

# Research & Reviews: Journal of Statistics and Mathematical Sciences

## Bayesian Estimation in Shared Compound Negative Binomial Frailty Models

David D Hanagal<sup>1\*</sup> and Asmita T Kamble<sup>2</sup>

Department of Statistics, University of Pune, Pune 411007, India

### Research Article

Received date: 29/08/2015

Accepted date: 30/12/2015

Published date: 06/01/2016

#### \*For Correspondence

David D Hanagal, Department of Statistics, University of Pune, Pune-411007, India, Tel: (020) 25698752

**E-mail:** david\_hanagal@yahoo.co.in

**Keywords:** Bayesian estimation, Compound negative binomial frailty, Markov chain monte carlo (MCMC), Shared frailty.

#### ABSTRACT

Frailty models are used in survival analysis to model unobserved heterogeneity. To study such heterogeneity by the inclusion of a random term called the frailty is assumed to multiply hazards of all subjects in the shared frailty. We study compound negative binomial distribution as frailty distribution and two different baseline distributions namely Pareto and linear failure rate distribution in this paper. A simulation study is done to compare the true values of parameters with the estimated value. We develop the Bayesian estimation procedure using Markov Chain Monte Carlo (MCMC) technique to estimate the parameters of the proposed models. We try to fit the proposed models to a real life bivariate survival data set of McGrilchrist and Aisbett related to kidney infection. Also, we present a comparison study for the same data by using model selection criterion, and suggest a better model.

### INTRODUCTION

Frailty is a random component designed to account for heterogeneity caused by unobserved individual-level factors that are otherwise neglected by the other predictors in the model. Vaupel et al. <sup>[1]</sup> suggested frailty models to account for the variations due to unobserved covariates. Several distributions like gamma, inverse Gaussian, positive stable distribution, power variance function, Weibull, compound poisson are used as frailty models for heterogeneity in the populations.

A class of random effect models which proved useful in survival analysis of related individuals is a class of frailty models which are based on the modifying the hazard function of individuals by introducing multiplicative effect on the baseline hazard function. Thus the frailty model is a random effect model for time to event data which is an extension of the Cox's proportional hazards model. Vaupel et al. <sup>[1]</sup>, Keyfitz and Littman <sup>[2]</sup> showed that ignoring individual heterogeneity lead to incorrect conclusions.

Let  $T$  be a survival time with an absolutely continuous distribution. A non-negative random variable random variable  $Z$  is called 'frailty' if the conditional hazard functions given  $Z=z$  is given by

$$h(t|z) = zh_0(t); t > 0 \quad (1.1)$$

Where  $h_0(t)$  is called the baseline hazard function. Then the conditional survival function is given by,

$$S(t|z) = e^{-zH(t)}; t > 0 \quad (1.2)$$

Where  $H(t) = \int_0^t h_0(u)du$  is the cumulative baseline hazard.

And the marginal survival function

$$S(t) = E[S(t|z)] = E[e^{-zH(t)}] = L_Z(H(t)); t > 0 \quad (1.3)$$

Where  $L_Z(\cdot)$  is the Laplace transform of the frailty distribution.

In this paper, we consider the shared frailty model with the compound negative binomial distribution as a frailty distribution

and the Pareto and linear failure rate distribution as baseline distributions. The remainder of the paper is organized as follows. In Section 2, we give the properties of the general shared frailty models. In Section 3, we introduce the shared compound negative binomial frailty model. The baseline distributions are given in Section 4. In Section 5, we propose two different compound negative binomial frailty models. In Section 6, we discuss Bayesian method is used to estimate the parameters of the proposed models. We discuss the different model selection criterion by Bayesian approach in section 7. In Section 8 we present the simulation study. We present analysis of kidney infection data set and suggest a better model from these two proposed models in Section 9. Finally in Section 10, we discuss the conclusion of the study.

## GENERAL SHARED FRAILTY MODEL

Shared frailty models explain correlations within groups (family, litter or clinic) or for recurrent events facing the same individual. i.e., the different events within each community share a common frailty, shared by each individual within the community and each unit belongs to precisely one category. The shared gamma frailty model was suggested by Clayton <sup>[3]</sup> for the analysis of the correlation between clustered survival times in genetic epidemiology.

Bivariate survival data arise when each subject understudy experiences two events. For e.g., failure times of paired organs like kidneys, eyes, ears or any other paired organs of an individual, In industrial applications, the breakdown times of dual generators in a power plant or failure times of two engines in a two-engine airplane, recurrences of a given disease. Both monozygotic and dizygotic twins share date of birth and common pre-birth environment are the illustrations of bivariate survival data with the shared frailty <sup>[4]</sup>.

In order to build the shared frailty model for such kind of the data it is assumed that survival times are conditionally independent, for a given shared frailty. i.e., there is an association between survival times only due to frailty.

Let a bivariate random variable  $(T_{1j}, T_{2j})$  be the survival time of the  $i^{\text{th}}$  ( $i=1, 2$ ) component of the  $j^{\text{th}}$  individual ( $j=1,2,\dots,n$ ). Given the unobserved  $Z_j$  the hazard function for the  $(T_{1j}, T_{2j})$  is given by,

$$h_i(t_{ij}|z_j, \underline{X}_j) = z_j h_{0i}(t_{ij}) \exp(\underline{X}_j' \underline{\beta}), i=1,2 \quad (2.1)$$

Where  $z_j$  represents frailty acting as multiplicative effect at an individual level,  $h_{0i}(t_{ij})$  is the baseline hazard at time  $t_{ij} > 0$  and  $\underline{\beta}$  is the vector of regression coefficients with  $k$  components and  $\underline{X}_j$  is the vector of observed covariates having  $k$  components.

Integrating the hazard function  $h_i(t_{ij}|z_j, \underline{X}_j)$  we get, the conditional cumulative hazard function for the  $j^{\text{th}}$  individual at the  $i^{\text{th}}$  component survival time  $t_{ij} > 0$  for the given frailty  $Z_j = z_j$  is,

$$H_i(t_{ij}|z_j, \underline{X}_j) = z_j H_{0i}(t_{ij}) \eta_j \quad (2.2)$$

Where  $\eta_j = \exp(\underline{X}_j' \underline{\beta})$  and  $H_{0i}(t_{ij})$  is the cumulative baseline hazard function at time  $t_{ij} > 0$ . The conditional survival function for the  $j^{\text{th}}$  individual at the  $i^{\text{th}}$  component survival time  $t_{ij} > 0$  for the given frailty  $Z_j = z_j$  is,

$$S(t_{ij}|z_j, \underline{X}_j) = e^{-H_i(t_{ij}|z_j, \underline{X}_j)} = e^{-z_j H_{0i}(t_{ij}) \eta_j} \quad (2.3)$$

When  $T_{1j}$  and  $T_{2j}$  for  $j=1,2,\dots,n$  are independent, the bivariate conditional survival function of  $(T_{1j}, T_{2j})$  for the given frailty  $Z_j = z_j$  is the product to conditional survival function of  $T_{1j}$  and  $T_{2j}$  for the given frailty  $Z_j = z_j$ . Thus, we have

$$S(t_{1j}, t_{2j}|z_j, \underline{X}_j) = e^{-z_j (H_{01}(t_{1j}) + H_{02}(t_{2j})) \eta_j} \quad (2.4)$$

Where  $H_{01}(t_{1j})$  and  $H_{02}(t_{2j})$  be the cumulative baseline hazard functions of the first and the second component at  $t_{1j} > 0$  and  $0$  respectively.

Integrate out the bivariate conditional survival function of  $(T_{1j}, T_{2j})$  over the frailty variable  $Z_j$  having the probability function,  $f_z(Z_j)$  for the  $j^{\text{th}}$  individual in order to obtain the unconditional bivariate survival function at time  $T_{ij} > 0$ .

$$\begin{aligned} S(t_{1j}, t_{2j} | \underline{X}_j) &= \int_{Z_j} e^{-z_j (H_{01}(t_{1j}) + H_{02}(t_{2j})) \eta_j} f_z(z_j) dz_j \\ &= L_{Z_j} \left( (H_{01}(t_{1j}) + H_{02}(t_{2j})) \eta_j \right) \end{aligned} \quad (2.5)$$

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

## SHARED COMPOUND NEGATIVE BINOMIAL FRAILTY MODEL

A compound distribution is a model for a random sum  $Z = X_1 + X_2 + \dots + X_N$  where the number of terms  $N$  is uncertain. We assume

that the variables  $X_i$  are independent and identically distributed and that each  $X_i$  is independent of  $N$ . The random sum  $Z$  can be interpreted the sum of all the measurements that are associated with certain events that occur during a fixed period of time. For example, compound distribution is the random variable of the aggregate claims generated by an insurance policy or a group of insurance policies during a fixed policy period. In this setting,  $N$  is the number of claims generated by the portfolio of insurance policies and  $X_1$  is the amount of the first claim and  $X_2$  is the amount of the second claim and so on. The random sum  $Z$  is said to have a compound Poisson distribution if  $N$  follows the Poisson distribution.

The Compound Poisson variable  $Z$  is defined as

$$Z = \begin{cases} X_1 + X_2 + \dots + X_N & ; N > 0 \\ 0 & ; N = 0 \end{cases}$$

Where  $N$  is Poisson distributed with mean  $\rho$  while  $X_1, X_2, \dots, X_N$  are independent gamma distributed with scale parameter  $\nu$  and shape parameter  $\xi$ .

The distribution of  $Z$  can be partitioned into two parts, one a discrete part corresponding to the probability of zero susceptibility and second is due to continuous part on the positive real line.

The discrete part is given as

$$P(Z = 0) = \exp(-\rho)$$

Which is decreases as  $\rho$  decreases.

The distribution of the continuous part is given by conditioning  $Z$  and with  $X$  having gamma distribution.

Aalen <sup>[5,6]</sup> introduced compound Poisson distribution as a mixing distribution in survival models. In many situations hazard rates or intensities are raising at the start, reaching a maximum value and then declining, that's why the intensity has a unimodal shape with finite mode.

For e.g. death rates for cancer patients, divorce rates.

The reason to start decline in the population intensity is that the high risk individuals have already died or been divorced in case of above examples.

Also in is often seen that the total integral under the intensity or hazard rate is to be finite. It occurs due to the distribution is defective. It means that some individuals have zero susceptibility; they will survive forever.

In the case of above examples some patients survive their cancer, some people never marry, and some marriages are not prone to be dissolved. In such kind of data compound Poisson distribution plays an important role of mixing distribution.

Even though the compound Poisson distribution has many attractive properties, in non-susceptible or zero susceptibility type of data a convenient frailty distribution. Some other examples of zero susceptibility type of data if  $N > 0$  then we can interpret  $Z$  as aggregate heterogeneity due to failures before we get first success or in general  $r^{\text{th}}$  success can be given as, in case of marriage data,  $Z$  may represents heterogeneity due to difficulties in finding a marriageable partner before individual meet first suitable partner. In case of fertility,  $Z$  may be heterogeneity due to miss-carriages observed or unable to conceive a child with a couple before they have their first child or second child. Some mothers would like to deliver babies until she delivers a baby boy or two baby boys. Politicians go on contesting elections until they win once or twice and so on. Individuals go on changing the jobs until he/she gets a suitable job. In such situations negative binomial distribution or geometric distribution is a suitable choice of distribution for variate  $N$ .

So, we decide to consider compound negative binomial distribution as a frailty distribute-ion to model zero susceptibility type data in our study.

When the number of successes is equal to one, the compound negative binomial distribution reduces to compound geometric. There is a relation between compound negative binomial distribution and compound Poisson distribution. If  $Z_1$  and  $Z_2$  are respectively follows compound Poisson and compound negative binomial distribution with  $Z_1 = X_1 + X_2 + \dots + X_{N_1}$  and  $Z_2 = X_1 + X_2 + \dots + X_{N_2}$  where  $N_1$  is Poisson random variate with intensity  $\lambda$  and  $N_2$  is and negative binomial random variate with parameters, number of successes  $r$  and probability

of success  $p$  then  $Z_1$  and  $Z_2$  are identically distributed if  $\lambda = -r \log p$ .

The compound negative binomial distribution is defined as

$$Z = \begin{cases} X_1 + X_2 + \dots + X_N & ; N > 0 \\ 0 & ; N = 0 \end{cases}$$

Where  $N$  is negative binomial variate with parameters  $r$  and  $p$ ;  $r$  and  $p$  denotes respectively, the number of successes and

the probability of success while  $X_1, X_2, \dots, X_N$  are independent gamma distributed with scale parameter  $v$  and shape parameter  $\xi$ .

The distribution of  $Z$  consists of two parts; a discrete part which corresponds to the probability of zero susceptibility, and a continuous part on the positive real line.

The discrete part is,

$$P(Z = 0) = p^r$$

The distribution of the continuous part can be found by conditioning  $N$  and using the fact that the  $X$ 's are gamma distributed. It can be written as

$$f(Z) = p^r \frac{1}{Z} e^{-vZ} \sum_{n=0}^{\infty} \binom{n+r-1}{r-1} \frac{q^n (vZ)^{n\xi}}{\tilde{\Lambda}(n\xi)} \quad ; \quad \begin{array}{l} Z > 0, v > 0, \xi > 0, n = 0, 1, 2, \dots \\ r = 1, 2, \dots \\ 0 < p < 1 \end{array} \quad (3.1)$$

is the density function of compound negative binomial distribution.

The expectation and variance of  $Z$  are

$$E(Z) = rq\xi / pv \text{ and } var(Z) = \sigma^2 = rq\xi(\xi + p) / (p^2v^2)$$

The Laplace transformation of  $Z$  is given by,

$$\begin{aligned} L_Z(s) &= E(e^{-sZ}) = E(e^{-s(X_1 + X_2 + \dots + X_N)}) \\ &= E(L_X(s)^N) \\ &= L_N(-\ln(L_X(s))) \\ &= \left\{ \frac{p}{1 - q \left(1 + \frac{s}{v}\right)^{-\xi}} \right\}^r \end{aligned}$$

In order to solve the non-identifiability problem, we take  $E(z) = 1$  which leads to a  $var(Z) = \sigma^2 = (\xi + p) / (rq\xi)$ . Thus, we have the following form of Laplace transformation.

$$L_Z(s) = \left\{ \frac{p}{1 - q \left(1 + \frac{ps}{rq\xi}\right)^{-\xi}} \right\}^r \quad (3.2)$$

The unconditional bivariate survival function for the  $j$ th individual at the time  $t_{1j} > 0$  and  $t_{2j} > 0$  by replacing the above mentioned Laplace transformation in equation (2.5) we have,

$$S(t_{1j}, t_{2j}) = \left\{ \frac{p}{1 - q \left(1 + \frac{p(H_{01}(t_{1j}) + H_{02}(t_{2j}))\eta_j}{rq\xi}\right)^{-\xi}} \right\}^r \quad (3.3)$$

Where  $H_{01}(t_{1j})$  and  $H_{02}(t_{2j})$  are the cumulative baseline hazard functions of lifetime random variables  $T_{1j}$  and  $T_{2j}$  respectively.

In this paper, we consider the two baseline distributions namely Pareto and linear failure rate distribution which yield two compound negative binomial frailty models.

## BASELINE DISTRIBUTIONS

The Pareto distribution is a skewed, heavy-tailed distribution that is sometimes used to model the distribution of incomes. This distribution is not limited to describing wealth or income, but to many situations in which an equilibrium is found in the distribution of the "small" to the "large".

The first baseline distribution is the Pareto distribution [7]. A continuous random variable  $T$  is said to follow the Pareto distribution with the scale parameter  $\lambda$  and the shape parameter  $\alpha$  if its survival function is,

$$S_0(t) = \begin{cases} (\lambda t + 1)^{-\alpha} & ; t > 0, \lambda > 0, \alpha > 0 \\ 0 & ; \text{otherwise} \end{cases} \quad (4.1)$$

and the hazard function and the cumulative hazard function as

$$h_0(t) = \begin{cases} \frac{\alpha \lambda}{(1 + \lambda t)} & ; t > 0, \lambda > 0, \alpha > 0 \\ 0 & ; \text{otherwise} \end{cases} \quad (4.2)$$

$$H_0(t) = \begin{cases} \alpha \log(\lambda t + 1) & ; t > 0, \lambda > 0, \alpha > 0 \\ 0 & ; \text{otherwise} \end{cases} \quad (4.3)$$

Observe that  $h_0(t)$  decreases with  $t$ ;  $\lambda > 0, \alpha > 0$ . Hence this distribution belongs to the Decreasing Failure Rate class.

The exponential and Rayleigh are the two most commonly used distributions for analyzing lifetime data. These distributions have several desirable properties and nice physical interpretations. Unfortunately the exponential distribution only has constant failure rate and the Rayleigh distribution has increasing failure rate. The linear failure rate distribution generalizes both these distributions. We consider this is the second baseline distribution.

The linear failure rate distribution of a continuous random variable  $T$  with the parameters  $\alpha > 0$  and  $\lambda > 0$ , will be denoted by LFRD  $(\alpha, \lambda)$  has the following survival function

$$S_0(t) = \begin{cases} \exp\left(-\alpha t - \frac{\lambda}{2} t^2\right) & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; \text{otherwise} \end{cases} \quad (4.4)$$

It is easily observed that the exponential distribution (ED  $(\alpha)$ ) and the Rayleigh distribution (RD  $(\lambda)$ ) can be obtained from LFRD  $(a, b)$  by putting  $\lambda = 0$  and  $\alpha = 0$  respectively. Moreover, the probability density function (PDF) of the LFRD  $(\alpha, \lambda)$  can be decreasing or unimodal but the failure rate function is either constant or increasing only. See for example Bain <sup>[8]</sup>, Sen and Bhattacharya <sup>[9]</sup>, Lin et al. <sup>[10]</sup>, Ghitany and Kotz <sup>[11]</sup> and the references cited therein.

The hazard function and the cumulative hazard function of linear failure rate distribution are respectively as stated below:

$$h_0(t) = \begin{cases} \alpha + \lambda t & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; \text{otherwise} \end{cases} \quad (4.5)$$

$$H_0(t) = \begin{cases} \alpha t + \lambda \frac{t^2}{2} & ; t > 0, \alpha > 0, \lambda > 0 \\ 0 & ; \text{otherwise} \end{cases} \quad (4.6)$$

## PROPOSED MODELS

Here we present the two compound negative binomial frailty models say Model I and Model II by putting respectively the cumulative hazard function of the baseline distributions namely Pareto and linear failure rate distribution in the unconditional survival function of bivariate random variable  $(t_{1j}, t_{2j})$  given in equations (4.3) and (4.6).

$$S(t_{1j}, t_{2j}) = \left\{ \frac{p}{1 - q \left( 1 + \frac{p(\alpha_1 \log(\lambda_1 t_{1j} + 1) + \alpha_2 \log(\lambda_2 t_{2j} + 1)) \eta_j}{r q \xi} \right)^{-\xi}} \right\}^r ; t_{1j} > 0, t_{2j} > 0 \quad (5.1)$$

$$S(t_{1j}, t_{2j}) = \left\{ \frac{p}{1 - q \left( 1 + \frac{p \left( \left( \alpha_1 t_{1j} + \lambda_1 \frac{t_{1j}^2}{2} \right) + \left( \alpha_2 t_{2j} + \lambda_2 \frac{t_{2j}^2}{2} \right) \right) \eta_j}{r q \xi} \right)^{-\xi}} \right\}^r ; t_{1j} > 0, t_{2j} > 0 \quad (5.2)$$

Here onwards, equations (5.1) and (5.2) as Model I and Model II which correspond to compound negative binomial frailty models with baseline Pareto and linear failure rate distributions respectively.

## ESTIMATION OF PARAMETERS BY BAYESIAN APPROACH

For the bivariate life time distribution, we use the bivariate censoring scheme given by Hanagal and Dabade <sup>[12]</sup>.

Suppose that there are  $n$  independent pairs of components under study and the  $j^{\text{th}}$  pair of the component have lifetimes  $(t_{1j}, t_{2j})$  i.e. there are  $n$  individuals with a pair of components having lifetimes  $(t_{1j}, t_{2j})$  for  $j=1, 2, \dots, n$ . The life time times associated with  $j^{\text{th}}$  individual is given by,

$$(T_{1j}, T_{2j}) = \begin{cases} (t_{1j}, t_{2j}) & ; t_{1j} < c_{1j}, t_{2j} < c_{2j} \\ (t_{1j}, c_{2j}) & ; t_{1j} < c_{1j}, t_{2j} < c_{2j} \\ (c_{1j}, t_{2j}) & ; t_{1j} > c_{1j}, t_{2j} < c_{2j} \\ (c_{1j}, c_{2j}) & ; t_{1j} > c_{1j}, t_{2j} > c_{2j} \end{cases}$$

Where  $c_{1j}$  and  $c_{2j}$  be the observed censoring times for  $j^{\text{th}}$  individual ( $j=1,2,\dots,n$ ) with a pair of components respectively. We assume that the lifetimes and censoring times are independently distributed.

Now the likelihood of the sample of size  $n$  is given by,

$$(\underline{\theta}, \underline{\beta}, \sigma^2) = \left( \prod_{j=1}^{n_1} f_1(t_{1j}, t_{2j}) \right) \left( \prod_{j=1}^{n_2} f_2(t_{1j}, c_{2j}) \right) \left( \prod_{j=1}^{n_3} f_3(c_{1j}, t_{2j}) \right) \left( \prod_{j=1}^{n_4} f_4(c_{1j}, c_{2j}) \right) \quad (6.1)$$

Where  $\underline{\theta}, \underline{\beta}, \sigma^2$  are the vector of parameters of the baseline distributions, the vector of regression coefficients and the frailty parameter respectively. Let the counts  $n_1, n_2, n_3$  and  $n_4$  be the number of individuals for which the first and the second components 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

$$\begin{aligned} f_1(t_{1j}, t_{2j}) &= \frac{\partial^2 S(t_{1j}, t_{2j})}{\partial t_{1j} \partial t_{2j}} \\ &= \frac{p^2 \eta_j^2 h_{01}(t_{1j}) h_{02}(t_{2j}) S(t_{1j}, t_{2j}) q (-1 + r\xi) + ((\xi + 1)(\phi(t_{1j}, t_{2j}))^\xi)}{r q \xi \left[ (\phi(t_{1j}, t_{2j}))^2 (q - (\phi(t_{1j}, t_{2j}))^\xi) \right]^2} \\ f_2(t_{1j}, c_{2j}) &= -\frac{\partial S(t_{1j}, c_{2j})}{\partial t_{1j}} = \frac{p h_{01}(t_{1j}) S(t_{1j}, c_{2j}) \eta_j}{\left[ \phi(t_{1j}, c_{2j}) \right]^{1+\xi} \left[ 1 - q(\phi(t_{1j}, c_{2j}))^{-\xi} \right]} \\ f_3(c_{1j}, t_{2j}) &= -\frac{\partial S(c_{1j}, t_{2j})}{\partial t_{2j}} = \frac{p h_{02}(t_{2j}) S(c_{1j}, t_{2j}) \eta_j}{\left[ \phi(c_{1j}, t_{2j}) \right]^{1+\xi} \left[ 1 - q(\phi(c_{1j}, t_{2j}))^{-\xi} \right]} \\ f_4(c_{1j}, c_{2j}) &= S(c_{1j}, c_{2j}) \end{aligned} \quad (6.2)$$

Where

$$\phi(t_{1j}, t_{2j}) = 1 + \frac{p(H_{01}(t_{1j}) + H_{02}(t_{2j})) \eta_j}{r q \xi}$$

Thus, we get the two likelihood functions for the two proposed compound negative binomial frailty models namely Model I, Model II by substituting the corresponding hazard functions and cumulative hazard functions in the likelihood function given by equation (6.1) with  $S(t_{1j}, t_{2j})$  stated in equation (5.1) and (5.2).

The likelihood function is obtained for the

Model I with

$$f_1(t_{1j}, t_{2j}) = \frac{p^2 \eta_j^2 \frac{\alpha_1 \lambda_1}{(1 + \lambda_1 t_{1j})} \frac{\alpha_2 \lambda_2}{(1 + \lambda_2 t_{2j})} S(t_{1j}, t_{2j}) q (-1 + r\xi) + ((\xi + 1)(\phi(t_{1j}, t_{2j}))^\xi)}{r q \xi \left[ (\phi(t_{1j}, t_{2j}))^2 (q - (\phi(t_{1j}, t_{2j}))^\xi) \right]^2}$$

$$f_2(t_{1j}, c_{2j}) = \frac{p \frac{\alpha_1 \lambda_1}{(1 + \lambda_1 t_{1j})} S(t_{1j}, c_{2j}) \eta_j}{[\phi(t_{1j}, c_{2j})]^{1+\xi} [1 - q(\phi(t_{1j}, c_{2j}))^{-\xi}]}$$

$$f_3(c_{1j}, t_{2j}) = \frac{p \frac{\alpha_2 \lambda_2}{(1 + \lambda_2 t_{2j})} S(c_{1j}, t_{2j}) \eta_j}{[\phi(c_{1j}, t_{2j})]^{1+\xi} [1 - q(\phi(c_{1j}, t_{2j}))^{-\xi}]}$$

$$f_4(c_{1j}, c_{2j}) = S(c_{1j}, c_{2j})$$

Where

$$\phi(t_{1j}, t_{2j}) = 1 + \frac{p(\alpha_1 \log(\lambda_1 t_{1j} + 1) + \alpha_2 \log(\lambda_2 t_{2j} + 1)) \eta_j}{r q \xi} \tag{6.3}$$

Model II with

$$f_1(t_{1j}, t_{2j}) = \frac{p^2 \eta_j^2 (\alpha_1 + \lambda_1 t_{1j})(\alpha_2 + \lambda_2 t_{2j}) S(t_{1j}, t_{2j}) q(-1 + r \xi) + ((\xi + 1)(\phi(t_{1j}, t_{2j}))^\xi)}{r q \xi \left[ (\phi(t_{1j}, t_{2j}))^2 \left( q - (\phi(t_{1j}, t_{2j}))^\xi \right) \right]^2}$$

$$f_2(t_{1j}, c_{2j}) = \frac{p(\alpha_1 + \lambda_1 t_{1j}) S(t_{1j}, c_{2j}) \eta_j}{[\phi(t_{1j}, c_{2j})]^{1+\xi} [1 - q(\phi(t_{1j}, c_{2j}))^{-\xi}]}$$

$$f_3(c_{1j}, t_{2j}) = \frac{p(\alpha_2 + \lambda_2 t_{2j}) S(c_{1j}, t_{2j}) \eta_j}{[\phi(c_{1j}, t_{2j})]^{1+\xi} [1 - q(\phi(c_{1j}, t_{2j}))^{-\xi}]}$$

$$f_4(c_{1j}, c_{2j}) = S(c_{1j}, c_{2j})$$

Where

$$\phi(t_{1j}, t_{2j}) = 1 + \frac{p \left( \left( \alpha_1 t_{1j} + \lambda_1 \frac{t_{1j}^2}{2} \right) + \left( \alpha_2 t_{2j} + \lambda_2 \frac{t_{2j}^2}{2} \right) \right) \eta_j}{r q \xi} \tag{6.4}$$

Each of the proposed models consists of eleven parameters, computing the maximum likelihood estimators (MLEs) involves solving a eleven dimensional optimization problem for these model. As the method of maximum likelihood fails to estimate the several parameters due to convergence problem in the iterative procedure, so 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<sup>[13]</sup>. 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 Bayesian approach for the estimation of parameters of the frailty models. Some of them are, Ibrahim et al.<sup>[14]</sup> and Santos and Achcar<sup>[15]</sup>. Santos and Achcar<sup>[15]</sup> considered the parametric models with Weibull and the generalized gamma distribution as the baseline distributions and gamma and log-normal as frailty distributions. Ibrahim et al.<sup>[14]</sup> and references therein considered Weibull model and the piecewise exponential model with gamma frailty. Therefore we proposed Bayesian inferential approach in this study to estimate the parameters of the model, which is a popularly used method, because computation of the Bayesian analysis becomes feasible due to advances in computing technology.

To apply markov chain monte carlo (MCMC) methods, we assume that, conditional on observed covariates and on the entire set of parameters, observations are independent and prior distributions for all parameters are mutually independent. We used the Metropolis-Hastings algorithm within Gibbs sampler technique which is the most basic MCMC method used in Bayesian Inference. Convergence of Markov chain to a stationary distribution is observed by the trace plots, the coupling from the past plots, the Gelman-Rubin convergence statistic, and the Geweke test. The trace plots are used to check the behavior of the chain and the coupling from the past plots can be used to decide the burn-in period. The Gelman-Rubin convergence statistic values are approximately equal to one then sample can be considered to be come from the stationary distribution. The Geweke test examine the convergence of a Markov chain based on the sub parts of a chain at the end and at the beginning of the convergence period.

The large standardized difference between ergodic averages at the beginning and at the end of the convergence period indicates non convergence. The sample autocorrelation plots can be used to decide the autocorrelation lag.

In Bayesian paradigm the parameters of the model are viewed as random variables with some distribution known as prior distribution. It enables us to combine both the prior information and the data at hand to update the information of parameter. Thus, posterior density of a parameter is the distribution of a parameter updated by combining its prior distribution and the likelihood function. We assume that, conditional on explanatory variables and on the entire set of parameters, observations are independent and prior distributions for all parameters are mutually independent while applying MCMC methods.

Let  $L(\theta|y)$  be the likelihood function and  $p(\theta)$  be the prior density of a parameter then posterior density function of a parameter  $\pi(\theta|y)$  is given by,

$$\pi(\theta|y) \propto L(\theta|y)p(\theta) \tag{6.5}$$

In our case the joint posterior density function of a parameter for given failure times  $(t_{1j}, t_{2j})$  is  $\pi(\underline{\theta} = (\alpha_1, \lambda_1, \alpha_2, \lambda_2, \sigma^2, \underline{\beta}|(t_{1j}, t_{2j})))$  as

$$\begin{aligned} &\pi(\underline{\theta} = (\alpha_1, \lambda_1, \alpha_2, \lambda_2, \sigma^2, \underline{\beta}|(t_{1j}, t_{2j}))) \\ &\propto L(\underline{\theta} = (\alpha_1, \lambda_1, \alpha_2, \lambda_2, \sigma^2, \underline{\beta}|(t_{1j}, t_{2j}))) \left( g_1(\alpha_1) g_2(\lambda_1) g_3(\alpha_2) g_4(\lambda_2) g_5(\sigma^2) \right) \prod_{i=1}^k p_i(\beta_i) \end{aligned} \tag{6.6}$$

Where  $g_i(\cdot)$  for  $i = 1, 2, \dots, 5$  denotes the prior density function with known hyper parameters of corresponding argument for baseline parameters and frailty variance;  $p_i(\beta_i)$  is the prior density function for regression coefficients  $\beta_i$  for  $i = 1, 2, \dots, k$ . And the likelihood function  $L(\cdot)$  is given by equation (6.1).

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. Since the full conditional distributions are not easy to integrate out therefore full conditional distributions are obtained by considering that they are proportional to the joint distribution of the parameters of the model. The samples are then obtained from these full conditional distributions.

## MODEL SELECTION CRITERION BY BAYESIAN APPROACH

Bayesian model comparison is commonly performed by computing posterior model probabilities. In order to compare proposed models we use the akaike information criterion (AIC), the Bayesian information criterion (BIC), the deviance information criterion (DIC) and the Bayes factor. These are the most common methods of Bayesian model assessment.

Akaike <sup>[16]</sup> suggested that, given a class of competing models for a data set, one choose the model that minimizes

$$AIC = D(\hat{\underline{\theta}}) + 2p \tag{7.1}$$

where  $p$  represents number of parameters of the model.  $D(\hat{\underline{\theta}})$  represents an estimate of the deviance evaluated at the posterior mean,  $\hat{\underline{\theta}} = E(\underline{\theta}|data)$ . The deviance is given by,  $D(\underline{\theta}) = -2\log L(\underline{\theta})$ , Where  $\underline{\theta}$  is a vector of unknown parameters of the model and  $L(\underline{\theta})$  is the likelihood function of the model.

Bayesian information criterion (BIC) was suggested by Schwarz <sup>[17]</sup>. Shibata <sup>[18]</sup> and Katz <sup>[19]</sup> have shown that the AIC tends to overestimate the number of parameters needed, even asymptotically. The Schwarz criterion indicates that the model with the highest posterior probability is the one that minimizes

$$BIC = D(\hat{\underline{\theta}}) + p \log(n) \tag{7.2}$$

Where  $n$  is the number of observations, or equivalently, the sample size.

DIC, a generalization of AIC, is introduced by Spiegelhalter et al. <sup>[20]</sup> and is defined as;

$$DIC = D(\hat{\underline{\theta}}) + 2p_D \tag{7.3}$$

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

Models with smaller values of the AIC, BIC and DIC are preferred.

Kadane and Lazar <sup>[21]</sup> review model selection from Bayesian and frequent perspectives. The Bayes factor  $B_{jk}$  for a model  $M_j$  against  $M_k$  or given data  $D = (t_{1j}, t_{2j}); j = 1, 2, \dots, n$  is

$$\text{Where } P(D|M_k) = \int_S P(D|M_k, \theta_k) \pi(\theta_k | M_k) d\theta_k \tag{7.4}$$

$\theta_k$  is the parameter vector under model  $M_k$  and  $\pi(\theta_k | M_k)$  is prior density and  $S$  is the support of the parameter  $\theta_k$



. Raftery [22], following Jeffreys [23], 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  $I_k = P(D | M_k)$ . By considering one of the approaches given in Kass and Raftery [24], we obtain the following MCMC estimate of  $I_k$  which is given by,

$$I_k = \left\{ \frac{\sum_{i=1}^N P(D | \theta^{(i)})^{-1}}{N} \right\}^{-1} \quad (7.5)$$

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.

## SIMULATION STUDY

A simulation study is done to evaluate the performance of the Bayesian estimation procedure. For the simulation purpose we have considered only one covariate  $X_1$  and we assume that it follows normal distribution. The frailty variable  $Z$  is assumed to have compound negative binomial distribution. Life times  $(T_{1j}, T_{2j})$  for the  $j^{\text{th}}$  individual are conditionally independent for the given frailty  $Z_j = z_j$ . We considered that  $T_{ij}$  ( $i=1,2; j=1,2,\dots,n$ ) follows one of the baseline distributions, namely, Pareto (Model-I) or linear failure rate (Model-II) distribution respectively. As the Bayesian methods are time consuming, we generate only twenty, forty and sixty pairs of lifetimes using inverse transform technique. Here we have generated different random samples of size  $n=20, 40$  and  $60$  for lifetimes  $T_{1j}$  and  $T_{2j}$ . But here we are giving procedure for sample generation of only one sample size, say,  $n=20$ . Samples are generated using the following procedure:

1. Generate a random sample of size 20 from the compound negative binomial distribution with shape  $\xi$  and the scale parameter  $v$  as shared frailties ( $z_j$ ) for  $j^{\text{th}}$  ( $j=1,2,\dots,20$ ) individual.

To generate random observation from compound negative binomial distribution, we firstly generate a random observation  $N=n$  from negative binomial distribution with parameters

$$r=1, p=0.2.$$

Then we consider the following two cases:

(i) If  $N=0$ ; take frailty  $Z=0$ .

(ii) If  $N>0$ ; generate  $N$  gamma variates say  $X_1, X_2, \dots, X_N$  with the shape parameter  $\xi=0.2$  and scale parameter  $\nu = \frac{(rq\xi)}{p} = 0.8$  then frailty is taken as  $Z=X_1+X_2+\dots+X_N$ .

2. Generate 20 covariate values for  $X_1$  from the normal distribution.

3. Compute  $\eta = e^{(-X_1\beta)}$  with the regression coefficient  $\beta=0.5$ .

4. Generate 20 pairs of lifetimes  $(t_{1j}, t_{2j})$  for the given frailty ( $z_j$ ) using the following generators,

for Model-I and Model-II respectively,

$$t_{1j} = \frac{1}{\lambda_1} \left( \exp\left(\frac{A_{1j}}{\alpha_1}\right) - 1 \right), t_{2j} = \frac{1}{\lambda_2} \left( \exp\left(\frac{A_{2j}}{\alpha_2}\right) - 1 \right) \quad (8.1)$$

$$t_{2j} = \frac{1}{\lambda_1} \left( -\alpha_1 - \sqrt{\alpha_1^2 - 4\left(\frac{\lambda_1}{2}\right)A_{1j}} \right) \quad (8.2)$$

Where  $A_{1j} = \frac{-\log r_1}{z_j \eta_j}$  and  $A_{2j} = \frac{-\log r_2}{z_j \eta_j}$  and  $r_{ij}$   $i=1,2$  are random sample from  $U(0,1)$ .

5. Generate the censoring times  $(c_{1j}$  and  $c_{2j})$  from the exponential distribution.

6. Observe the  $i^{\text{th}}$  survival time  $t_{ij} = \min(t_{ij}, c_{ij})$  and the censoring indicator  $\delta_{ij}$  for the  $j^{\text{th}}$  individual ( $i= 1;2$  and  $j=1,2,\dots,20$ ), Where

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

Thus we have data consists of 20 pairs of the survival times  $(t_{1j}; t_{2j})$  and the censoring indicators  $\delta_{ij}$ .

We run two parallel chains for the proposed model with the different starting points using Metropolis-Hastings algorithm within Gibbs sampler based on normal transition kernels.

We iterate both the chains for 95,000 times. In our study we use non-informative prior for the frailty parameter  $\sigma_2$  and the regression coefficient  $\beta_1$ . Since we do not have any prior information about baseline parameters,  $\alpha_1, \lambda_1, \alpha_2$ , and  $\lambda_2$  prior distributions are assumed to be flat. A widely used prior for frailty parameter  $\sigma^2$  is the gamma distribution with mean one and large variance,  $G(\phi, \phi)$  say with a small choice of  $\phi$  and for the regression coefficients  $\beta_1, \dots, \beta_k$ , we use the normal prior with mean zero and large variance  $\varepsilon^2$ . Similar types of prior distributions are used in Ibrahim et al. [14], Sahu et al. [25] and Santos and Achcar [15]. We set hyper-parameters  $\phi=0.0001$ ,  $\varepsilon^2=1000$ . We consider the non-informative prior distribution for baseline parameters as the Gamma (1,0.0001).

For both the chains the results were somewhat similar so we present here the analysis for only one chain (i.e. chain 1) for the resulting model. Also due to lack of space we are not providing graphs. Simulated values of the parameters have the autocorrelation of lag k, so every  $k^{\text{th}}$  iteration is selected as a sample from posterior distribution. The posterior mean and standard error with credible intervals for different sample sizes are reported in **Tables 1 and 2** for Model-I, Model-II respectively. From these Tables, it can be observed that the estimated values of the parameters are close to the true values of the parameters and the standard errors decrease as the sample size increases.

**Table 1.** Posterior summary for simulation study of Model-I.

Parameter	$\alpha_1$	$\lambda_1$	$\alpha_2$	$\lambda_2$	$\rho$	$\xi$	$\beta_1$
True values	5.5	0.5	5.5	0.5	0.2	0.2	0.5
<b>Sample size=20</b>							
<b>Burn in period =2160</b>				<b>Autocorrelation lag=180</b>			
estimates	5.495265	0.4825503	5.519549	0.5052762	0.192014	0.2092385	0.4716779
standard error	0.1779357	0.05413933	0.170599	0.05805617	0.02155833	0.0546049	0.170429
lower credible limit	5.212513	0.4034647	5.213434	0.4063945	0.1613895	0.1095084	0.2124834
upper credible limit	5.78779	0.591469	5.786101	0.5966154	0.2363334	0.2962317	0.7802635
<b>Sample size=40</b>							
<b>Burn in period=3180</b>				<b>Autocorrelation lag=175</b>			
estimates	5.512785	0.5162308	5.481029	0.4817246	0.1780525	0.2431996	0.5472664
standard error	0.1689113	0.05611794	0.1722812	0.0514482	0.01668905	0.04252427	0.1600458
lower credible limit	5.214905	0.4081556	5.208617	0.402791	0.1603109	0.1479856	0.2352484
upper credible limit	5.79289	0.5972417	5.776535	0.5842291	0.2227113	0.2983601	0.7861975
<b>Sample size=60</b>							
<b>Burn in period = 3110</b>				<b>Autocorrelation lag=800</b>			
estimates	5.476585	0.4719628	5.529323	0.5212741	0.1748693	0.2539005	0.4671208
standard error	0.1789366	0.05523213	0.1701266	0.05339679	0.01332308	0.03512942	0.1510194
lower credible limit	5.215819	0.4018427	5.226985	0.4110774	0.1603077	0.1728933	0.2225904
upper credible limit	5.78585	0.5900244	5.788103	0.5962147	0.2118279	0.2982968	0.7729958

**Table 2.** Posterior summary for simulation study of Model-II.

Parameter	$\alpha_1$	$\lambda_1$	$\alpha_2$	$\lambda_2$	$\rho$	$\xi$	$\beta_1$
True values	5.5	0.5	5.5	0.5	0.2	0.2	0.5
<b>Sample size=20</b>							
<b>Burn in period=2590</b>				<b>Autocorrelation lag=175</b>			
estimates	5.482741	0.4997339	5.509842	0.4993959	0.1854986	0.2204464	0.4549848
standard error	0.1739062	0.05775292	0.1697222	0.05738042	0.0201024	0.05059383	0.1586061
lower credible limit	5.215199	0.4055352	5.221458	0.4037915	0.1609234	0.1176498	0.2115565
upper credible limit	5.783958	0.5951022	5.783743	0.5946693	0.2335115	0.2962933	0.7648207
<b>Sample size=40</b>							
<b>Burn in period=1950</b>				<b>Autocorrelation lag=800</b>			
estimates	5.518256	0.4961739	5.497526	0.4986062	0.1747031	0.246543	0.4915105
standard error	0.1584283	0.06212257	0.167403	0.05550285	0.01471476	0.0415707	0.1597369

lower credible limit	5.210711	0.4019986	5.207941	0.4086488	0.1602437	0.1530928	0.2092465
upper credible limit	5.783837	0.5967012	5.789753	0.5939387	0.219531	0.297743	0.7606974
<b>Sample size=60</b>							
<b>Burn in period=1940</b>				<b>Autocorrelation lag=900</b>			
estimates	5.446878	0.4899263	5.525997	0.5001326	0.1729319	0.2578155	0.4514856
standard error	0.1722099	0.05835875	0.1735858	0.05696716	0.01154164	0.0157387	0.1518377
lower credible limit	5.212004	0.4024134	5.214769	0.4045295	0.1603077	0.1909802	0.2208491
upper credible limit	5.771953	0.5971016	5.7874	0.5923991	0.2035887	0.2983601	0.7444531

## ANALYSIS OF KIDNEY INFECTION DATA

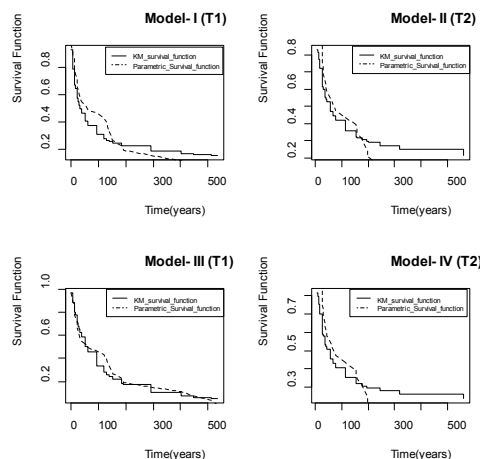
To study the Bayesian estimation procedure we use kidney infection data of McGilchrist and Aisbett [4]. The data is regarding recurrence times to infection at point of insertion of the catheter for 38 kidney patients using portable dialysis equipment. For each patient, first and second recurrence times (in days) of infection from the time of insertion of the catheter until it has to be removed owing to infection is recorded. The catheter may have to be removed for reasons other than kidney infection and this regard as censoring. So survival time for a patient given may be first or second infection time or censoring time. After the occurrence or censoring of the first infection sufficient (ten weeks interval) time was allowed for the infection to be cured before the second time the catheter was inserted. So the first and second recurrence times are taken to be independent apart from the common frailty component. The data consists of three risk variables age, sex and disease type GN, AN and PKD where GN, AN and PKD are short forms of glomerulo neptiritis, acute neptiritis and polycyatic kidney disease.

Let  $T_1$  and  $T_2$  be represents first and second recurrence time to infection. Five covariates age, sex and presence or absence of disease type GN, AN and PKD are denoted by  $X_1, X_2, X_3, X_4,$  and  $X_5$ . To analyze kidney infection data success is defined as getting infection first time so we define  $r=1$ .

First we check goodness of fit of the  $T_1$  and  $T_2$ . If marginal distributions of  $T_1$  and  $T_2$  for two proposed distributions fit well then the bivariate distribution of  $T_1$  and  $T_2$  may be fit well for the same. We used Kolmogorov-Smirnov Goodness-of-Fit Test. Thus from p-values of K-S test we can say that there is no statistical evidence to reject the hypothesis that data are from proposed models in the univariate case and we assume that the models also fit for the bivariate case. **Table 3** gives the p-values of Kolmogorov-Smirnov test for the proposed models. **Figure 1** shows the parametric versus non-parametric plots.

**Table 3.** Goodness-of-Fit Test: p-values K-S statistic for kidney infection data.

Model	K-S statistic		p-value	
	T1	T2	T1	T2
Model I	0.154412	0.2164123	0.44909	0.18199
Model II	0.1513085	0.2357491	0.58947	0.11738



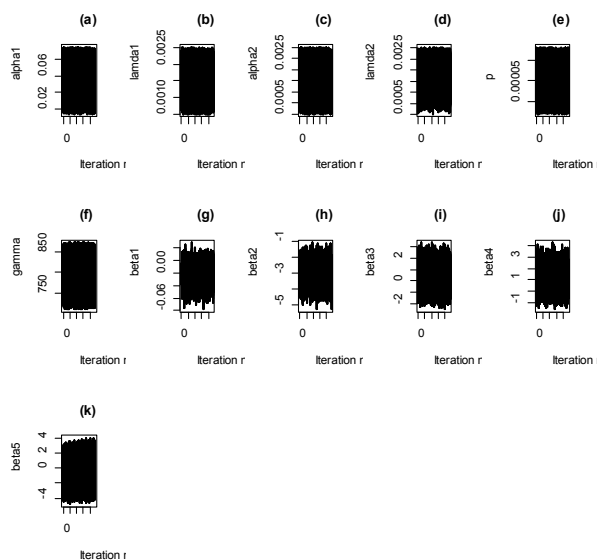
**Figure 1.** Survival function plots for (K-M survival and parametric survival).

To analyze kidney data set, various models have been applied by different researchers.

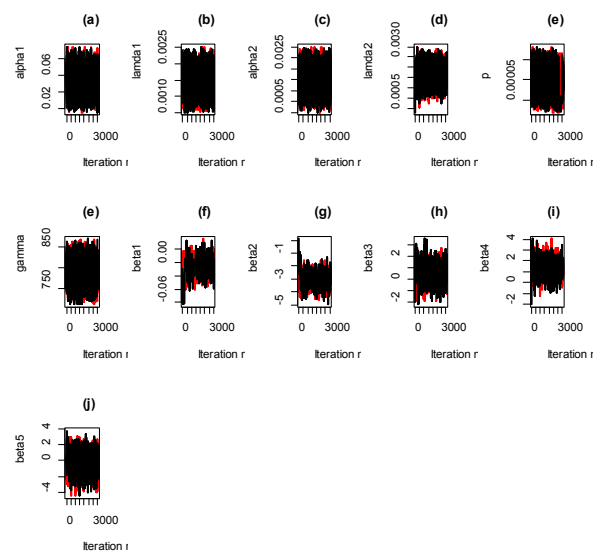
Some of them are, McGilchrist and Aisbett [4] McGilchrist [26], Sahu et al. [25], Boneg [27], Yu [28] and Santos and Achcar [15]. McGilchrist and Aisbett considered semi-parametric Cox proportional hazards model with log-normal frailty distribution and

applied Newton-Raphson iterative procedure to estimate the parameters of the model. McGilchrist [26] and Yu [28] both considered the same model as in McGilchrist and Aisbett but McGilchrist estimated the parameters of the model using BLUP, ML and REML methods and Yu proposed modified EM algorithm and penalized partial likelihood method. Sahu et al. considered four parametric models first two are piecewise exponential model with constant baseline hazard say  $\lambda_k$  within each interval having gamma prior for  $\lambda_k$  for Model-I and for Model-II normal prior to. Both the models have frailty distribution as gamma. Other two models are multiplicative gamma frailty as Model-III and additive frailty as Model-IV with Weibull as baseline distribution. Santos and Achcar used MCMC method to estimate the parameters of parametric regression model with Weibull and generalized gamma distribution as baseline and gamma and log-normal as frailty distributions. Boneg considered Cox proportional hazards model and also parametric frailty models. In parametric frailty models he considered Weibull distribution as the baseline and log-normal, Weibull as frailty distributions. He applied MHL and RMHL methods to estimate the parameters of the models.

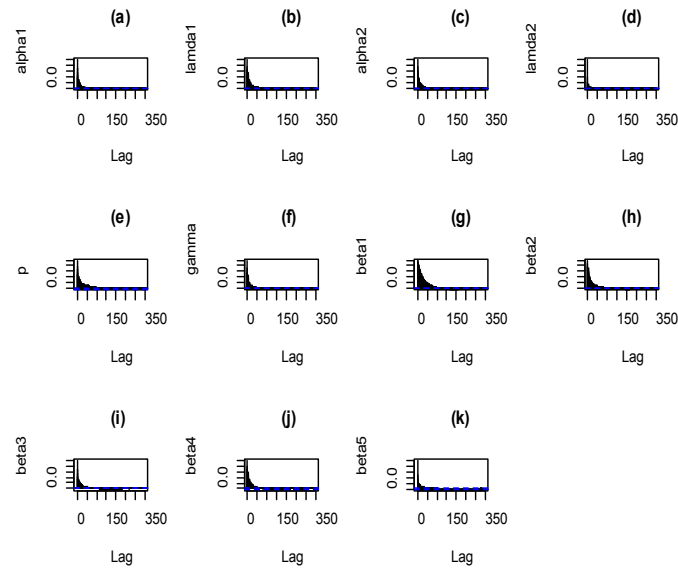
We run two parallel chains for both models using two sets of prior distributions with the different starting points using Metropolis-Hastings algorithm and Gibbs sampler based on normal transition kernels. On the similar line of simulation, here also we assume same set of prior distributions. We iterate both the chains for 95000 times. We present here the analysis for only one chain with  $G(a_1; a_2)$  as prior for baseline parameters, for both the proposed models. Due to lack of space we are presenting trace plots, coupling from the past plots and sample autocorrelation plots for the parameters of Model I only as shown in **Figures 2-4**. Gelman-Rubin convergence statistic values are nearly equal to one and Geweke test statistic values are quite small and corresponding p-values are large enough to say the chains attain stationary distribution. Simulated values of parameters have autocorrelation of lag  $k$ , so every  $k^{\text{th}}$  iteration is selected as sample. The posterior mean and standard error with 95% credible intervals for baseline parameters, frailty parameter and regression coefficients are presented in **Table 4**. The AIC, BIC and DIC values for both the models are given in **Table 5**. The Bayes factor for the proposed models is also computed.



**Figure 2.** Trace plot (Chain I) for model II.



**Figure 3.** Coupling from the past plots for model II.



**Figure 4.** Autocorrelation graphs for model II.

**Table 4.** Posterior summary for kidney infection data.

Parameter	Estimates	Standard errors	Credible Intervals Lower Upper	
<b>Model-I</b>				
<b>Burn in period=4290; Autocorrelation lag=165</b>				
$\alpha_1$	188.677	5.852149	179.5511	198.4807
$\lambda_1$	0.001499921	0.0004127996	0.001018017	0.002643678
$\alpha_1$	83.66022	5.617377	74.49408	93.32466
$\lambda_2$	0.002519979	0.000608146	0.001433446	0.00377513
$\rho$	8.324508e-05	2.647687e-05	4.247646e-05	0.0001352153
$\xi$	0.09982988	0.005377886	0.09099296 [1] 0.1094011	0.1094011
$\beta_1$	-0.03098565	0.01499194	-0.06100589	0.001318125
$\beta_2$	-2.232593	0.5281014	-3.262621	-1.185981
$\beta_3$	0.2489173	0.7285625	-1.259335	1.633234
$\beta_4$	0.8977515	0.7253961	-0.4573143	2.309644
$\beta_5$	-0.1771244	1.087419	-2.360675	1.969119
<b>Model-II</b>				
<b>Burn in period=3360; Autocorrelation lag=345</b>				
$\alpha_1$	0.05648459	0.0207349	0.02031303	0.09277614
$\lambda_1$	0.001774532	0.0005236752	0.001025109	0.00289744
$\alpha_2$	0.002125575	0.0008264149	0.0005656352	0.003393323
$\lambda_2$	0.00205937	0.000452556	0.001109531	0.003051102
$\rho$	8.372903e-05	4.18004e-05	2.187317e-05	0.0001699641
$\xi$	791.7109	43.40519	717.4634	866.5391
$\beta_1$	-0.006614582	0.01437798	-0.03282809	0.02471562
$\beta_2$	-3.059486	0.5444863	-4.304183	-1.934005
$\beta_3$	0.368245	0.6982481	-1.021127	1.74352
$\beta_4$	1.13813	0.7731091	-0.3421089	2.798761
$\beta_5$	-0.3338178	1.175335	-2.707662	1.93735

**Table 5.** AIC, BIC and DIC values for kidney infection data.

Model	Distribution	AIC values	BIC values	DIC values
I	Pareto	705.8871	723.9006	696.5783
II	linear failure rate	690.1407	708.1542	681.47

From the **Table 4**, we can observe that for these two models, the value zero is not a credible value for the credible interval of the regression coefficient  $X_2$  so  $X_2$  that is sex variable seems significant. Negative value of  $\beta_2$  indicates that the female patients have a slightly lower risk of infection. The remaining all the covariates  $X_4$  i.e. age,  $X_5$  i.e. disease type PKD  $X_3$  i.e. disease type GN and covariate  $X_4$  i.e. disease type AN are insignificant for both the Models.

The estimate of  $\sigma^2$  from two models shows that there is a strong evidence of high degree of heterogeneity in the population of patients. Some patients are expected to be very prone to infection compared to others with same covariate value. This is not surprising, as seen in the data set there is a male patient with infection time 8 and 16, and there is also male patient with infection time 152 and 562.

The comparison between two proposed models is done using AIC, BIC and DIC values given in **Table 5**. It is observed that both Model I and Model II have AIC, BIC except DIC values are near about same. An alternate way to take the decision about the best model between the proposed Model I and Model II, we use the Bayes factor which is defined as,

$$M_{ij} = 2 \log \left( \frac{I_i}{I_j} \right) \quad (9.1)$$

Where  $I_i$  is as defined as in equation (7.5)

Using Equation (9.1) we computed Bayes factor  $M_{21}$  which is 6.3162 for our proposed models.

We can observe that the Bayes factor of Model II vs Model I is 6.3162 implies that, there is positive evidence against Model-I, so Model-II is better than Model-I. Thus, Model II is best model of the proposed compound negative binomial frailty models. Now, we are in a position to say that, we have suggested a new shared compound negative binomial frailty model with linear failure rate distribution as the baseline distribution is best for modeling of kidney infection data. For simulation study and to analysis the kidney infection data we used R software.

## CONCLUSIONS

In the present study we discuss results for the two proposed models of compound negative binomial frailty namely Pareto, linear failure rate distribution as the baseline distributions. Our aim is to find the model which fit best between proposed models. For maximum likelihood estimate, likelihood equations do not converge and method of maximum likelihood fails to estimate the parameters so we use Bayesian approach. Using the Bayesian approach we perform simulation study and analyze kidney infection data. The estimate of frailty variance from different models (Model-I=1.0009; Model-II=1.0001) shows that there is a strong evidence of high degree of heterogeneity in the population of patients. The covariate sex is significant for both models. Negative value of the regression coefficient ( $\beta_2$ ) of covariate sex indicates that the female patients have a slightly lower the risk of infection. Negative value of the regression coefficient ( $\beta_5$ ) of covariate, the disease type PKD indicates that the patient with the absence of this diseases have a slightly lower the risk of infection. On the basis of AIC, BIC, DIC and Bayes factor, the Model II, the shared compound negative binomial frailty with linear failure rate distribution is best model for kidney infection.

## REFERENCES

1. Vaupel JW, et al. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979;1:439-454.
2. Keyfitz N and Littman G. Mortality in a heterogeneous population. *Population Studies*. 1979;33:333-42.
3. Clayton DG. A model for association in bivariate life tables and its applications to epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*. 1978;65:141-151.
4. McGilchrist CA and Aisbett CW. Regression with frailty in survival analysis. *Biometrics*. 1991;47:461-466.
5. Aalen OO. Heterogeneity in survival analysis. *Statistics in Medicine*. 1988;7:1121-37.
6. Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*. 1992;2:951-72.
7. Deshpande JV and Purohit SG. *Life Time Data Statistical Models and Methods*. World Scientific, New Jersey. 2005.
8. Bain LJ. Analysis for the Linear Failure-Rate Life-Testing Distribution. *Technometrics*. 1974;164:551-559.
9. Sen A and Bhattacharya GK. Inference procedure for the linear failure rate model", *Journal of Statistical Planning and Inference*. 1995;46:59-76.
10. Lin CT, et al. Monte Carlo methods for Bayesian inference on the linear hazard rate distribution. *Communications in Statistics -Theory and Methods*. 2006;35:575-590.
11. Ghitany ME and Kotz S. Reliability properties of extended linear failure-rate distributions. *Probability in the Engineering and Information Sciences*. 2007;21:441-450.

12. Hanagal DD and Dabade. Modeling inverse Gaussian frailty model for bivariate survival data. *Communications in Statistics, Theory and Methods*.2013;4220:3744-69.
13. Kheiri S, et al. Bayesian analysis of an inverse Gaussian correlated frailty model. *Computational Statistics and Data Analysis*.2007;51:5317-5326.
14. Ibrahim JG, et al. *Bayesian Survival Analysis*. Springer, Verlag. 2001.
15. Santos CA and Achcar JA. A Bayesian analysis for multivariate survival data in the presence of covariates. *Journal of Statistical Theory and Applications*.2010;9:233-253.
16. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrox BN and Caski F. *Second International Symposium on Information Theory*. Budapest Akademiai Kiado. 1973.
17. Schwarz GE. Estimating the Dimension of a Model. *Annals of Statistics*. 1978; 62:461-464.
18. Shibata R. Selection of the Order of an Autoregressive Model by Akaike's Information Criterion. *Biometrika*. 1976;63:117-126.
19. Katz RW. On Some Criteria for Estimating the Order of a Markov Chain. *Technometrics*. 1981;23:243-249.
20. Spiegelhalter D, et al. Bayesian Measures of Model Complexity and Fit. *J Roy Statist Soc Ser B*. 2002;64:583-639.
21. Kadane J and Lazar N. Methods and Criteria for Model Selection. *Journal of the American Statistical Association*. 2004;99465:279-90.
22. Raftery AE. Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models. *Biometrika*. 1994;832:251-266.
23. Jeffreys H. *Theory of Probability*. (3rd edn) Oxford Oxford University Press. 1961.
24. Kass RA and Raftery AE. Bayes' factor. *Journal of the American Statistical Association*. 1995;90:773-795.
25. Sahu SK, et al. A Weibull regression model with gamma frailties for multivariate survival data. *Life time data analysis*. 1997;3:123-137.
26. McGilchrist CA. REML estimation for survival models with frailty. *Biometrics*. 1993;49:221-225.
27. Boneg YT. Weibull frailty for modelling heterogeneity. Unpublished Ph.D. thesis, University of Gueiph. 2001.
28. Yu B. Estimation of shared gamma frailty models by a modified EM algorithm.*Computational Statistics and Data Analysis*. 2006;50:463-474.