0.0058	0.0006	0.00836	0.5033	1.005
$\beta_1$	0.0076	0.0247	-0.04398	0.0475	6.009e-05	0.5000	1.008
$\beta_2$	-0.0648	0.0737	-0.2136	0.0740	0.001229	0.5004	1.001

Table 8: Posterior summary for Australian twin data set Model V

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 4000; autocorrelation lag = 100							
$\theta$	6.179392	1.0882	4.148583	8.21719	0.0138028	0.50550	1.0025
$\beta_0$	0.000116	0.0055	-0.009277	0.00938	0.0166271	0.50663	1.0000
$\beta_1$	0.006033	0.2350	-0.374677	0.37676	0.0038946	0.50155	1.0005
$\beta_2$	0.013092	0.3378	-0.565301	0.57159	0.0043858	0.50174	1.0003

Table 9: Posterior summary for Australian twin data set Model VI

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6800; autocorrelation lag = 100							
$\beta_0$	0.00118684	0.005628	-0.009245	0.009676	0.000465	0.50018	1.0008
$\beta_1$	0.00185114	0.230795	-0.377088	0.374764	0.015327	0.50611	1.0026
$\beta_2$	0.01744018	0.338022	-0.557324	0.575124	-0.002975	0.49881	1.0043

Table 10: AIC, BIC and DIC Comparison

Model	AIC	BIC	DIC
Model- I	8480.493	8519.972	8474.005
Model- II	8900.544	9057.116	9128.43
Model-III	8780.672	8820.346	8774.574
Model- IV	8907.057	8942.74	8903.658
Model- V	13002.67	13021.88	13058.43
Model- VI	19082.24	19092.07	19904.33

Table 11: Bayes Factor for six models

$B_{ij}$	$M_{1,2}$	$M_{1,3}$	$M_{1,4}$	$M_{1,5}$	$M_{1,6}$	$M_{2,3}$	$M_{2,4}$	$M_{2,5}$
Bayes Factor	561.55	297.96	428.206	4761.69	16264.66	-263.58	-133.34	4200.14
$B_{ij}$	$M_{2,6}$	$M_{3,4}$	$M_{3,5}$	$M_{3,6}$	$M_{4,5}$	$M_{4,6}$	$M_{5,6}$	-
Bayes Factor	15703.107	130.238	4463.728	15966.69	4333.489	15836.45	11502.96	-

$$M_{i,j} = 2ln \left[ \frac{\hat{I}_i}{\hat{I}_j} \right]$$

Table 12: Predictive interval for frailty models.

Model	99%	95%	90%	75%	50%
Model I	1128	1103	1070	940	850
Model III	1125	1102	1076	998	881

Table 13: Model selection criteria (5.7) \* 10<sup>-3</sup> for testing frailty  $\theta$

Model	Penalty term	Goodness-of-fit term	$D_1$	$D_5$	$D_{10}$	$D_\infty$
Model I	201.527	221.759	312.407	386.327	403.127	423.287
Model III	238.147	230.037	353.166	429.845	447.273	468.185

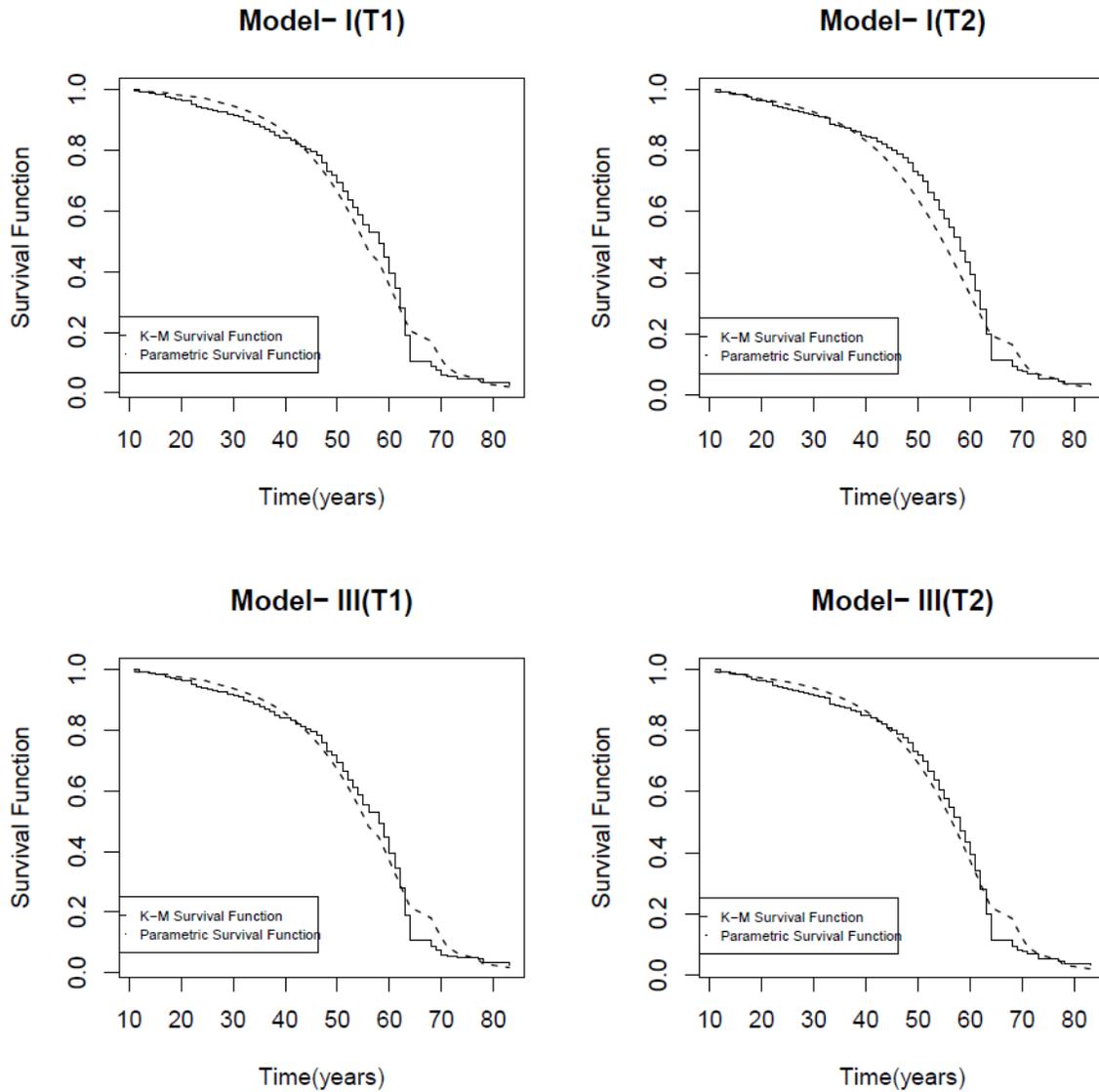


Figure 1: Survival function plots for (K-M survival and parametric survival).

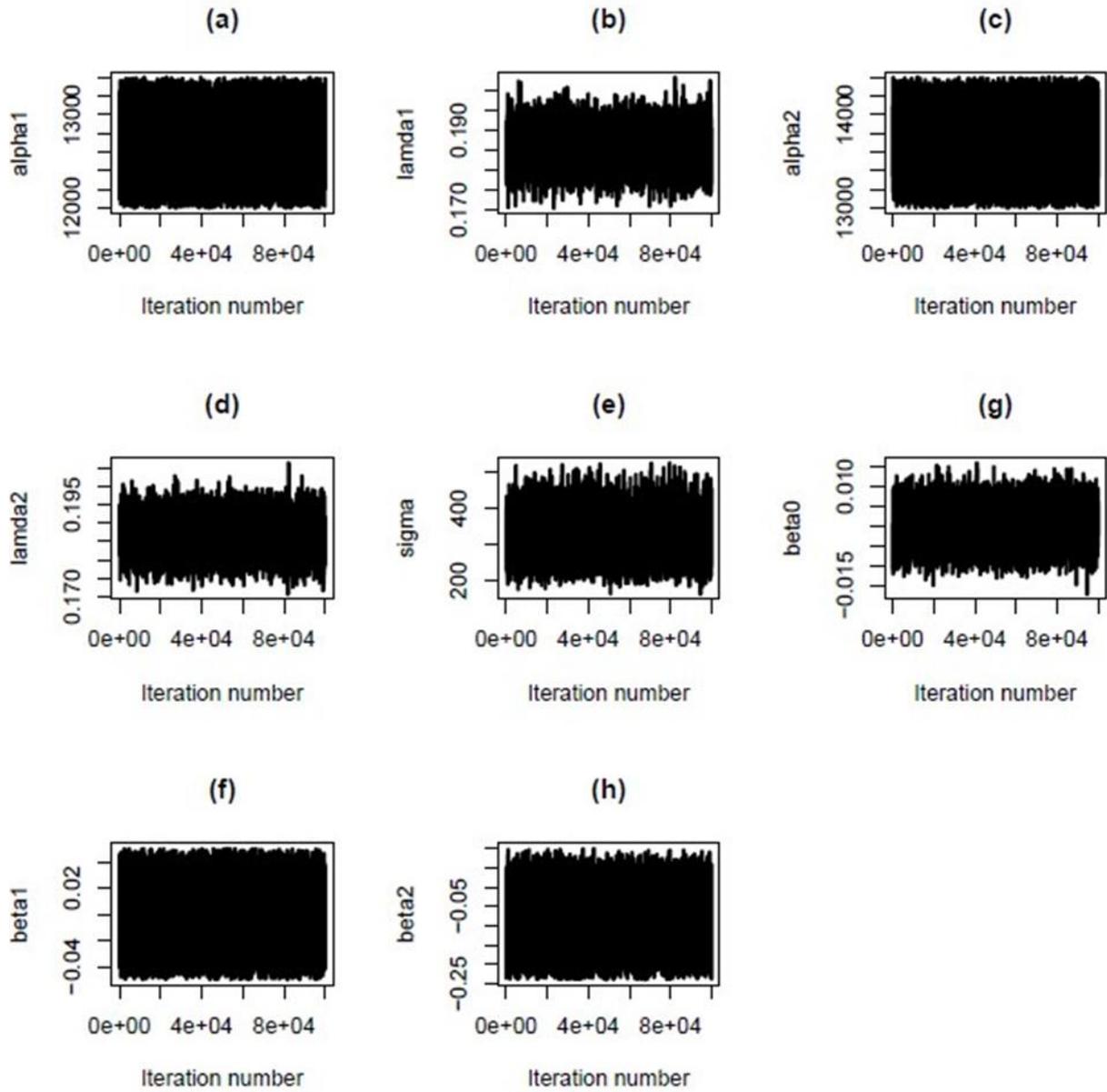


Figure 2: Trace plots for Model I

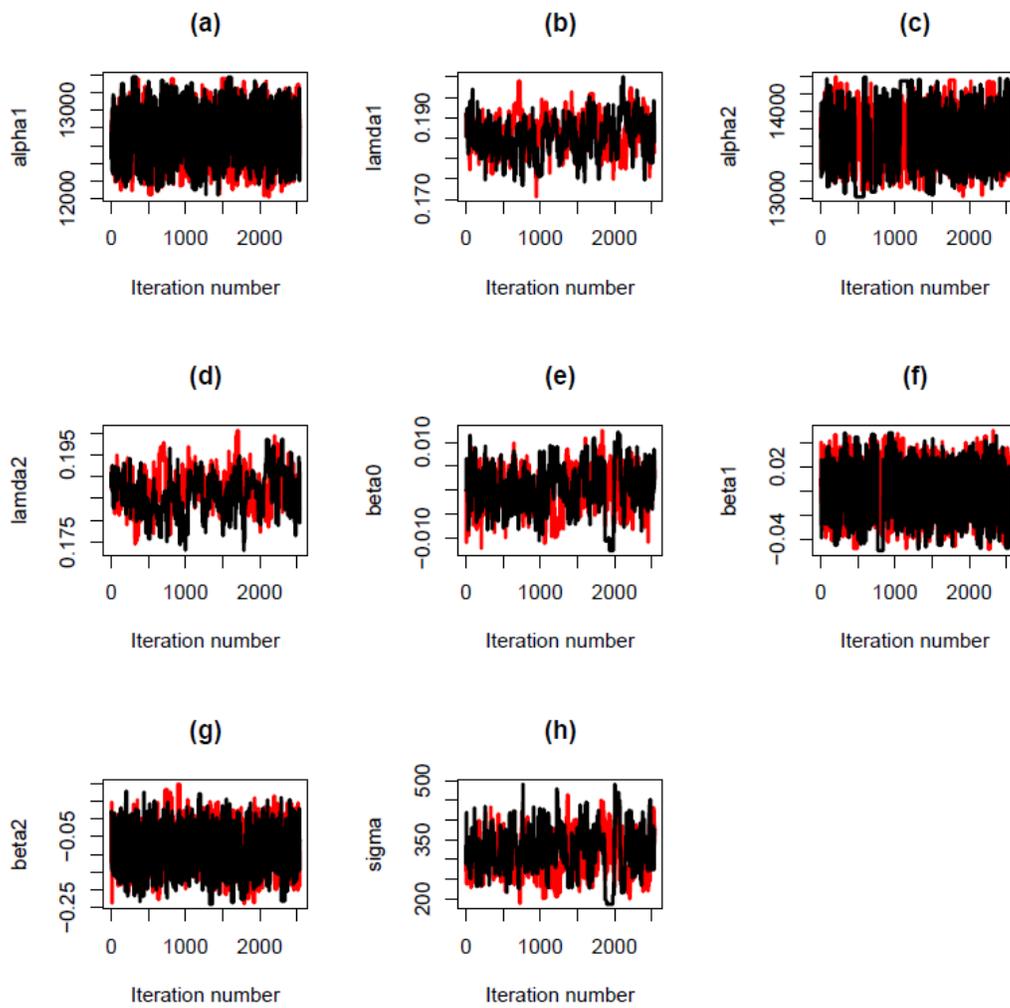


Figure 3: Coupling from the past plots for Model I.

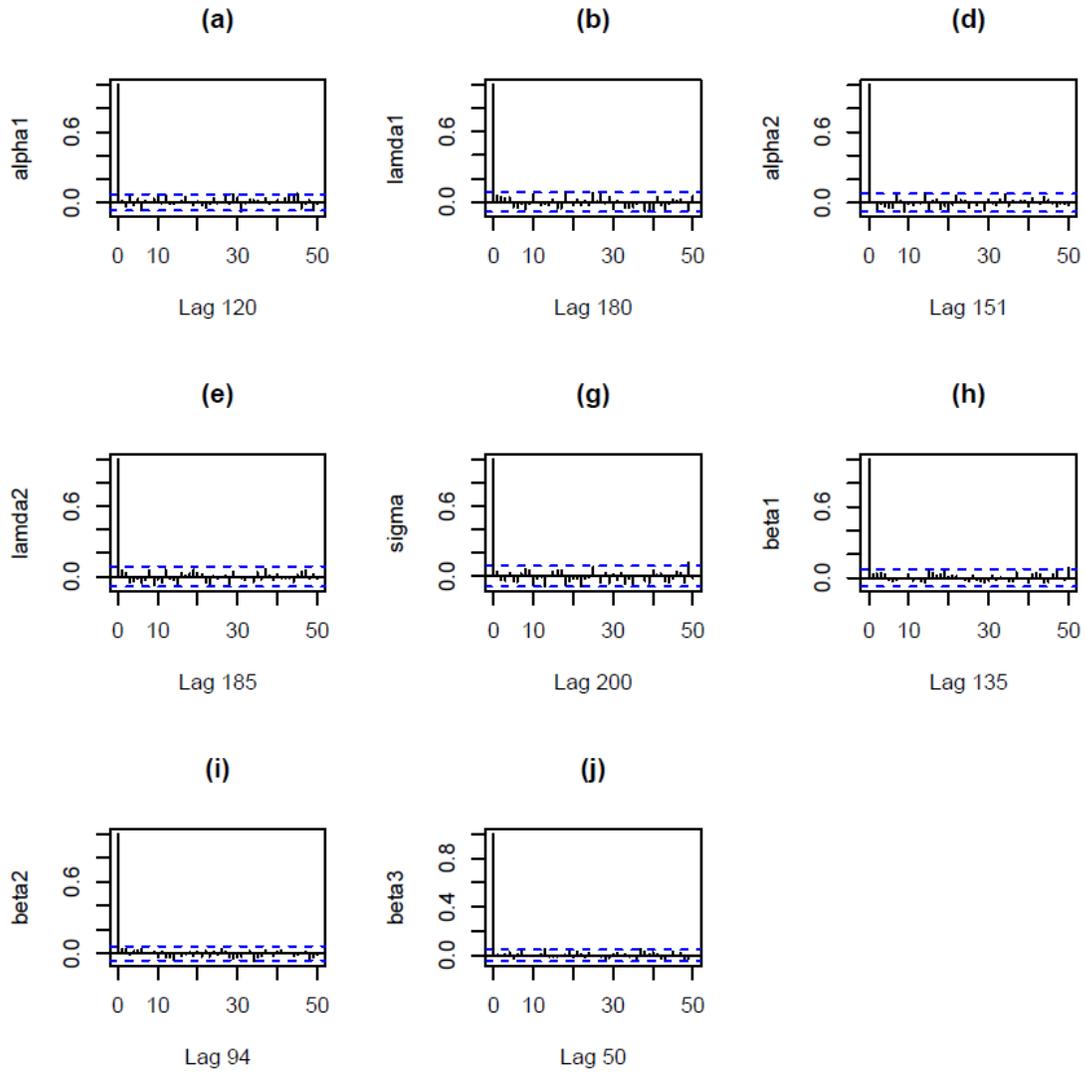


Figure 4: ACF after thinning plot for Model I

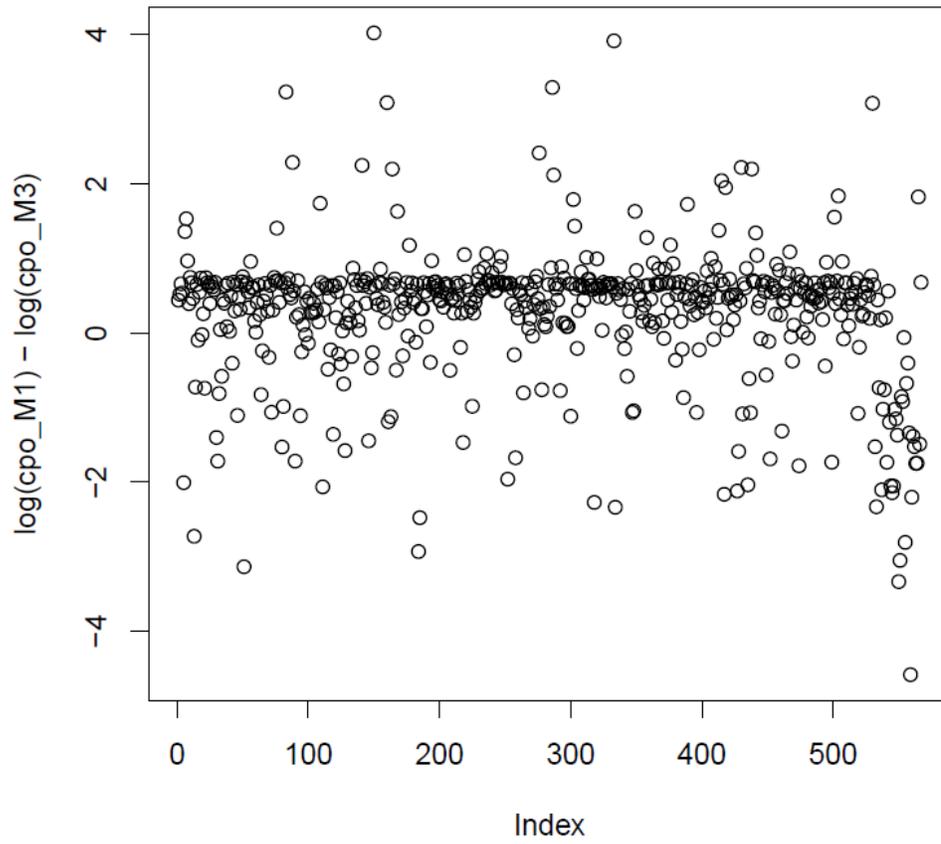


Figure 5: CPO plots for Model comparison.

