International Journal of Statistics and Management System
Vol. 4 No. 2 (July-December, 2019)

Correlated compound Poisson frailty model for bivariate survival data*

David D. Hanagal†

Received: 23rd July 2018 Revised: 14th December 2018 Accepted: 15th January 2019

Abstract

We propose Weibull model with heterogeneity (frailty or random effect) which is generated by correlated compound Poisson distribution with random scale for the bivariate survival data. There are some interesting situations like survival times in genetic epidemiology, dental implants of patients and twin births (both monozygotic and dizygotic) where genetic behavior (which is unknown and random) of patients follows a known frailty distribution. These are the situations which motivate to study this particular model. We propose maximum likelihood estimation procedure for the parameters in the proposed model.

1 Introduction. The shared gamma frailty models were suggested by Clayton [4] for the analysis of the correlation between clustered survival times in genetic epidemiology. An advantage is that without covariates its mathematical properties are convenient for estimation (see Oakes ([16, 17]). However, when adjusting for environment risk factors the analysis of the clustering is more difficult (see Parner [18]).

In a frailty model, it is absolutely necessary to be able to include some known explanatory variables to be able to estimate the aspects of the frailty distribution which represents the effect of unknown covariates. The reason is that the frailty describes the influence of

Key words and phrases: frailty model, bivariate survival data.

†*Mailing Address:* Department of Statistics, University of Pune, Pune 411007, India. *E–mail:* david_hanagal@yahoo.co.in

common unknown factors. If some covariates are included in the model, the variation owing to unknown covariates should be reduced.

For monozygotic twins, examples are gender and any other genetically based covariate. Both monozygotic and dizygotic twins share date of birth and common pre-birth environment. By measuring some potentially important covariates, we can examine the influence of the covariates, and we can examine whether they explain the dependence, that is, whether the frailty has no effect (or more correctly, no variation), when the covariate is included in the model.

In genetic studies where the outcome is the time to the event of interest, failure times among family members may not be independent. In this case, conventional survival analysis may yield consistent estimates of the marginal hazard if the marginal hazard is correctly modeled. However, variance estimates overestimate the true variance when the independent variables vary within a unit, and underestimate when the independent variables are constant within a unit, leading to incorrect inferences.

The natural parametric distribution to consider is the Weibull, because it allows for both the proportional hazard model and the accelerated failure time model. Hanagal [7, 8, 9, 10, 11] proposed bivariate Weibull regression models with gamma, positive stable, power variance function, lognormal frailties. Hanagal [12] proposed bivariate Weibull regression model with compound poisson frailty.

We propose compound Poisson frailty model for the bivariate survival life time. In Section 2, we introduce the compound Poisson distribution as a frailty and in Section 3, we introduce compound Poisson distribution with random scale. In Section 4, we present correlated compound Poisson frailty model for the bivariate life times. In Section 5, we give maximum likelihood estimate of the parameters. We present a simulation study in Section 6 and we conclude with discussions in Section 7.

2 Compound Poisson Frailty. Aalen [2, 3] introduced a compound Poisson distribution as a mixing distribution in survival models which is an extension of one studied by Hougaard [13]. The compound Poisson distribution plays a prominent role in this extension, being used here as a mixing distribution. Quite often hazard rates or intensities be raising at the start, reaching a maximum and then declining. Hence the intensity has an unimodal shape with finite mode. For example, 1) death rates for cancer patients, meaning that the longer the patient lives, beyond a certain times, the more improved are his or her

chances, 2) divorce rates, the maximal rate of divorce which occurs after a few years which means most marriages are going through crisis and then improving (Aaberge et al. [1]]. The population intensity starts to decline simply because the high-risk individuals have already died or been divorced, and so forth.

An additional feature which is often seen is that the total integral under the intensity(hazard rate) to be finite; that is, the distribution is defective. In practical terms this means that some individuals have zero susceptibility; they will "survive forever". For instance, some patients survive their cancer, some people never marry, some marriages are not prone to be dissolved, and so on. In medicine, there are several examples of diseases primarily attacking people with a particular susceptibility, for instance genetic kind, other people having virtually zero susceptibility of getting the disease. Another example is fertility. Some couples are unable to conceive children, so that the distribution of times to first child birth for a population of couples will be defective. In an unemployment data, one is also faced with the fact that some people may be completely unable to get a job.

The use of the compound Poisson distribution for Y is not only mathematically convenient, but might also be seen as natural in a more substantial sense. The distribution arises as a sum of a random number of independent gamma variables, where the number of terms in the sum is Poisson distributed. This might be viewed as a kind of shock model, where the vulnerability of the subject has been shaped by a random number of shocks, each of random size. The compound Poisson variable (Y) can be defined as follows.

$$Y = \begin{cases} X_1 + X_2 + \cdots + X_N, & N > 0 \\ 0, & N = 0 \end{cases} \tag{2.1}$$

where N is Poisson distributed with mean ρ, while $X_1, X_2, \ldots$ are independent and gamma distributed with scale parameter ν and shape parameter η. The distribution of Y 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

$$P(Y = 0) = \exp(-\rho), \tag{2.2}$$

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

$$f_Y(y; \eta, \nu, \rho) = \exp[-(\rho + \nu y)] \frac{1}{y} \sum_{n=1}^{\infty} \frac{\rho^n (\nu y)^{n\eta}}{\Gamma(n\eta) n!}. \tag{2.3}$$

The parameter set for the compound Poisson distribution is η, ν, $\rho > 0$. The expectation and variance are given by

$$E(Y) = \rho\eta/\nu \quad \text{and} \quad Var(Y) = \rho\eta(\eta+1)/\nu^2. \tag{2.4}$$

The Laplace transforms of the gamma and Poisson distributions are given by $L_X(s) = [\nu/(\nu+s)]^\eta$ and $L_N(s) = \exp(-\rho+\rho e^{-s})$, respectively. Now Laplace transform of compound Poisson distribution is

$$L_Y(s) = \exp\left\{-\rho + \rho\left(\frac{\nu}{\nu+s}\right)^\eta\right\}. \tag{2.5}$$

The survival function given the frailty Y is given by

$$S_{T|Y}(t|y) = \exp(-yM(t)) \tag{2.6}$$

where $M(t)$ is the integrated hazard of T. The unconditional survival function is given by

$$S_T(t) = \exp\left(-\rho\left\{1 - \left[\frac{\nu}{\nu+M(t)}\right]^\eta\right\}\right). \tag{2.7}$$

The hazard rate $\gamma(t)$ is given by

$$\gamma_T(t) = \frac{\rho\eta\nu^\eta m(t)}{[\nu+M(t)]^{\eta+1}}. \tag{2.8}$$

3 Compound Poisson distribution with random scale. An extension of the compound Poisson frailty model to family data, is to apply a probability distribution to the parameter ρ which was proposed by Moger and Aalen [15]. A probability density of the parameter ρ expresses the variation between families. The individuals of a given family are characterized by having a specific value of ρ, so they will have correlated frailties, while individuals from different families are independent. This yields a two level model, where the frailty has two components: A familial component, for instance relating to shared genes and environment, and an individual component, which could relate to exposure to individual environment. Thus the model does not fit into the traditional dichotomy of shared frailty models. We would like to stress the importance of frailty models having clear biological content, corresponding to understand a problem from a substance point of view, as opposed to just making mathematical assumptions. Since compound Poisson distribution

is included in the power variance function (PVF) distributions, this corresponds to randomizing a scale parameter in the PVF distributions. Hanagal [12] proposed compound Poisson frailty model when ρ is non-random for the bivariate Weibull baseline model. He also estimated the parameters in the model.

This paper will focus on densities for ρ which are included in the PVF distribution family. Specifically we consider the gamma, inverse Gaussian and positive stable distributions. As given in Hougaard [14], the distributions can be united in a three-parameter family with parameter set $\alpha \leq 1$, $\epsilon > 0$, with $\theta \geq 0$ for $\alpha > 0$, and $\theta > 0$ for $\alpha \leq 0$. For $\alpha = 0$ the gamma distributions are obtained. The inverse Gaussian distributions are obtained for $\alpha = 1/2$, and for $\theta = 0$ one gets the positive stable distributions. The positive stable distributions are absolutely continuous and nonnegative, with unimodal densities (Hougaard [13]. For $\alpha = 1$, a degenerate distribution is obtained, at ϵ, independent of θ. This corresponds to independence within families. This is given by (2.3), with $\eta = -\alpha$, $\rho = -(\epsilon/\alpha)\theta^{\alpha}$ and $\nu = \theta$. Hence, values of $\alpha < 0$ yield the compound Poisson distributions. The parametrization used in Section 2 is the most appropriate for $\alpha < 0$, while the parametrization by Hougaard is more easy to use for $\alpha > 0$. The expectation and variance of the distribution of ρ are

$$E(\rho) = \epsilon\theta^{\alpha-1} \quad \text{and} \quad Var(\rho) = \epsilon(1-\alpha)\theta^{\alpha-2}. \tag{3.1}$$

Thus, the positive stable distribution has no finite expectation or variance. The Laplace transform of ρ is given by

$$L_{\rho}(s) \quad = \exp\{-\tfrac{\epsilon}{\alpha}[(\theta+s)^{\alpha} - \theta^{\alpha}]\}. \tag{3.2}$$

In the case of the mixed compound Poisson distribution, the unconditional discrete part of Y is given by

$$P(Y=0) \quad = \quad E(\exp(-\rho)] = L_{\rho}(1). \tag{3.3}$$

The density of the unconditional continuous part of Y can be calculated in a similar manner by noting that

$$E[\rho^{n}\exp(-\rho)] \quad = \quad (-1)^{n} L_{\rho}^{(n)}(1), \tag{3.4}$$

where $L^{(n)}(s)$ denotes the n-th derivative of the Laplace transform. By inserting the density

of ρ into (2.3) and integrating out ρ, the density of Y may be put on the following form:

$$h_Y(y;\eta,\nu,\alpha,\theta,\epsilon) = \exp(-\nu y)\frac{1}{y}\sum_{n=1}^{\infty}\frac{(\nu y)^{n\eta}}{\Gamma(n\eta)n!}(-1)^n L_\rho^{(n)}(1). \tag{3.5}$$

The derivatives of the Laplace transform for the power variance function distribution for ρ are of the form

$$L_\rho^{(n)}(s) = (-1)^n L_\rho(s)\sum_{j=1}^{n} c_{n,j}(\alpha)\epsilon^j(\theta+s)^{j\alpha-n}, \tag{3.6}$$

as shown in Hougaard [14]. The coefficients $c_{n,j}(\alpha)$ are given by the recursive formula

$$\begin{aligned} c_{n,1}(\alpha) &= \Gamma(n-\alpha)/\Gamma(1-\alpha), \qquad c_{n,n}(\alpha)=1, \\ C_{n,j}(\alpha) &= C_{n-1,j-1}(\alpha) + c_{n-1,j}(\alpha)[(n-1)-j\alpha]. \end{aligned}$$

The Laplace transform of ρ, $L_\rho(s)$, combined with (2.5) yield the expression

$$L_Y(s) = L_\rho\left(1-\left(\frac{\nu}{\nu+s}\right)^\eta\right) \tag{3.7}$$

for the Laplace transform of Y. For the PVF distributed ρ, this equals

$$L_Y(s) = \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta+1-\left(\frac{\nu}{\nu+s}\right)^\eta\right]^\alpha-\theta^\alpha\right\}\right) \quad \text{if} \quad \alpha\le 1, \alpha\neq 0, \tag{3.8}$$

$$L_Y(s) = \left(\frac{\theta}{\theta+1-\left(\frac{\nu}{\nu+s}\right)^\eta}\right)^\epsilon \quad \text{if} \quad \alpha=0. \tag{3.9}$$

The Laplace transform of the gamma mixture distribution ($\alpha=0$) is obtained by taking the limit of the general Laplace transform. The positive stable mixture distribution ($\theta=0$) gives some nice properties when used as a frailty distribution. The Laplace transform of Y in this case is

$$L_Y(s) = \exp\left\{-\frac{\epsilon}{\alpha}\left[1-\left(\frac{\nu}{\nu+s}\right)^\eta\right]^\alpha\right\}. \tag{3.10}$$

Apart from the exponent α, this is of the same form as the Laplace transform of a compound Poisson distribution given in Equation (2.5), with ϵ/α playing the role of ρ.

Now the survival function of the lifetime T is given by

$$\begin{aligned} S_T(t) &= L_Y(M(t)) \\ &= \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta+1-\left(\frac{\nu}{\nu+M(t)}\right)^\eta\right]^\alpha-\theta^\alpha\right\}\right) \quad \text{if} \quad \alpha\le 1, \alpha\neq 0, \\ S_T(t) &= \left(\frac{\theta}{\theta+1-\left(\frac{\nu}{\nu+M(t)}\right)^\eta}\right)^\epsilon \quad \text{if} \quad \alpha=0. \end{aligned}$$

The mean and variance of Y can easily be found by means of (2.4), by noting that $E(Y) = E(E(Y|\rho))$ and that $\mathrm{Var}(Y) = \mathrm{Var}[\mathrm{E}(Y|\rho)] + \mathrm{E}[\mathrm{Var}(Y|\rho)]$:

$$\beta = E(Y) = \frac{\epsilon\eta}{\nu\theta^{1-\alpha}} \quad \text{and} \quad Var(Y) = \frac{\epsilon\eta[\theta + \eta(1-\alpha+\theta)]}{\nu^2\theta^{2-\alpha}}. \tag{3.11}$$

Note that when a positive stable mixture distribution is used, the frailty distribution Y has no finite expectation or variance. When ρ is not stable distributed ($\theta > 0$), the Laplace transform in (3.8) can be reparameterized by using the expectation β from (3.11) and the squared coefficient of variation

$$d \;=\; \frac{Var(Y)}{E(Y)^2} = \frac{[\theta + \eta(1-\alpha+\theta)]}{\theta^{\alpha}\epsilon\eta}$$

as new parameters. The value $d = 0$ corresponds to no heterogeneity.

4 Correlated compound Poisson frailty for the bivariate survival lifetimes. Moger and Aalen [15] developed correlated compound Poisson frailty model for the survival time of the two individuals in a family. Let Y_1 and Y_2 be the frailty variables of two individuals in a family with joint distribution $h_{Y_1,Y_2}(y_1, y_2)$. Let their marginal distribution given ρ, $f_{Y_1}(y_1|\rho)$ and $f_{Y_2}(y_2|\rho)$, be independent identically distributed compound Poisson with parameters η and ν. The parameter ρ, which is common for both Y_1 and Y_2, is assumed to be PVF distributed with parameters α, ϵ and θ. The joint discrete part of (Y_1, Y_2) is

$$P(Y_1 = 0, Y_2 = 0) \;=\; E_\rho[\exp(-2\rho)] = L_\rho(2).$$

The joint density of the continuous part of the distribution can be found analogously to (3.5), by using the ρ, in the distributions $f_{Y_1|\rho}(y_1|\rho)$ and $f_{Y_2|\rho}(y_2|\rho)$ to get derivatives of $L_\rho(s)$. It is given by

$$\begin{aligned} h_{Y_1,Y_2}(y_1, y_2; \eta, \nu, \alpha, \theta, \epsilon) \;=\; & \frac{1}{y_1 y_2}\exp[-\nu(y_1+y_2)]\sum_{n=2}^{\infty}(-1)^n L_\rho^{(n)}(2) \\ & \cdot\sum_{k=1}^{n-1}\frac{n u^{n\eta} y_1^{(n-k)\eta} y_2^{k\eta}}{\Gamma((n-k)\eta)\Gamma(k\eta)(n-k)!k!}, \end{aligned}$$

where $L_\rho^{(n)}(s)$ is defined in (3.6). Since the marginal distributions are compound Poisson given ρ, the joint distribution has an interesting feature: it is possible to have two related

individuals where one has zero frailty and the other has a positive frailty. The probability is given by

$$P(Y_1 = 0, Y_2 > 0) \;=\; E_\rho[\exp(-\rho) - \exp(-2\rho)] = L_\rho(1) - L_\rho(2).$$

In some situations, this is an aspect that may make the model fit better than a shared frailty model. Also, it is interesting for the interpretation. For instance, testicular cancer is hypothesized to be caused by some sort of damage in foetal life (Henderson et al., 1988). This damage could be due to genetics, mothers or pregnancies. If there is a mother effect, it may not be natural with the possibility of $Y_1 = 0$ and $Y_2 > 0$.

By using (3.7), one easily finds their joint Laplace transform

$$L_{Y_1,Y_2}(s,t) \;=\; L_\rho\left(2 - \left(\frac{\nu}{\nu+s}\right)^\eta - \left(\frac{\nu}{\nu+t}\right)^\eta\right),$$

which in the case of PVF distributed ρ is

$$L_{Y_1,Y_2}(s,t) = \begin{cases} \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta + 2 - \left(\frac{\nu}{\nu+s}\right)^\eta - \left(\frac{\nu}{\nu+t}\right)^\eta\right]^\alpha - \theta^\alpha\right\}\right) & \text{if} \quad \alpha \le 1, \alpha \neq 0, \\ \left[\frac{\theta}{\theta+2-\left(\frac{\nu}{\nu+s}\right)^\eta-\left(\frac{\nu}{\nu+t}\right)^\eta}\right]^\epsilon & \text{if} \quad \alpha = 0. \end{cases} \tag{4.1}$$

Note that the univariate Laplace transform in (3.8) appears by setting $t = 0$.

By noting that $Cov(Y_1, Y_2) = COV(E(Y_1|\rho), E(Y_2|\rho))$ and using (2.4), the correlation coefficient between frailties of two individuals in a family obtained by Moger and Aalen [15] is

$$Corr(Y_1, Y_2) \;=\; \frac{\eta(1-\alpha)}{\theta + \eta(1-\alpha+\theta)} \quad \text{if} \quad \theta > 0. \tag{4.2}$$

The parameter θ determines the degree of correlation. Since none of the moments exist (when $\theta = 0$), the correlation coefficient cannot be used as a measure of dependence for the compound Poisson-positive stable distribution. For values of θ close to zero, the correlation between two related individuals is approaching one. It is evident that the correlation has to be larger than zero, so the model can not handle negative dependencies.

Let T_1 and T_2 be the lifetimes of the two individuals which are independent. The survival function of (T_1, T_2) given the two dependent frailties (Y_1, Y_2) is given by

$$S_{T_1,T_2|Y_1,Y_2}(t_1, t_2|y_1, y_2) \;=\; \exp(-M_1(t_1)y_1 - M_2(t_2)y_2). \tag{4.3}$$

The unconditional survival function of (T_1, T_2) is obtained by integrating (Y_1, Y_2) out

$$\begin{aligned} S_{T_1,T_2}(t_1,t_2) &= E(\exp(-M_1(t_1)Y_1 - M_2(t_2)Y_2)) \\ &= \begin{cases} \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta+2-\left(\frac{\nu}{\nu+M_1(t_1)}\right)^{\eta}-\left(\frac{\nu}{\nu+M_2(t_2)}\right)^{\eta}\right]^{\alpha}-\theta^{\alpha}\right\}\right) \\ \qquad\qquad \text{if} \quad \alpha \le 1, \alpha \neq 0, \\ \left[\frac{\theta}{\theta+2-\left(\frac{\nu}{\nu+M_1(t_1)}\right)^{\eta}-\left(\frac{\nu}{\nu+M_2(t_2)}\right)^{\eta}}\right]^{\epsilon} \quad \text{if} \quad \alpha = 0. \end{cases} \end{aligned} \tag{4.4}$$

Let (T_1, T_2) are independent Weibull distributions with $\mathrm{W}(\lambda_1, c_1)$ and $\mathrm{W}(\lambda_2, c_2)$ respectively, where λ_i's scale parameters and c_i's are shape parameters of Weibull distributions. The survival function of T_i is

$$S_{T_i}(t_i) = \exp(-\lambda_i t_i^{c_i}). \tag{4.5}$$

Now the unconditional survival function of (T_1, T_2) with correlated compound Poisson frailties is given by

$$S_{T_1,T_2}(t_1,t_2) = \begin{cases} \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta+2-\left(\frac{\nu}{\nu+\lambda_1 t_1^{c_1}}\right)^{\eta}-\left(\frac{\nu}{\nu+\lambda_2 t_2^{c_2}}\right)^{\eta}\right]^{\alpha}-\theta^{\alpha}\right\}\right) \\ \qquad\qquad \text{if} \quad \alpha \le 1, \alpha \neq 0, \\ \left[\frac{\theta}{\theta+2-\left(\frac{\nu}{\nu+\lambda_1 t_1^{c_1}}\right)^{\eta}-\left(\frac{\nu}{\nu+\lambda_2 t_2^{c_2}}\right)^{\eta}}\right]^{\epsilon} \quad \text{if} \quad \alpha = 0 \end{cases} \tag{4.6}$$

In order to solve the identifiability problem, we assume a mean of 1 for the frailty distributions. For the gamma distribution, this can be achieved by setting $\theta = \epsilon$. In the shared PVF model, $E(Y) = 1$ is achieved by setting $\epsilon = \theta^{1-\alpha}$. The shared frailty models are compared to a compound Poisson model where ρ is gamma distributed, yielding a compound Poisson-gamma model. To secure a unit mean for the frailty, we get $\epsilon = \nu\theta/\eta$. In this paper, we assume the distribution of frailty as compound Poisson-gamma (with $\alpha = 0$) distribution for the bivariate survival data. The bivariate survival function based on this frailty is given by

$$S(t_1,t_2) = \left[\frac{\theta}{\theta+2-\left(\frac{\nu}{\nu+\lambda_1 t_1^{c_1}}\right)^{\eta}-\left(\frac{\nu}{\nu+\lambda_2 t_2^{c_2}}\right)^{\eta}}\right]^{\nu\theta/\eta}. \tag{4.7}$$

5 Estimation of Parameters. For the bivariate life time distribution, we use univariate censoring scheme given by Hanagal [5, 6] because the individuals do not enter at the same time and withdrawal or death of an individual or termination of the study will censor both life times of the components. Here the censoring time is independent of the life times of both components. This is the standard univariate right censoring for both failure times T_1 and T_2.

Suppose that there are n independent pairs of components under study and i-th pair of the components have life times (t_{1i}, t_{2i}) and a censoring time (w_i). The life times associated with i-th pair of the components are given by

$$\begin{aligned}(T_{1i}, T_{2i}) &= \{t_{1i}, t_{2i}\}, && \max(t_{1i}, t_{2i}) < w_i \\ &= \{t_{1i}, w_i\}, && t_{1i} < w_i < t_{2i} \\ &= \{w_i, t_{2i}\}, && t_{2i} < w_i < t_{1i} \\ &= \{w_i, w_i\}, && z_i < \min(t_{1i}, t_{2i}). \end{aligned} \tag{5.1}$$

Discarding factors which do not contain any of the parameters, we want to estimate the parameters in the proposed model. Now the likelihood of the sample of size n is given by

$$L = (\prod_{i=1}^{n_1} f_{1i})(\prod_{i=1}^{n_2} f_{2i})(\prod_{i=1}^{n_3} f_{3i})(\prod_{i=1}^{n_4} \overline{F_i}) \tag{5.2}$$

where

$$\begin{aligned} f_{1i} &= \frac{\partial^2 S(t_{1i}, t_{2i})}{\partial t_{1i} \partial t_{2i}}, && 0 < t_{1i} < t_{2i} < w_i \\ f_{2i} &= -\frac{\partial S(t_{1i}, w_i)}{\partial t_{1i}}, && 0 < t_{1i} < w_i < t_{2i} \\ f_{3i} &= -\frac{\partial S(w_i, t_{2i})}{\partial t_{2i}}, && 0 < t_{2i} < w_i < t_{1i} \\ \overline{F_i} &= S(w_i, w_i), && 0 < w_i < min(t_{1i}, t_{2i}) \end{aligned}$$

n_1, n_2, n_3 and n_4 are the number of observations observed to fail in the range space corresponding to f_{1i}, f_{2i}, f_{3i} and $\overline{F_i}$, respectively. f_{1i} is the conditional pdf with respect to Lebesque measure in R^2 and f_{2i} and f_{3i} are the conditional pdf with respect to Lebesque measure in R^1 in their respective regions.

The likelihood equations can be obtained by taking first order partial derivatives of the log-likelihood with respect to the parameters and equating to zero. We use the Newton-

Raphson method to solve the MLE. The observed Fisher information matrix with appropriate second order partial derivatives of the log-likelihood with respect to the parameters is I_1. The inverse of the observed information matrix (I_1) is the observed variance-covariance matrix ($\hat{\mathbf{\Sigma}}_{\mathbf{11}} = I_1^{-1}$) of the MLEs $\underline{\hat{\gamma}}' = (\hat{\eta}, \hat{\theta}, \hat{\nu}, \hat{c}_1, \hat{c}_2, \hat{\lambda}_1, \hat{\lambda}_2)'$ of the parameters $\underline{\gamma}' = (\eta, \theta, \nu, c_1, c_2, \lambda_1, \lambda_2)'$.

We expect that $\sqrt{n}(\underline{\hat{\gamma}} - \underline{\gamma})$ has asymptotic multivariate normal distribution with mean vector zero and variance-covariance matrix $\mathbf{\Sigma}_{11}$, where $\mathbf{\Sigma}_{11}$ is 7×7 variance covariance matrix of $\underline{\hat{\gamma}}' = (\hat{\eta}, \hat{\theta}, \hat{\nu}, \hat{c}_1, \hat{c}_2, \hat{\lambda}_1, \hat{\lambda}_2)'$

6 Simulation study. We generate 1000 samples of sizes $n = 50$, 100, and 200 from BVW model and obtain MLEs of the parameters. We observed from the simulation study as in Table 1 that MLEs are very close to the known values of the parameters in the proposed model and standard errors(se) of the MLEs are very low. The empirical standard errors(ese) and se are very close.

7 Discussions: We have simulated 1000 samples each of size $n = 50$, 100, and 200. If I take smaller sample sizes for the simulation, there is a problem of convergence of estimates of the parameters in N-R procedure. In the survival data, one should remember that the number of failures is always lower than the sample size. In the simulation process, the percentage of censoring changes from sample to sample for fixed sample size. So the effective sample size for the parametric model is the number of failures. In our case, we have a frailty model with seven parameters under the base line model. The sample sizes 20 and 40 are very small for this model with seven parameters and censoring scheme and so we have taken sample sizes as $n = 50$, 100, and 200. The efficiency and convergence of estimators depend on three things as follows:

1) sample size,
2) percentage of censoring,
3) number of parameters in the model.

When the sample size is very small and it is highly censored and there are higher number of parameters in the model, the probability of convergence (in N-R procedure) of the estimates of the parameters is very low as compared to larger sample sizes [by law of large numbers(LLN)].

Table 1: MLEs of the Parameters in Correlated Compound Poisson Frailty under Weibull Baseline Model.

Parameters	η	θ	ν	c_1	c_2	λ_1	λ_2
values	0.4	0.5	0.6	0.7	0.8	0.9	1.0
	$n = 50$						
Est.	0.3886	0.5219	0.6192	0.6882	0.7875	0.9215	0.9865
se	0.0257	0.0228	0.0267	0.0278	0.0241	0.0226	0.0242
ese	0.0269	0.0242	0.0271	0.0292	0.0258	0.0240	0.0259
	$n = 100$						
Est.	0.4106	0.5089	0.6118	0.6919	0.7921	0.9170	0.9901
se	0.0211	0.0202	0.0241	0.0244	0.0198	0.0219	0.0214
ese	0.0224	0.0221	0.0256	0.0269	0.0211	0.0225	0.0216
	$n = 200$						
Est.	0.4036	0.5039	0.6075	0.6981	0.7966	0.9064	0.9971
se	0.0118	0.0120	0.0116	0.0113	0.0122	0.0115	0.0127
ese	0.0126	0.0132	0.0121	0.0123	0.0120	0.0128	0.0138
Parameters	η	θ	ν	c_1	c_2	λ_1	λ_2
values	0.6	0.7	0.8	0.9	1.1	1.2	1.3
	$n = 50$						
Est.	0.5815	0.7221	0.8186	0.8825	1.1224	1.1837	1.3219
se	0.0279	0.0268	0.0271	0.0256	0.0259	0.0244	0.0235
ese	0.0283	0.0276	0.0277	0.0269	0.0263	0.0251	0.0241
	$n = 100$						
Est.	0.5883	0.7129	0.8125	0.8899	1.1161	1.2148	1.3138
se	0.0215	0.0208	0.0226	0.0211	0.0197	0.0202	0.0196
ese	0.0227	0.0216	0.0235	0.0222	0.0204	0.0216	0.0203
	$n = 200$						
Est.	0.5921	0.7048	0.8042	0.8956	1.1026	1.1972	1.3023
se	0.0121	0.0127	0.0124	0.0114	0.0129	0.0111	0.0115
ese	0.0128	0.0135	0.0128	0.0121	0.0127	0.0116	0.0119

Acknowledgments. I thank Council of Scientific and Industrial Research, New Delhi for providing financial assistance to carry out this research work under the major research project number 25(0176)/09/EMR-II. I thank Associate Editor and the referees for the valuable suggestions.

References

[1] Aaberge, R.; Kravdal, O. and Wennemo, T. (1989). *Unobserved heterogeneity in models of marriage dissolution.* Discussion Paper 42, Central Bureau of Statistics Norway.

[2] Aallen, O.O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine.* **7** 1121–37.

[3] Aallen, O.O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *The Annals of Applied Probability.* **2** 951–972.

[4] Clayton, D.G. (1978). A model for association in bivariate life tables and its applications to epidemiological studies of familial tendency in chronic disease incidence. *Biometrika.* **65** 141–51.

[5] Hanagal, D.D. (1992a). Some inference results in bivariate exponential distributions based on censored samples. *Communications in Statistics, Theory-Methods.* **21** 1273–95.

[6] Hanagal, D.D. (1992b). Some inference results in modified Freund's bivariate exponential distribution. *Biometrical Journal.* **34** 745–56.

[7] Hanagal, D.D. (2005). A positive stable frailty regression model in bivariate survival data. *Journal of the Indian Society for Probability and Statistics.* **9** 35–44.

[8] Hanagal, D.D. (2006). Weibull extension of bivariate exponential regression model with gamma frailty for survival data. *Economic Quality Control.* **21** 165–174.

[9] Hanagal, D.D. (2007). A power variance function frailty regression model in bivariate survival data. *IAPQR Transactions.* **32(2)** 117–129.

[10] Hanagal, D.D. (2008). Modelling heterogeneity for bivariate survival data by the log-normal distribution. *Statistics and Probability Letters.* **78(9)** 1101–09.

[11] Hanagal, D.D. (2009a). Weibull extension of bivariate exponential regression model with different frailty distributions. *Statistical Papers.* **50** 29–49.

[12] Hanagal, D.D. (2009b). Modeling heterogeneity flor bivariate survival data by compound Poisson distribution. *Model Assisted Statistics and Applications.* **4** To appear.

[13] Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometika*. **73** 387-396.

[14] Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer-Verlag New York, Inc.

[15] Moger, T.A. and Aalen, O.O. (2005). A distribution for multivariate frailty based on the compound Poisson distribution with random scale. *Lifetime Data Analysis*. **11** 41–59.

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

[17] Oakes, D. (1986). Semiparametric inference in bivariate survival data. *Biometrika*. **73** 353–61.

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