Thailand Statistician
October 2024; 22(4): 803-820
http://statassoc.or.th
Contributed paper

Compound Poisson Correlated Frailty Model Based on Modified Weibull Baseline Distribution for Bivariate Survival Data

Arvind Pandey [a] and Lal Pawimawha [b]*
[a] Department of Statistics, Central University of Rajasthan, Rajasthan, India
[b] Department of Statistics, Pachhunga University College, Mizoram University, Aizawl, India
*Corresponding author; e-mail: raltelalpawimawha08@gmail.com

Received: 6 February 2022
Revised: 25 April 2022
Accepted: 25 April 2022

Abstract

Frailty models are survival models that are used to investigate the features of unobserved heterogeneity in people as it relating to disease and death. Despite their drawbacks, shared frailty models are frequently utilized. Several correlated frailty models were developed over the previous decade to solve these drawbacks. The performance of a compound Poisson correlated frailty model by considering modified Weibull distribution as the model baseline distribution is investigated in this work. The parameters in the proposed models are estimated by adopting Bayesian estimation procedure under the Markov chain Monte Carlo (MCMC) method. In addition, a comparison of the parameters' true values with estimated values is done using a simulation study.The data from Kidney infection was then used to test the proposed models. Models are compared to existing models using different information criteria and the Bayes factor. Accordingly, the best model for infected patient's data that have had their catheters inserted has been proposed.

Keywords: Bayesian comparison techniques, correlated frailty model, compound Poisson distribution, MCMC, modified Weibull distribution.

1. Introduction

A univariate frailty model (Vaupel et al. 1979) was introduced to assess the heterogeneity among individuals due to unobserved covariates. The univariate frailty model was extended in the early studies of Clayton (1978) and Holt and Prentice (1974), and can be considered to overcome the dependency problem between event times. Frailty models have been used frequently for modeling dependence in multivariate time-to-event data (Oakes, 1972; Yashin et al., 1995). To tackle bivariate data, a shared frailty model was introduced by Clayton (1978), which provides a useful way to describe the correlation within a group, but it has some limitations (Vaupel et al. 1979; Xue and Ding, 1999). To solve this problem, a correlated frailty approach was developed. Butler et al. (1986) explained the importance of examining the interaction between heterogeneous variables from the same individual for the study population. Correlated frailty model allows for more flexibility than shared frailty model, because it allows for the inclusion of frailty in a pair of persons with differing (but correlated) frailty. In order to understand the relationship between these two variables, it is useful to evaluate the correlation coefficient between them. As in the case of the shared frailty model, the two lifetimes of the pair are assumed to be conditionally independent given the frailties.

For the correlated frailties W_1 and W_2, the conditional survival function in the absence of observe covariates for survival times x_1 and x_2 appears as follows

$$S(x_1, x_2|W_1, W_2) = S(x_1|W_1)S(x_2|W_2) = e^{-W_1 M_{01}(x_1)} e^{-W_2 M_{02}(x_2)}.$$

Each individual's frailty is described by a measure of relative risk in the correlated frailty model (bivariate). Given the frailties, we assume that the frailties have a multiplicative effect on the baseline hazard function and that the data in a pair are conditionally independent. As a result, the hazard of the individual j ($j = 1, 2$) in pair k ($k = 1, \ldots, n$) takes the following form.

$$m(x|Yjk, W_{jk}) = W_{jk} m_{0k}(x) e^{\beta' Y_{jk}},$$

where x represents time or age, Y_{jk} stands for individuals j ($j = 1, 2$) observed covariates of pair k ($k = 1, \ldots, n$), β represents the regression coefficients, m_{0k} represents the baseline hazard function and W_{jk} stands for frailties. Frailties are related and occur in joint distribution. The occurring of independence between two frailties results in independent lifetimes, and the model has no clustering. If the two frailties are equal, the shared frailty model is obtained as a special case of the correlated frailty model with correlation one between the frailties (Wienke, 2010).

For given the frailty, the marginal likelihood function is constructed by assuming the conditional independent of the lifetimes. The conditional likelihood function in connection with censoring indicator ω_{jk} for individual j ($j = 1, 2$) in pair k ($k = 1, \ldots, n$). The conditional likelihood function for individual j in pair k with censoring indicator ω_{jk} having value 1 if the person has gone through the event of interest, and 0 otherwise is given by

$$[W_{jk} m_{0k}(x) e^{\beta' Y_{jk}}]^{\omega_{jk}} e^{W_{jk} M_{0k}(x_{jk})} e^{\beta' Y_{jk}},$$

where x_{jk} denotes the time spent observing individual j from pair k.

We may get marginal likelihood function by integrating the conditional likelihood function with respect to frailties and is given as

$$\prod_{k=1}^{n} \int_{R^2} \int [W_{1k} m_{01}(x_{1k}) e^{\beta_k x_{1k}}]^{\omega_{1k}} e^{-W_{1k} m_{o1}(x_{1k})} e^{\beta_k Y_{1k}}$$
$$[W_{2k} m_{02}(x_{2k}) e^{\beta_k Y_{2k}}]^{\omega_{2k}} e^{-W_{2k} m_{02}(x_{2k})} e^{\beta_k Y_{2k}} g(w_{1k}, w_{2k}) dw_{1k} dw_{2k},$$

where $g(\cdot, \cdot)$ stands for the probability density function related to frailty distribution.

In this work, we examine the properties of the different correlated frailty models using the data with right censored by taking compound Poisson as the frailty distribution under modified Weibull baseline distribution. Because the hazard function has a flexible property, which is one of the most significant qualities in real-life data analysis, the modified Weibull distribution was chosen as the baseline distribution. Yashin et al. (1995) and Pickles et al. (1994) established the correlated gamma frailty model, which has been applied to related lifetimes in many various scenarios. Bayesian analysis was initially utilized by Kheri et al. (2007) to analyze inverse Gaussian correlated frailty models. An analysis of bivariate survival data by Hanagal et al. (2017) also included the use of correlated frailty models with gamma as a frailty distribution. For inverse Gaussian correlated frailty models, Hanagal and Pandey (2018) used the procedure of Bayesian techniques for the estimation of the parameters and compare models under different baseline distributions. Hanagal (2021) developed a positive stable correlated frailty model that was used to analyze infectious data resulting from catheter insertion. Under the logistic exponential baseline distribution, Pandey and Lalpawimawha (2020) examined gamma and inverse Gaussian used as frailty distribution in correlated frailty models. According to the variance value of the frailty distribution, the amount of heterogeneity presented in the study population is determined. A degenerate distribution is one in which the variance of the frailty distribution is zero.

Estimation techniques such as the maximum likelihood and Bayesian approaches are both widely employed. Frailty, on the other hand, may provide a problem for the standard maximum likelihood strategy. Furthermore, standard maximum likelihood based inference methods may not be appropriate due to limited sample sizes and excessive censoring. As a result, in our scenario, a Bayesian approach is the most appropriate strategy, as it is free of such issues. Using the dataset obtained from kidney infection, all estimation processes and model comparisons are demonstrated.

In Sections 2, 3 and 4 introductions of the compound Poisson frailty model compound Poisson correlated frailty model, correlated gamma frailty model are given, followed by baseline distribution and proposed models in Sections 5 and 6. Parameters estimation and simulation study are given in Sections 7 and 8. Analysis of kidney infection data and their conclusions are discussed in Sections 9 and 10.

2. Compound Poisson Frailty Model

There are many other frailty distributions, but the gamma distribution is the most widely employed because of its mathematical ease. On the downside, it may reduce the effect of covariates (Kheri et al., 2007). In that case, gamma distribution may not be most suitable to analyse such data, Aalen (1988, 1992) employed the compound Poisson for the frailty distribution. An interesting feature is that it creates a zero-vulnerability in a subgroup of a population, the events have never been experienced under the study. Though the density is continuous with infinite series, it is also necessary to calculate numerically. Since the Laplace transform can be expressed as a simple form which makes the distribution mathematically suitable. The expression of the compound Poisson variable (X) is given as

$$X = \begin{cases} Y_1 + Y_2 + ... + Y_M & ; M > 0 \\ 0 & ; M = 0. \end{cases}$$

where M is Poisson distributed random variable with mean v and having Laplace transform as $L_M(\epsilon)$=$\exp(-v + ve^{-\epsilon})$, while Y_1, Y_2,... represents independent and identically distributed gamma distribution having the k and λ as the parameters and Laplace transform is $L_Y(\epsilon) = (1 + \frac{\epsilon}{\lambda})^{-k}$. One appealing interpretation of this model is that each individual is subjected to a large number of hits that cause damage. The effects of those hits compound over time, causing a person's frailty to rise. It is worth noting that this treatment occurs before follow-up begins, possibly as early as childhood, with the assumption that follow-up frailties are repaired from the outset.

The distribution of W includes a discrete part with zero susceptible with probability equal to $e^{-\mu}$ i.e. P($W = 0$)= $e^{-\mu}$, which is decreases as μ increases and a continuous part having the distribution as

$$g(v; \nu, \phi, \mu) = \begin{cases} \exp[-(\mu + \phi v)] \frac{1}{2} \sum_{n=1}^{\infty} \frac{\mu^n (\phi v)^{n\nu}}{\Gamma(n\nu) n!} & ; \ v > 0, \phi > 0, \mu > 0, \nu > 0 \\ 0 & ; \text{ otherwise.} \end{cases}$$

The expectation is $E(W)$= $\frac{\mu\nu}{\phi}$ and the variance becomes $V(W)$ = $\frac{\mu\nu(\nu+1)}{\phi^2}$.

The Laplace transform, on the other hand, is deduced as follows: $L_w(\epsilon)$= $\exp\left\{-\mu[1 - (\frac{\phi}{\phi+\epsilon})^{\nu}]\right\}$.

For identifiability purposes, the mean of the distribution is equal to one, which implies that ϕ=$\mu\nu$. The simplified form of Laplace transform using the above condition becomes $L_w(\epsilon) = \exp\left\{-\mu[1 - (1 + \frac{\epsilon}{\mu\nu})^{-\nu}]\right\}$ and frailty variance becomes τ^2=$\frac{\nu+1}{\mu\nu}$.

3. Compound Poisson Correlated Frailty Model

The Laplace transform of W_1, W_2 by assuming there are frailty variable W with infinitely divisible possessing Laplace transform $L_W(\epsilon)$ and ρ belong to (0,1), then the random variables W_1, W_2 are existed each with univariate Laplace transform $L_W(\epsilon)$ is given as (Hanagal, 2011)

$$L(\epsilon_1, \epsilon_2) = L_W^{\rho}(\epsilon_1 + \epsilon_2) L_W^{1-\rho}(\epsilon_1) L_W^{1-\rho}(\epsilon_2) \quad (1)$$

If the frailty variable W has a variance, then the $Corr(W_1, W_2) = \rho$.

Under mild regularity criteria on W, the respective bivariate survival model can be identified if $\rho > 0$. In the case, when $\rho = 1$ is known as the shared Frailty model.

Let W_k be the compound Poisson distributed with mean 1, variance τ, and Laplace transform is given as

$$L(\epsilon_k, \cdot) = \exp\left\{-\mu\left[1 - \left(1 + \frac{\epsilon k}{\mu\nu}\right)^{-\nu}\right]\right\}.$$

The bivariate Laplace transform for the compound Poisson correlated frailty model by using Eqn. (1) is given by

$$\begin{aligned} L(\epsilon_1, \epsilon_2, \mu, \nu, \rho) = & \exp\left\{\rho\mu\left[1 - \left(1 + \frac{\epsilon_1 + \epsilon_2}{\mu\nu}\right)^{-\nu}\right]\right\} \exp\left\{(1-\rho)\mu\left[1 - \left(1 + \frac{\epsilon_1}{\mu\nu}\right)^{-\nu}\right]\right\} \\ & \exp\left\{(1-\rho)\mu\left[1 - \left(1 + \frac{\epsilon_2}{\mu\nu}\right)^{-\nu}\right]\right\}, \end{aligned}$$

where $Corr(W_1, W_2) = \rho$. The correlated frailty model with compound Poisson frailty distribution is characterized by the bivariate survival function of the form:

$$\begin{aligned} S(x_{1k}, x_{2k}) &= L(M_{01}(x_{1k}), M_{02}(x_{2k}), \mu, \nu, \rho) \\ &= \exp\left\{\rho\mu\left[1 - \left(1 + \frac{\eta_k(M_{01}(x_{1k}) + M_{02}(x_{2k}))}{\mu\nu}\right)^{-\nu}\right]\right\} \\ &\quad \exp\left\{(1-\rho)\mu\left[1 - \left(1 + \frac{\eta_k(M_{01}(x_{1k}))}{\mu\nu}\right)^{-\nu}\right]\right\} \\ &\quad \exp\left\{(1-\rho)\mu\left[1 - \left(1 + \frac{\eta_k(M_{02}(x_{2k}))}{\mu\nu}\right)^{-\nu}\right]\right\}, \end{aligned} \quad (2)$$

where $\eta_k = e^{\underline{Y}_k\underline{\beta}}$, $M_{01}(x_{1k})$ and $M_{02}(x_{2k})$ represents the cumulative baseline hazard functions of the life time random variables X_{1k} and X_{2k}respectively.

4. Correlated Gamma Frailty Model

The unconditional survival function is derived for the correlated gamma frailty model, which is developed by Yashin et al. (1995) for $x_{1k} > 0$ and $x_{2k} > 0$ is given as

$$\begin{aligned} S(x_1, x_2) = & (1 + \eta_k\tau_1^2 M_{01}(x_{1k}) + \eta_k\tau_2 M_{02}(x_{2k}))^{\frac{-\rho}{\tau_1\tau_2}} \times (1 + \eta_k\tau 1^2 M_{01}(x_{1k}))^{\frac{-1+\frac{\tau_1}{\tau_2}\rho}{\tau_1^2}} \\ & (1 + \eta_k\tau_2^2 M_{02}(x_{2k}))^{\frac{-1+\frac{\tau_2}{\tau_1}\rho}{\tau_2^2}}, \end{aligned} \quad (3)$$

where ${\tau_1}^2$ and ${\tau_2}^2$ denote the variances of the frailty variables W_1 and W_2.

When the frailty variable is degenerate, the unconditional survival function is given by

$$S(x_{1k}, x_{2k}) = e^{-[(M_{01}(x_{1k}) + M_{02}(x_{2k}))\eta_k]}. \quad (4)$$

5. Modified Weibull as Baseline Distribution

The baseline distribution used in this study is the modified Weibull distribution presented by Sarhan and Zaindin (2008), which is an extension of the Weibull distribution used in data relating to fatigue life with the hazard function for time X as

$$r(x) = \alpha + \lambda\gamma x^{\gamma} \ ; x > 0, \alpha > 0, \lambda > 0, \gamma > 0.$$

The cumulative hazard function and survival function respectively, are as follows:

$$\begin{aligned} M(x) &= \alpha x + \lambda x^{\gamma} \ ; x > 0, \alpha > 0, \lambda > 0, \gamma > 0, \\ S(x) &= e^{-\alpha y - \lambda x^{\gamma}} \ ; x > 0, \alpha > 0, \lambda > 0, \gamma > 0, \end{aligned} \tag{5}$$

where α, λ and γ represent the modified Weibull distribution parameters. Depending on the values of the parameters, the distribution generalizes many other distributions. The hazard function might be constant, increasing or decreasing according to the value of the parameter.

6. Models Description

The unconditional survival function is derived by substituting the survival function of modified Weibull distribution in Eqns. (2), (3) and (4). Then,

$$\begin{aligned} S(x_{1k}, x_{2k}) &= \exp\left\{\rho\mu\left[1 - \left(1 + \frac{\eta_k(\alpha_1 x_{1k} + \lambda_1 x_{1k}^{\gamma_1} + \alpha_2 x_{2k} + \lambda_2 x_{2k}^{\gamma_2})}{\mu\nu}\right)^{-\nu}\right]\right\} \\ &\quad \exp\left\{(1-\rho)\mu\left[1 - \left(1 + \frac{\eta_k(\alpha_1 x_{1k} + \lambda_1 x_{1k}^{\gamma_1})}{\mu\nu}\right)^{-\nu}\right]\right\} \\ &\quad \exp\left\{(1-\rho)\mu\left[1 - \left(1 + \frac{\eta_k(\alpha_2 x_{2k} + \lambda_2 x_{2k}^{\gamma_2})}{\mu\nu}\right)^{-\nu}\right]\right\} \end{aligned} \tag{6}$$

$$\begin{aligned} S(x_{1k}, x_{2k}) &= (1 + \eta_k\sigma_1^2(\alpha_1 x_{1k} + \lambda_1 x_{1k}^{\gamma_1}) + \eta_k\sigma_2(\alpha_2 x_{2k} + \lambda_2 x_{2k}^{\gamma_2}))^{\frac{-\rho}{\sigma_1\sigma_2}} \times \\ &\quad (1 + \eta_k\sigma_1^2(\alpha_1 x_{1k} + \lambda_1 x_{1k}^{\gamma_1}))^{\frac{-1+\frac{\sigma_1}{\sigma_2}\rho}{\sigma_1^2}} \\ &\quad (1 + \eta_k\sigma_2^2(\alpha_2 x_{2k} + \lambda_2 x_{2k}^{\gamma_2}))^{\frac{-1+\frac{\sigma_2}{\sigma_1}\rho}{\sigma_2^2}} \end{aligned} \tag{7}$$

$$S(x_{1k}, x_{2k}) = \exp[-[(\alpha_1 x_{1k} + \lambda_1 x_{1k}^{\gamma_1} + \alpha_2 x_{2k} + \lambda_2 x_{2k}^{\gamma_2})\eta_k]]. \tag{8}$$

Eqns. (6) and (7) are compound Poisson correlated frailty model and correlated gamma frailty model based on modified Weibull as baseline distribution and called Model-I and Model-II. Eqn. (8) is without frailty model with modified Weibull as baseline distribution, called Model-III.

7. Parameters Estimation and Models Comparison

Assume that m individuals are being studied, and each individual is supposed to have a first and second failure time, which are represented by (x_{1k}, x_{2k}) and let $(\omega_{1k}, \omega_{2k})$ be the times of censoring in connection with the times of failure for the k^{th} individual $(k = 1, 2, 3, ..., m)$. Further, censorship schemes and life times of individuals are believed to be independent of one another.

The likelihood function for the k^{th} individual's bivariate life time random variable is given by,

$$L_k(x_{1k}, x_{2k}) = \begin{cases} g_1(x_{1k}, x_{2k}), & ; \ x_{1k} < \omega_{1k}, x_{2k} < \omega_{2k}, \\ g_2(x_{1k}, \omega_{2k}), & ; \ x_{1k} < \omega_{1k}, x_{2k} > \omega_{2k}, \\ g_3(\omega_{1k}, x_{2k}), & ; \ x_{1k} > \omega_{1k}, x_{2k} < \omega_{2k}, \\ g_4(\omega_{1k}, \omega_{2k}), & ; \ x_{1k} > \omega_{1k}, x_{2k} > \omega_{2k}, \end{cases}$$

and in the presence of frailty, the likelihood function is,

$$L(\underline{\phi}, \underline{\beta}, \tau) = \prod_{k=1}^{m_1} g_1(x_{1k}, x_{2k}) \prod_{k=1}^{m_2} g_2(x_{1k}, \omega_{2k}) \prod_{k=1}^{m_3} g_3(\omega_{1k}, x_{2k}) \prod_{k=1}^{m_4} g_4(\omega_{1k}, \omega_{2k}), \quad (9)$$

where τ stands for the frailty parameter, $\underline{\phi}$ and $\underline{\beta}$ are the vector representation of baseline parameters and regression coefficients respectively.

In the absence of frailty, the likelihood function is

$$L(\underline{\phi}, \underline{\beta}) = \prod_{k=1}^{m_1} g_1(x_{1k}, x_{2k}) \prod_{k=1}^{m_2} g_2(x_{1k}, \omega_{2k}) \prod_{k=1}^{m_3} g_3(\omega_{1k}, x_{2k}) \prod_{k=1}^{m_4} g_4(\omega_{1k}, \omega_{2k}), \quad (10)$$

where m_1, m_2, m_3 and m_4 denote the random number of observations found to fall within the ranges $x_{1k} < \omega_{1k}, x_{2k} < \omega_{2k}$; $x_{1k} < \omega_{1k}, x_{2k} > \omega_{2k}$; $x_{1k} > \omega_{1k}, x_{2k} < \omega_{2k}$ and $x_{1k} > \omega_{1k}, x_{2k} > \omega_{2k}$ respectively and the k^{th} individual's contribution to likelihood function as follows:

$$\begin{aligned} g_1(x_{1k}, x_{2k}) &= \frac{\partial^2 S(x_{1k}, x_{2k})}{\partial x_{1k} \partial x_{2k}} \\ g_2(x_{1k}, \omega_{2k}) &= -\frac{\partial S(x_{1k}, \omega_{2k})}{\partial x_{1k}} \\ g_3(\omega_{1k}, x_{2k}) &= -\frac{\partial S(\omega_{1k}, x_{2k})}{\partial x_{2k}} \\ g_4(\omega_{1k}, \omega_{2k}) &= S(\omega_{1k}, \omega_{2k}). \end{aligned}$$

The likelihood functions for correlated frailty models and without frailty models are obtained by replacing the survival functions, differentiating them, and substituting them in Eqns. (9) and (10).

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

$$\pi(\alpha_1, \lambda_1, \alpha_2, \lambda_2, \sigma, \underline{\beta}) \propto L(\alpha_1, \lambda_1, \alpha_2, \lambda_2, \sigma, \underline{\beta}) \times f_1(\alpha_1) f_2(\lambda_1) f_3(\alpha_2) f_4(\lambda_2) f_5(\sigma) \prod_{i=1}^{5} q_j(\underline{\beta}_i),$$

where $f_j(\cdot)$ $(j = 1, 2, ..., 5)$ denotes a prior density function with specified hyper-parameters for baseline parameters and frailty variance; $q_j(\cdot)$ represents prior distribution function for regression coefficient β_j; Except for β_j, $\underline{\beta}_j$ is a vector representation of regression coefficients, $i = 1, 2, ..., a$ and Eqn. (9) or (10) gives the likelihood function $\mathcal{L}(\cdot)$. All of the parameters are considered to be distributed independently in this case.

The Newton Raphson method is not easily used to solve the likelihood function expressions in Eqns. (9) and (10). Unfortunately, calculating maximum likelihood estimators (MLEs) necessitates solving an optimization problem with fourteen parameters for Models I and II, and an optimization problem with eleven parameters for Model III. In the Newton Raphson iterative technique, there is an issue with the convergence of estimates because we are estimating several parameters at once. As a result of a convergence difficulty, the ML technique fails to estimate the parameters. Then, parameters were estimated using the Bayesian technique, which does not suffer from this problem.

The following distributions are considered as the prior distribution in the study - gamma distribution with a mean of one and a high variance $\Gamma(\zeta, \zeta)$ with a small value of ζ is used as a prior distribution for frailty parameter. As a prior for the regression coefficient, a normal distribution with mean zero and huge variance, say ψ^2 is utilized. Prior distributions are of the same type as those analyzed by Ibrahim et al. (2001) and Sahu et al. (1997) and We assumed a non-informative prior for the baseline parameters because we don not know anything about them. $\Gamma(a_1, b_1)$and $U(a_2, b_2)$ are engaged as distributions for non-informative prior. In this case, it is assumed that all of the hyper-parameters ζ, ψ, a_1, a_2, b_1 and b_2 are known. Here Gamma distribution is represented by the shape

parameter a_1 and scale parameter b_1, whereas the uniform distribution is represented by U(a_2, b_2) throughout the interval from a_2 to b_2 respectively. We set the hyper-parameters to ζ equals 0.0001, ψ^2 equals 1000, a_1 equals 1, b_1 equals 0.0001, a_2 equals 0, and b_2 equals 100.

The Metropolis-Hastings (M-H) Algorithm and Gibbs Sampler are also employed in order to estimate parameters in models fitted using the above prior density function and likelihood functions (9) and (10). The Geweke (GK) test and Gelman-Rubin (GR) Statistics, as proposed by Geweke (1992) and Gelman and Rubin (1992) show that the Markov chain is convergent to a stationary distribution. The trace plot, coupling from the past plots, and sample autocorrelation plots were all utilized to verify the behavior of the chain, determine the burn-in period, and determine autocorrelation lag.

The algorithm involves obtaining samples continuously from the conditional distribution of each parameter from all other parameters of a given model. These distributions are called fully conditional distributions. Full conditional distributions are difficult to integrate out in our scenario. As a result, fully conditional distributions can be calculated by assuming that they are proportional to the joint distribution of the model's parameters.

It is important to determine the best model for a given dataset. Model comparisons were made using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Deviance Information Criterion (DIC), and Bayes Factor. On the basis of the posterior predictive density samples, we have calculated predictive intervals to test the models' adequacy. In recent years, the Gelfand and Ghosh (1998) model choice criterion has also been taken into account. The criterion is,

$$C_\eta = \sum_{k=1}^{m} \tau_k^2 + \frac{\eta}{\eta+1} \sum_{k=1}^{m} (\phi_k - \xi_k)^2,$$

where ϕ_k and τ_k^2 represents the Monte Carlo estimates of mean and variance of the k^{th} observation based on the posterior predictive density, $\eta > 0$ represents a constant and $\xi_k = x_{k,obs}$ if there is k^{th} observation in a time of failures and $\phi_k = \max(\phi_k, \omega_k)$ if ω_k is the censoring time in k^{th} observation. Among the models considered, the one with the lowest value of C_η is chosen as the best.

8. Models Checking with Simulation Study

A simulation study was conducted to test the performance of the Bayesian estimating method. As a result, just one covariate X_0 from a normal distribution was used as a covariate for analysis. Due to the time constraints of Bayesian methods, fifty sets of lifetimes were created using the inverse transform approach with α_1 = 0.0042, λ_1= 1.10, γ_1= 1.20 , α_2= 0.0042, λ_2= 1.10, γ= 1.20, ϕ = 2.64, μ= 6.60, ν= 2.5 and β= 0.002 for Model I; α_1 = 0.25, λ_1= 0.0021, γ_1= 0.42, α_2= 0.25, λ_2= 0.0021, γ = 0.42, σ = 0.004 and β= 0.005 for Model-II and The censoring times were generated using an exponential distribution, with a failure rate of 0.0021 and 0.02 for Models I and II. The frailty parameter was given a gamma distribution as a prior, and the regression coefficients were given a normal distribution [Ibrahim et al. (2001), Sahu et al. (1997), and Santos and Achcar (2010)]. There was no prior knowledge of baseline parameters, therefore non-informative prior distributions were utilized (i.e. $\Gamma(a_1, b_1)$ and U(a_2, b_2)) and all hyper-parameters were presumed to be known before the experiment. Both chains were iterated a hundred thousand times. Trace plots, coupling from the past plots, sample autocorrelation plots, and running mean plots are not available due to a shortage of space. The zigzag shape of trace plots indicates that parameters are flowing freely and suitably.There were large p-values for GK test values and GR scale reduction factor values close to one, showing that the chains reach stationary distribution in both prior sets. Furthermore, there was not much of variation in the convergence rate. The prior distribution had no effect on posterior summary because parameter estimates were almost identical. Considering that the results of both chains were rather comparable, the study was displayed as a single-chain analysis using the prior $\Gamma(a_1, b_1)$ for all models. The posterior results of compound Poisson correlated frailty model and correlated gamma frailty model with a modified Weibull as baseline distribution are presented in Tables 1 and 2.

9. Analysis of Real Life Data

The models' applicability was evaluated by applying them to infectious diseases data pertaining to kidney infection occurring after catheter insertion (McGilchrist and Aisbett, 1991). A total of 38 dialysis patients were studied to determine the first and second recurrence times of infection from the use of catheters on portable dialysis equipment. For each patient in a cluster, these two infection periods are grouped. Other pertinent data include infection period, patient's age, gender (0 for male and 1 for female), and kinds of symptoms such as Glomerulo Nephritis (GN), Acute Nephritis (AN), and Polycystic Kidney Disease (PKD) (PKD).

First, the Kolmogorov-Smirnov test is used to verify the goodness of fit of the kidney infection data, and the p- values obtained for the first and second recurrence are large enough, and it can be said that there is no reason to reject that the first and second recurrence times follow the univariate case, with the distribution of the survival function given in Eqn. (5), we suppose that this holds true in the bivariate situation as well. Table 3 shows the related p-values. Parametric and non-parametric survival function charts also become almost identical. Figure 1 shows how the two lines of parametric and non-parametric plots are near each other.

The M-H algorithm and Gibbs sampler with normal transition kernel is used to execute two parallel chains for all models, employing two sets of prior distributions and separate beginning points for each chain. Both chains were repeated 100000 times. It is also worth noting that both prior distributions produced roughly the same parameter estimates, as a result, estimates are not affected by the various prior distributions. The Gibbs sampler converges at nearly the same rate for both prior sets. Also, both chains produce identical findings, as a result, we offer here the analysis for only one chain with $\Gamma(a_1, b_1)$ as the baseline parameters prior for all models. We also compute the GR statistic and the Geweke test statistic to see if the Markov chain has converged to a stationary distribution. The statistics value of GR and Geweke test are near to one and zero, and the associated p-values are big enough to suggest that the chains achieve stationary distribution. All of the trace plots (Figure 2) display a zigzag pattern, indicating that the parameters move and mix more freely. As a result, the Markov chain appears to have arrived at a stationary state. The burn-in period is determined by coupling from the plots (Figure 3). Autocorrelation plots (Figure 4) are also used to choose autocorrelation lag. running mean plots (Figure 5) can also be used to determine how well our chains are mixing. Running mean plot for each parameter is convergent to the posterior mean of the parameter, indicating a successful mixing of the chain. As a result, our diagnostic plots indicate that the MCMC chains mix quite effectively. The estimate of parameters (posterior mean), standard error, credible limits are given in Tables 4-6. There are no zeros in the credible intervals, which means all factors are significant. According to β_1 having a positive value, age has a key role in the development of kidney infection, and as age grows, so does infection risk. As β_2 is negative for females, it indicates that females have a lower risk of contracting an infection than males. As can be seen in the data set, there is a male patient with infection times of 8 and 16, as well as 152 and 562. Diseases types such as Glomerulo Nephritis (β_3) and Acute Nephritis (β_4) are also major factors that increase the risk for infection.

Model-I is the best model in terms of AIC and DIC values since it has lower AIC and DIC values than Model-II and Model-III but not in BIC, as shown in Table 7. However, the difference in information criteria values for Model-I, Model-II, and Model-III is quite small, hence using different information criteria values to pick between models is not recommended. We are currently comparing models j and k using the Bayes factor. In Table 8, the value of $2\log(B_{jk})$ is more than 3, 6 and 3, which demonstrates that Model-I is superior to Model-II and Model-III for the given dataset. As a result of all of the demonstrated comparison criteria, we can conclude that Model-I, which is a compound Poisson correlated frailty based on modified Weibull as baseline distribution is superior to Model-II and Model-III, which are gamma correlated frailty model and without frailty model, respectively, for modeling kidney infection data. In order to determine the suitability of Models I and II. Predictive intervals of a produced random sample have been formed with 95% and 50% equal tails, and the total number of intervals in which r^{th} observation falls in each has been calculated. Out of

76 data, the 95 percent and 50 percent prediction ranges contain 76 and 76 observations, respectively, for Model-I and Model-II.

Finally, Table 9 displays the values for the C_η model choice criterion for the models. The second gives the penalty values and the third column and fourth column for the goodness-of-fit and C_η values with different values of η. The Model I has a minimum penalty term and goodness-of-fit term and for all η values, C_η value is the lowest for Model I, hence this criterion also shows that Model-I is the preferred model to choose. Thus, for modeling kidney infection data, Model-I, a compound Poisson frailty model with modified Weibull as baseline distribution, outperforms the gamma frailty model and the without frailty model with the same baseline distribution.

10. Conclusions

In this study, the results for compound Poisson, gamma, and without frailty models with modified Weibull as baseline distribution were discussed. Our study's primary goal is to determine which model (compound Poisson correlated frailty model, gamma correlated frailty model, or without frailty model) fits best. Models I and II have compound Poisson and gamma correlated frailty, while Model III has no frailty. Using our proposed models, we did a simulation study and assessed kidney infection data. To execute the analysis, we employed self-made programs created in the R statistical software.

When estimating parameters using maximum likelihood, the likelihood equations do not converge, thus we employed the Bayesian approach. The overall Bayesian estimate technique required a significant amount of calculation time, although the duration was roughly the same for all of the proposed models. The parameter estimations were the same for different priors. In our suggested model for kidney infection data, the convergence rate of the Gibbs sampling procedure is unaffected by these prior distribution selections. It is estimated that for Model-I the variance of frailty distribution is 0.2316, whereas for Model-II the variances are $\tau_1 = 0.5747$ and $\tau_2 = 0.6490$ respectively. Patients' populations are highly heterogeneous, as evidenced by the correlated frailty models that were developed. The frailty parameter $\tau_1 = \tau_2 = 0$ was tested using the Bayes factor, and it was found that frailty exists and frailty models fit better than those without frailty. For all models, all of the covariates are significant. The negative value of the covariate sex regression coefficient (β_2) revealed that female patients were at a somewhat lower risk of infection.

Using different information criteria values, a comparison of the three proposed models was made. The Model-I (compound Poisson correlated frailty model) has the smallest AIC and DIC values, but it does not have the smallest BIC value, but these discrepancies, on the other hand, are not that significantly different. The Bayes factor is used to choose amongst Models I, II, and III. Model-I is the best, according to the results. The compound Poisson correlated frailty model (Model-I) also outperforms the correlated gamma frailty model and the without frailty model Our results are also compared to correlated frailty models, which are based on generalized Weibull and generalized log-logistic distributions as the baseline proposed by Hanagal et al. (2017). It is clear that the compound Poisson correlated frailty model with modified Weibull as baseline distribution outperforms the correlated frailty model proposed by Hanagal et al (2017). All of the foregoing analysis has led us to the conclusion that a new compound Poisson correlated frailty model using the modified Weibull distribution as a baseline distribution that is the best of all proposed models for modeling kidney infection datasets. We also draw the conclusion that a strong correlation within clusters does not signify a better model for the given data.

References

Aalen OO. Heterogeneity in survival analysis. Stat Med. 1988; 7: 1121-1137.

Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. Ann Appl Probab. 1992; 4: 951-972.

Butler JS, Anderson HA, Burkhauser RV. Testing the relationship between work and health. Econ Lett. 1986; 20: 383-386.

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.

Gelfand AE, Ghosh SK. Model choice: a minimum posterior predictive loss approach. Biometrika. 1998; 85: 1-11.

Gelman A, Rubin DB. A single series from the Gibbs sampler provides a false sense of security. Bayesian statistics. 1992; 4: 625-31.

Geweke J. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. Bayesian Statist. 1992; 4: 641-9.

Hanagal DD. Modeling survival data using frailty models. New York: Chapman & Hall/CRC; 2011

Hanagal DD, Pandey A, Ganguly A. Correlated gamma frailty models for bivariate survival data. Commun Stat Simulat. 2017; 46(5): 3627-3644.

Hanagal DD, Pandey A. Correlated inverse Gaussian frailty models for bivariate survival data, Commun Stat Theory. 2018; 49(4): 845-863.

Hanagal DD. Correlated positive stable frailty models. Commun Stat Theory. 2021; 50(23): 5617-33.

Holt JD, Prentice RL. Survival analyses in twin studies and matched pair experiments. Biometrika. 1974; 61: 17-30.

Ibrahim JG, Chen M-H, Sinha, D. Bayesian Survival Analysis. Verlag: Springer; 2001.

Kheiri S, Kimber A, Meshkani MR. Bayesian analysis of an inverse Gaussian correlated frailty model. Comput Stat Data Anal. 2007; 51: 5317-5326.

McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. Biometrics. 1991; 47: 461-466.

Oakes D. A concordance test for independence in the presence of censoring. Biometrics. 1972; 38: 451-455.

Pandey A, Lalpawimawha R. Comparison of correlated frailty models. Commun Stat Simulat. 2020; 23:1-9.

Pickles A, Crouchley R, Simonoff E at al. Survival models for developmental genetic data: Age of onset of puberty and antisocial behavior in twins. Genet Epidemiol. 1994; 11(2): 155-70.

Sahu SK, Dey DK, Aslanidou H, Sinha D. A Weibull regression model with gamma frailties for multivariate survival data. Lifetime Data Anal. 1997; 3: 123-137.

Santos CA, Achcar JA. A Bayesian analysis for multivariate survival data in the presence of covariates. J Stat Theory Pract. 2010; 9: 233-253.

Sarhan AM, Zain-Din M. New generalized Weibull distribution. Appl Sci. 2008.

Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography. 1979; 16: 439-454.

Wienke A. Frailty models in survival analysis. Chapman and Hall/CRC; 2010 Jul 26.

Xue X, Ding Y. Assessing heterogeneity and correlation of paired failure times with the bivariate frailty model. Stat Med. 1999; 18: 907-918.

Yashin AI, Vaupel JW, Iachine IA. Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. Math Popul Stud. 1995; 5(2): 145-59.

Appendix : Tables and Figures

Table 1 Compound Poisson correlated frailty model with modified Weibull baseline distribution(Simulation for Model-I)

Parameter	Estimate	Standard error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6200;		autocorrelation lag = 120					
α_1 (0.0401)	0.0403	0.0029	0.0344	0.0468	0.0004	0.5041	1.0013
α_2 (0.0191)	0.0194	0.0031	0.0140	0.0265	0.0028	0.5011	1.0006
λ_1 (0.0006)	0.0006	5.3e-05	0.0005	0.0007	-0.0087	0.4964	1.0005
λ_2 (0.0007)	0.0007	4.8e-05	0.0006	0.0008	-0.0095	0.4962	1.0015
γ_1 (0.0218)	0.0219	0.0053	0.0128	0.0309	0.0059	0.5023	1.0032
γ_2 (0.0311)	0.0310	0.0050	0.0221	0.0402	0.0114	0.5045	1.0007
μ (8.8030)	8.8032	0.5087	7.8070	9.6584	-0.0123	0.4950	1.0020
ν (0.6541)	0.6542	0.0498	0.5593	0.7409	0.0013	0.5045	0.9999
ρ (0.8233)	0.8232	0.0492	0.7379	0.9170	-0.0072	0.4970	1.0007
β (0.0002)	0.0002	4.9e-05	0.0001	0.0003	0.0065	0.5026	1.0011

Table 2 Gamma correlated frailty model with modified Weibull baseline distribution(Simulation for Model-II)

Parameter	Estimate	Standard error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6700;		autocorrelation lag = 190					
α_1 (0.0463)	0.0465	0.0030	0.0407	0.0524	-0.0010	0.4995	1.0019
α_2 (0.0354)	0.0355	0.0047	0.0268	0.0443	-0.0035	0.4985	1.0032
λ_1(0.0006)	0.0006	5.2e-05	0.0005	0.0007	-0.0052	0.4979	1.0020
λ_2(0.0007)	0.0007	5.3e-05	0.0006	0.0008	-0.0052	0.4979	0.9999
γ_1 (0.0217)	0.0218	0.0053	0.0131	0.0309	-0.0095	0.4961	0.9999
γ_2 (0.0302)	0.0304	0.0050	0.0216	0.0401	0.0099	0.5039	1.0003
σ_1 (0.6481)	0.6483	0.0498	0.5438	0.7274	-0.0124	0.4950	1.0047
σ_2 (0.5868)	0.5869	0.0500	0.5127	0.6982	-0.0098	0.4960	1.0033
ρ (0.7282)	0.7283	0.0522	0.6440	0.8272	0.0017	0.5006	1.0003
β (0.0002)	0.0002	5.2e-05	0.0001	0.0003	0.0014	0.5045	1.0032

Table 3 p-values of K-S Statistics for goodness of fit test for Kidney Infection data set

	recurrence time	
Distribution	first	second
Compound Poisson Frailty	0.9999	0.9999
Gamma Frailty	0.9998	0.9999

Table 4 Posterior summary for kidney infection data set Model-I

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6300;		autocorrelation lag = 280					
α_1	0.0384	0.0030	0.0323	0.0447	-0.0021	0.4991	1.0000
λ_1	0.0004	5.7e-05	0.0003	0.0006	-0.0012	0.4995	1.0031
γ_1	0.0173	0.0055	0.0084	0.0273	-0.0018	0.4956	1.0010
α_2	0.0276	0.0049	0.0191	0.0373	-0.0047	0.4981	1.0004
λ_2	0.0006	5.5e-05	0.0005	0.0008	-0.0023	0.4911	0.9999
γ_2	0.0260	0.0054	0.0165	0.0354	0.0092	0.5036	1.0029
μ	10.3777	0.5516	9.4594	11.3418	0.0083	0.5033	1.0031
ν	0.7125	0.0556	0.6137	0.8044	0.0050	0.5020	1.0002
ρ	0.9526	0.0568	0.8551	1.0434	-0.0046	0.4981	1.0074
β_1	0.0002	5.6e-05	0.0001	0.0003	0.0065	0.5026	1.0082
β_2	-1.8016	0.2244	-2.2494	-1.3784	0.0018	0.5026	0.9999
β_3	0.0817	0.0053	0.0725	0.0914	-0.0014	0.4943	1.0014
β_4	0.6070	0.0504	0.5205	0.6935	-0.0012	0.4995	1.0011
β_5	0.0008	5.4e-05	0.0007	0.0009	0.0040	0.5016	1.0000

Table 5 Posterior summary for kidney infection data set Model-II

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6600;		autocorrelation lag = 195					
α_1	0.0517	0.0030	0.0457	0.0579	-0.0040	0.4983	1.0000
λ_1	0.0006	5.7e-05	0.0005	0.0007	-0.0041	0.4983	1.0041
γ_1	0.0223	0.0058	0.0124	0.0316	-0.0098	0.4960	1.0082
α_2	0.0359	0.0053	0.0255	0.0446	-0.0023	0.4990	1.0008
λ_2	0.0007	5.7e-05	0.0006	0.0008	-0.0025	0.4899	1.0015
γ_2	0.0310	0.0055	0.0214	0.0405	0.0051	0.5020	1.0017
τ_1	0.5747	0.0550	0.4826	0.6739	0.0049	0.5019	1.0022
τ_2	0.6490	0.0552	0.5538	0.7412	0.0027	0.5010	1.0034
ρ	0.7667	0.0545	0.6651	0.8544	-0.0072	0.4971	1.0013
β_1	0.0002	5.5e-05	0.0001	0.0003	0.0022	0.5008	1.0074
β_2	-1.9820	0.2242	-2.4262	-1.5585	0.0008	0.5008	1.0005
β_3	0.0114	0.0051	0.0016	0.0200	-0.0036	0.4945	1.0002
β_4	0.4143	0.0521	0.3161	0.5031	-0.0058	0.4976	1.0002
β_5	0.0032	0.0005	0.0022	0.0041	0.0051	0.5020	1.0000

Table 6 Posterior summary for kidney infection data set Model-III

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 5900;		autocorrelation lag = 240					
α_1	0.0050	0.0003	0.0043	0.0056	0.0044	0.5017	1.0007
α_2	0.0053	0.0004	0.0041	0.0059	0.0047	0.5018	1.0083
λ_1	0.0069	0.0005	0.0060	0.0079	-0.0052	0.4978	1.0208
λ_2	0.0031	0.0005	0.0022	0.0040	0.0012	0.5005	1.0061
γ_1	0.9938	0.0472	0.9232	1.0912	0.0020	0.5008	1.0074
γ_2	0.5042	0.0521	0.4123	0.5910	0.0046	0.5018	1.0005
β_1	0.0201	0.0046	0.0109	0.0285	0.0033	0.5013	1.0001
β_2	-1.4733	0.2793	-2.0170	-0.9352	-0.0041	0.5013	1.0007
β_3	0.0032	0.0005	0.0022	0.0041	-0.0031	0.4987	1.0003
β_4	0.6469	0.0510	0.5580	0.7423	-0.0096	0.4961	1.0010
β_5	-1.1094	0.4041	-1.8762	-0.2903	-0.0050	0.4979	1.0000

Table 7 AIC, BIC and DIC values for all the models fitted to kidney infection data set

Model No.	AIC	BIC	DIC
Model I	686.8472	706.4982	665.9469
Model II	689.3618	709.0129	668.4854
Model III	688.0174	706.0308	672.2273

Table 8 Bayes factor for three models

-	M12	M13	M23
Bayes Factor	3.3023	6.4411	3.1387

$$Mij = 2 * \log_e(\tfrac{Ii}{Ij})$$

Table 9 Model selection criteria for correlated frailty models fitted to the kidney data set.

Model	PEN	GF	C_1	C_5	C_{10}	C_∞
Model I	8510.357	16.3653	8518.539	8523.994	8525.234	8526.722
Model II	10036.94	18.2679	10046.07	10052.16	10053.55	10055.21

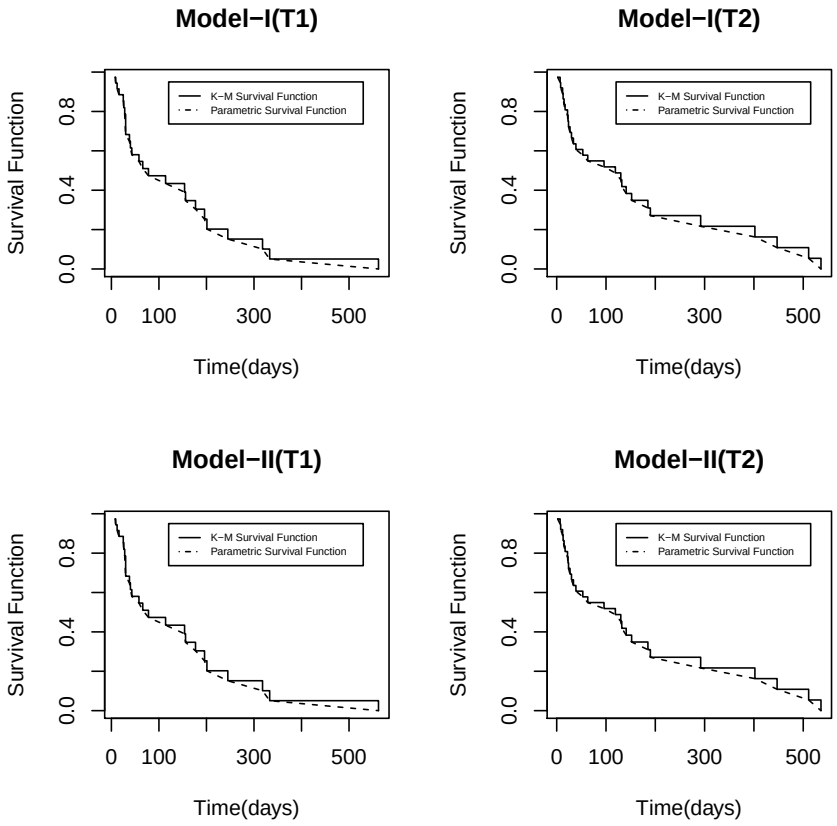

Figure 1 Survival function plots for (K-M survival and parametric survival)

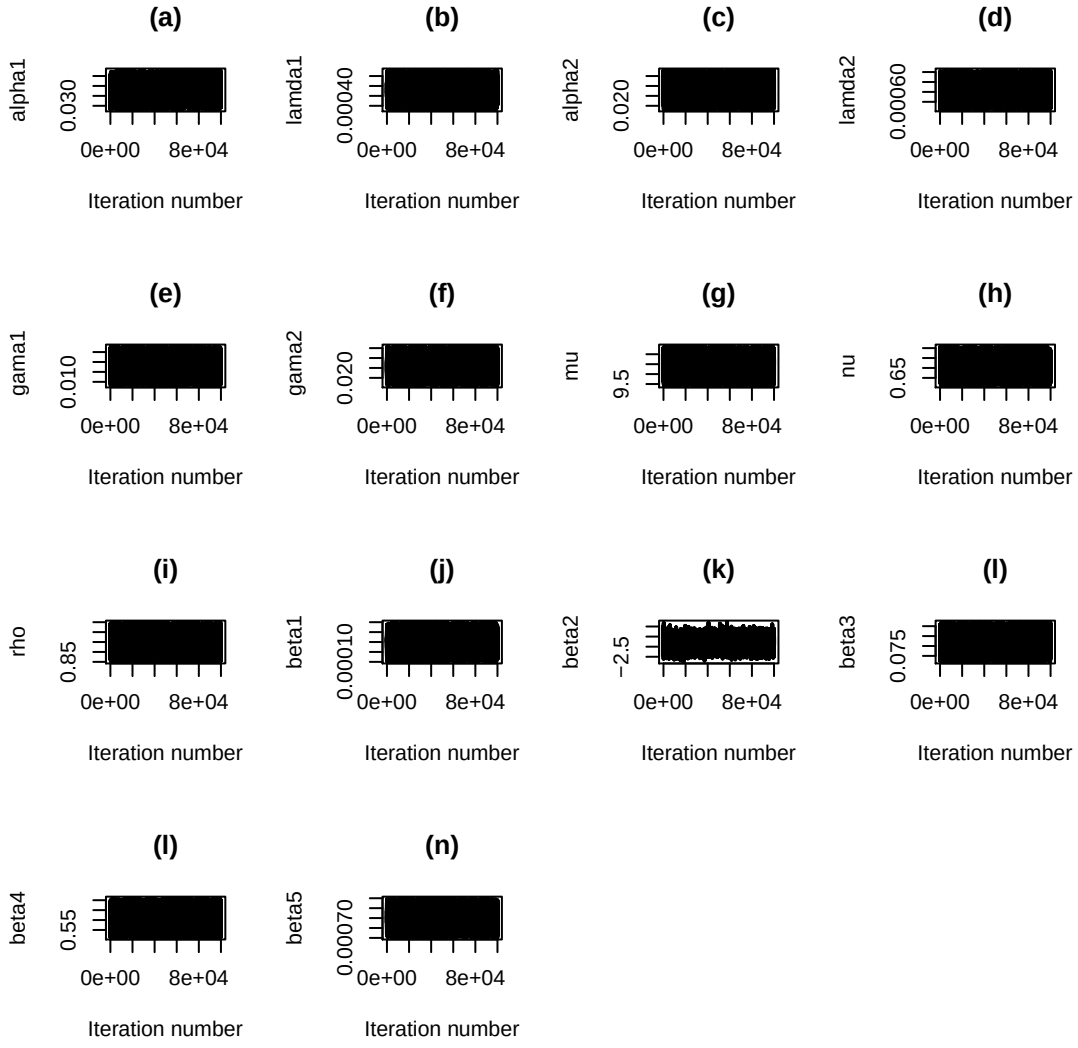

Figure 2 Trace plots for Model-I

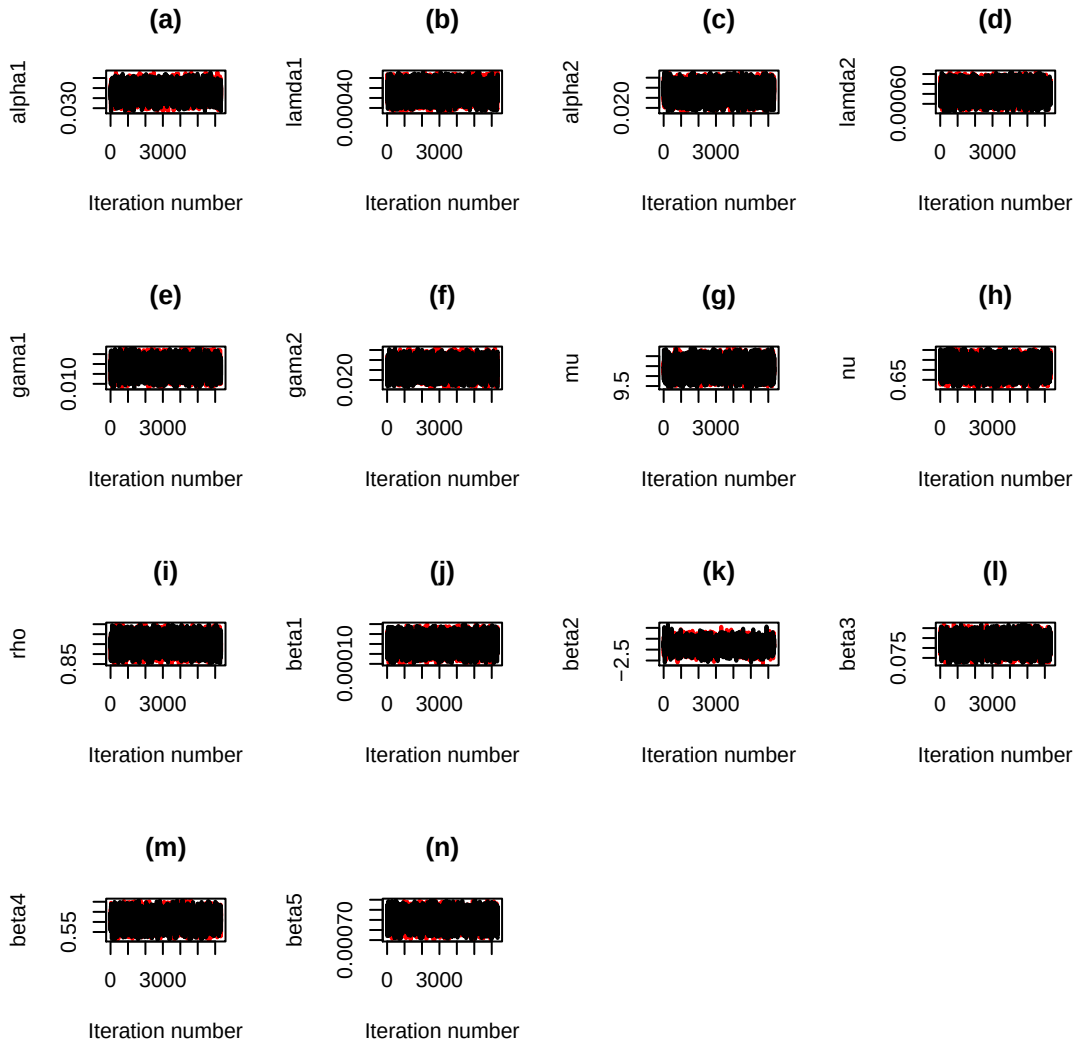

Figure 3 Coupling from the past plots for Model-I

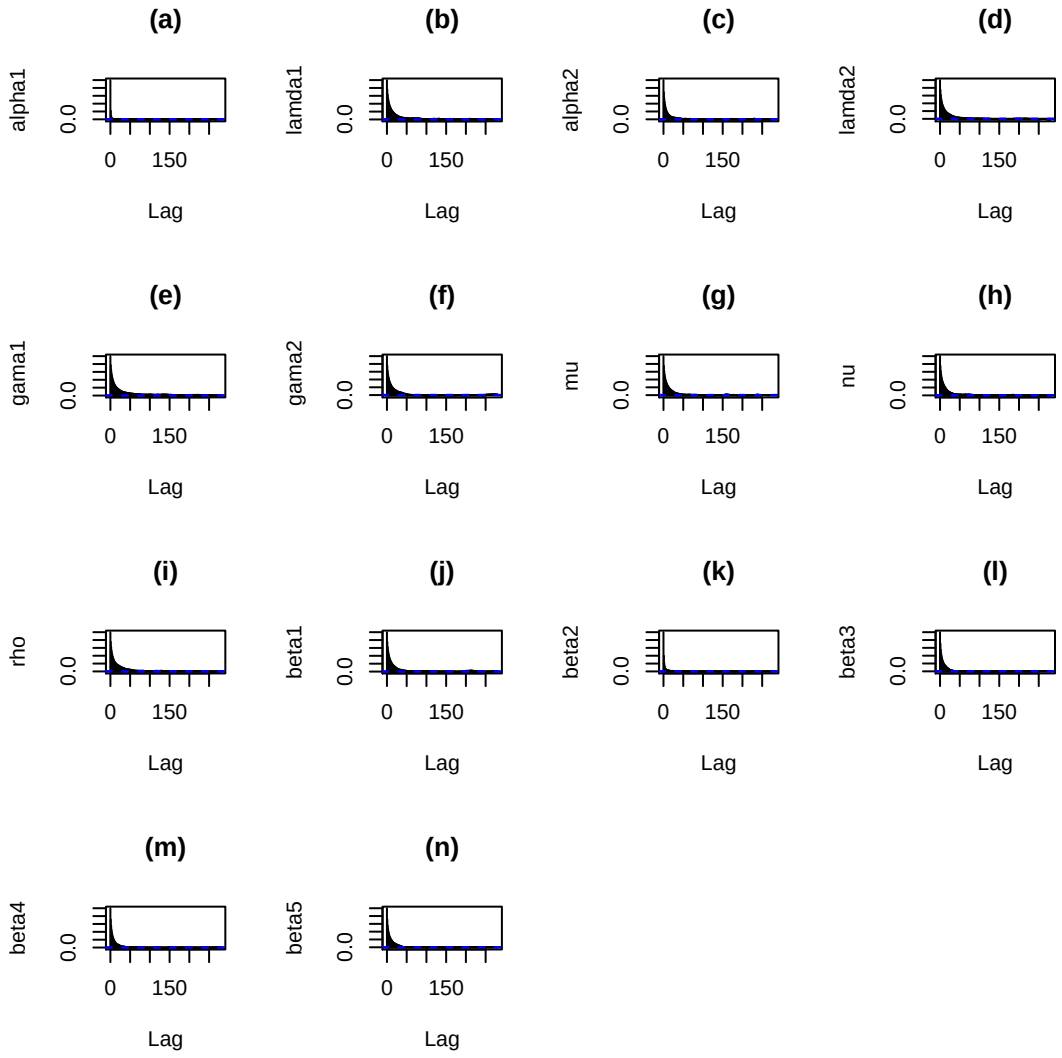

Figure 4 ACF plots for Model-I

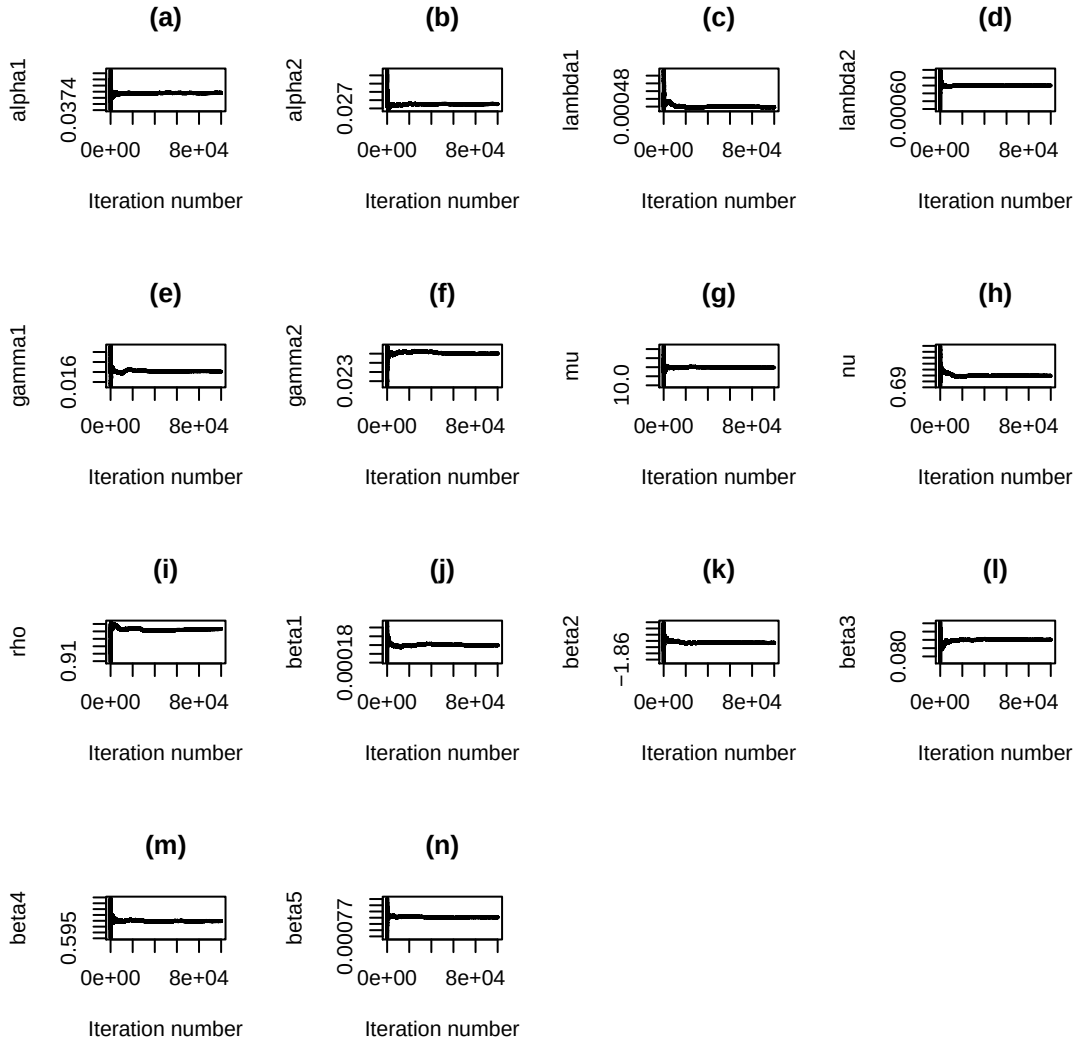

Figure 5 Running mean plots for Model-I