A Comparison of Frailty and Other Models for Bivariate Survival Data

Sujit K. Sahu
Faculty of Mathematical Studies,
University of Southampton,
Highfield, Southampton, UK.

Dipak K. Dey
Department of Statistics,
University of Connecticut,
Storrs, CT, USA.

September 13, 1999

Summary

Multivariate survival data arise when each study subject may experience multiple events or when the study involves several members of each group. Statistical analyses of such data need to account for the intra-cluster dependence through appropriate modeling. Frailty models are the most popular for such failure time data. However, there are other approaches which model the dependence structure directly. In this article, we compare the frailty models for bivariate data with the models based on bivariate exponential and Weibull distributions. Bayesian methods provide a convenient paradigm for comparing the two sets of models we consider. Our techniques are illustrated using two examples. One simulated example demonstrates model choice methods developed in this paper and the other example, based on a practical data set of onset of blindness for diabetic Retinopathy patients, considers Bayesian inference using different models.

Keywords: Bivariate exponential distribution; Bivariate Weibull distribution; Frailty models; Markov chain Monte Carlo methods; Proportional hazards model.

1 Introduction

Bivariate survival data arise when each study subject experiences two events. Failure times of paired human organs, e.g. kidneys, eyes; and double recurrences of a given disease are particular examples of such data. In a different context the data may consist of time to diagnosis or hospitalization and the time to eventual death from a fatal disease. In industrial applications these data types may come from systems whose survival depend on the survival of two very similar components. For example, the breakdown times of dual generators in a power plant or failure times of twin engines in a 2-engine airplane are illustrations of bivariate survival data.

Modeling dependence between the components of bivariate survival data is the focus of the current paper. We consider the following two very different parametric approaches to modeling such data. However, there are non- or semi-parametric modeling strategies available. See e.g. Gustafson (1997) where hierarchical Bayesian models with the partial likelihood at the first stage of hierarchy have been used.

The first approach we consider is based on what are called the frailty models. These models are increasingly being used to model bivariate (and multivariate) survival data, see e.g. Clayton and Cuzick (1985), Clayton (1978) and Oakes (1989). The models are built using unobserved covariates for each data point. The event times are conditionally independent given the *frailty*, an individual random effect. This can be seen as a strategy to model the dependence in multivariate data using univariate distributions. In this paper we assume multiplicative gamma frailty distribution with flexible Weibull baseline hazards.

An alternative parametric approach to the above modeling strategy is to consider models which directly address the dependence present in the data. Data points may be dependent because of simultaneous failure and/or due to failure of one component changing the life distribution of the other. The family of bivariate exponential distributions, see e.g. Freund (1961), Marshall and Olkin (1967), Block and Basu (1974), Proschan and Sullo (1974) and references therein, provide basic model elements. The exponential models can be generalized to more flexible Weibull cases, see e.g. Lee (1979) and Lu and Bhattacharyya (1990). Proportional hazards components can be built into these models, see e.g. Klein *et al.* (1989) and Ghosh and Gelfand (1998). In this paper we further develop and extend these models. We generalize the bivariate Weibull distribution to accommodate the changed intensity of the remaining lifetime of the surviving component after the failure of the other in Section 3. We then introduce the proportional hazards components for incorporating covariate information.

Which of the above two classes of models is more appropriate for a given data set? The dependence structure is introduced in the frailty models by an indirect way. Analyses using these models investigate the relationships between the observed covariates (or risk factors) and the hazards conditionally on the unobserved frailties. The bivariate exponential models, however, provide direct interpretation of the dependence present in the data. As an example consider the loss of one human organ (out of a pair). As a result of this the other organ is required to bear extra load which may decrease its remaining lifetime on the average.

To further investigate the dependence structures introduced by the above two modeling strategies with a concrete example we consider a local dependence measure which is the ratio of the joint bivariate survival function and the product of the marginal univariate survival functions, see e.g. Joe (1997). Values of this close to 1 show local independence. The left panel in Figure 1 shows the dependence structure induced by the frailty model. The components become more locally independent when they are both high. The right panel plots the same for the Marshall-Olkin bivariate exponential model. The measure of local independence for this model increases when both the components increase since the dependence between the components is introduced by the possibility of simultaneous failure and this decreases as both the components increase. The Marshall-Olkin model shows more local independence than the frailty model.

Although both types of models are suitable for many bivariate data sets, the latter models can intuitively be superior because of their direct interpretation of the nature of dependence and background biological justification. However, it is generally not possible to decide between the two strategies without undertaking further statistical investigation.

We develop modern Bayesian model choice techniques to compare the two sets of models. Model comparison using the classical likelihood ratio tests is not possible here since the two sets of models are not 'nested'. Bayesian methodologies are more attractive in this regard. The parameter spaces for the two sets of models are different and in general they are not comparable. However, the predictive distributions induced by the models are comparable because these can be seen as the distributions of future replicates of the same data set. Different model choice criteria based on suitable loss functions in the predictive space are developed in this paper to facilitate model comparison. Bayesian computation techniques using Markov chain Monte Carlo (MCMC) methods are used to compute the different model choice criteria.

The remainder of this paper is organized as follows. In Section 2 we present the frailty models with Weibull baseline hazards functions and gamma frailties. Section 3 sets out the models based on bivariate exponential and Weibull models. In Section 4 we develop different Bayesian model choice criteria. Section 5 illustrates the methodology with two examples: one simulated and one real data set. Computational details are placed in an Appendix.

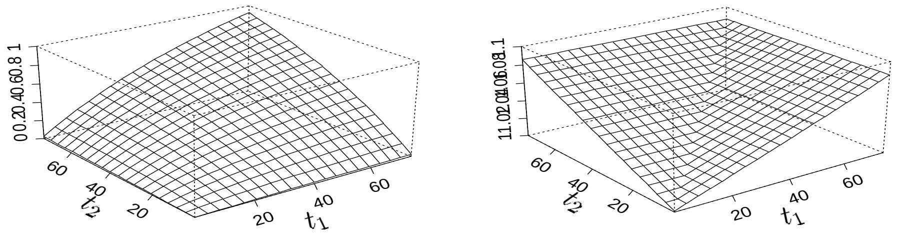

Figure 1: Local dependence structures induced by the frailty model (left panel) and the Marshall-Olkin model (right panel). The dependence measure, $d(t_1, t_2) = \frac{S(t_1,t_2)}{S(t_1)S(t_2)}$ where the function S is the survival function, is plotted for a grid of values of t_1 and t_2. The two components are locally independent if the height of the perspective plot is unity.

2 Frailty Models

Let t_{ij} denote the survival time of the jth component ($j = 1, 2$) for the ith subject ($i = 1, \ldots, n$). An individual level multiplicative effect $w_i > 0$, called the *frailty*, is associated with the ith subject. Given the unobserved w_i we assume a flexible Weibull baseline hazards function for t_{ij} (see e.g. Sahu *et al.*, 1997 and Sinha and Dey, 1997):

$$h(t_{ij}|\mathbf{z}_i, w_i) = \alpha t_{ij}^{\alpha-1}\mu \exp(\mathbf{z}_i^T\boldsymbol{\beta})w_i, \ \alpha, \mu > 0, \tag{1}$$

where $\boldsymbol{\beta}$ is the regression parameter and $\mathbf{z}_i$ is the covariate vector; α and μ are unknown parameters for which prior distributions are assumed below. For identifiability purposes the linear model $\mathbf{z}_i^T\boldsymbol{\beta}$ does not include any intercept term.

The frailty distribution w_i is taken as independent gamma distribution, (see e.g. Clayton, 1991) i.e.,

$$w_i \sim Gamma(\eta, \eta), \ i = 1, \ldots, n, \tag{2}$$

where η^{-1} is the unknown variance of w_i. (We write $X \sim Gamma(a, b)$ if its density is $\propto x^{a-1}\exp(-bx)$, for $x > 0$, $a > 0$, $b > 0$.)

We adopt a widely used conjugate prior for η, a Gamma distribution with mean 1 and large variance,

$Gamma(\phi, \phi)$ say with a small choice of ϕ. Components of $\boldsymbol{\beta}$ are assigned independent normal priors with large variances. Diffuse priors for μ and α are assumed in the final stage of hierarchical Bayesian model building. We assume $\mu \sim Gamma(\rho, \rho)$ and $\alpha \sim Gamma(\kappa, \kappa)$ for small values of ρ and κ for convenience. In our illustrative examples in Section 5 we take $\kappa = \rho = 0.1$. Although it is possible to use other suitable prior distributions if it is so desired, the Gamma prior distribution for μ is easy to use because in the Gibbs sampling implementation μ can be simulated from a suitable Gamma distribution.

Sahu *et al.* (1997) provide details on MCMC implementations of these models. However, these models can also be fitted using the general purpose Gibbs sampling software `BUGS`, publicly available from `http://www.mrc-bsu.cam.ac.uk/`.

Before concluding this section we obtain many important marginal survival functions and likelihood functions induced by the above conditionally independent hierarchical Bayesian model (1). The conditional hazards function $h(t_{ij}|\mathbf{z}_i, w_i)$ in (1) introduces the frailty parameter as a multiplicative random effect for the ith observation and in (2) a conjugate gamma distribution is assumed for the w_i. Due to this conjugacy the parameter w_i can be integrated out exactly. For example, the marginal bivariate survival function of t_{i1} and t_{i2}, denoted by $S(t_{i1}, t_{i2})$, is given by:

$$S(t_{i1}, t_{i2}) = \Gamma(\eta) \left\{ \eta + \mu \exp(\mathbf{z}_i^T \boldsymbol{\beta})(t_{i1}^\alpha + t_{i2}^\alpha) \right\}^{-\eta}, i = 1, \ldots, n, \quad (3)$$

where $\Gamma(\cdot)$ is the gamma function. In this way the marginal survival functions of the two components are also easy to obtain. The local dependence measure $d(t_1, t_2)$ defined in Figure 1 (without the data subscript i and without the covariates) is given by $\mu^{2\eta}(t_1 t_2)^{\eta\alpha}(\eta + \mu t_1^\alpha + \mu t_2^\alpha)^{-\eta}/\Gamma(\eta)$. The left panel of Figure 1 plots this dependence structure when $\eta = \alpha = 1$, $\mu = 0.03$ and $\boldsymbol{\beta} = 0$.

In what follows we also use the marginal bivariate density of the data points after integrating with respect to the frailty distribution (2). For future reference we write down the marginal likelihood in the presence of censoring. Let δ_{ij} denote the indicator variable taking value 1 if the jth component ($j = 1, 2$) of the ith ($i = 1, \ldots, n$) observation fails and value 0 otherwise. Let $\delta_{i+} = \sum_{j=1}^2 \delta_{ij}$. The marginal likelihood of $\eta, \mu, \alpha, \boldsymbol{\beta}$ is given by:

$$L(\eta, \mu, \alpha, \boldsymbol{\beta}; \mathbf{t}) = \prod_{i=1}^n \left[\Gamma(\eta + \delta_{i+}) \left\{ \eta + \mu \exp(\mathbf{z}_i^T \boldsymbol{\beta})(t_{i1}^\alpha + t_{i2}^\alpha) \right\}^{-\eta - \delta_{i+}} \prod_{j=1}^2 \left\{ \mu\alpha \exp(\mathbf{z}_i^T \boldsymbol{\beta}) t_{ij}^{\alpha - 1} \right\}^{\delta_{ij}} \right], \quad (4)$$

where $\mathbf{t}$ is the collection of n bivariate data points.

3 Bivariate Exponential and Weibull Models

3.1 Exponential Models

The bivariate exponential model due to Marshall and Olkin (1967) has the survival function of the form

$$S(t_{i1}, t_{i2}) = \exp\{-\theta_1 t_{i1} - \theta_2 t_{i2} - \theta_3 \max(t_{i1}, t_{i2})\}, \ \theta_j > 0, j = 1, 2, 3; i = 1, \ldots, n. \tag{5}$$

Some notable features of the above distribution are that: the marginal distribution of t_{ij} is exponential with parameter $\theta_j + \theta_3, j = 1, 2$; the distribution is singular for $\theta_3 > 0$ since it places mass along the line $t_{i1} = t_{i2}$; the components are independent if $\theta_3 = 0$. For this model $d(t_1, t_2) = \exp\{\theta_3 \min(t_1, t_2)\}$. See the right panel of Figure 1 for an illustration of this dependence structure when $\theta_3 = 0.0011$.

We consider an extension of the Marshall-Olkin (MO, henceforth) model (5). This is achieved by extending a distribution given by Freund (1961). Suppose that $\theta'_j + \theta_3$ is the changed intensity of component j after the failure of component $j' \neq j = 1, 2$ and $\theta = \theta_1 + \theta_2 + \theta_3$. The modified Freund (MF, henceforth) bivariate exponential distribution (see e.g. Proschan and Sullo, 1974; Hanagal, 1992) has the following pdf:

$$f(t_{i1}, t_{i2}) = \begin{cases} \theta_1(\theta'_2 + \theta_3) \exp\{-(\theta'_2 + \theta_3)(t_{i2} - t_{i1}) - \theta t_{i1}\}, \ t_{i1} < t_{i2} \\ \theta_2(\theta'_1 + \theta_3) \exp\{-(\theta'_1 + \theta_3)(t_{i1} - t_{i2}) - \theta t_{i2}\}, \ t_{i1} > t_{i2} \\ \theta_3 \exp\{-\theta t\}, \ t_{i1} = t_{i2} = t. \end{cases} \tag{6}$$

Properties of the above distribution are as follows. The distribution reduces to the Freund (1961) model under $\theta_3 = 0$; the MO model (5) is obtained by setting $\theta'_j = \theta_j, j = 1, 2$. The marginal distributions are not exponentials except for the above two cases. The distribution is symmetric (in the components) if $\theta_1 = \theta_2$ and $\theta'_1 = \theta'_2$. A schematic diagram for this model is given in Figure 2.

3.2 Weibull Models

Extensions of the above exponential distributions to more general cases, e.g. the Weibull's, have been considered by many authors, see e.g. Lee (1979), Lu (1989) and Lu and Bhattacharyya (1990). We consider the following MO type of model corresponding to (5) as mentioned in Lee (1979). The distribution has the survival function given by

$$S(t_{i1}, t_{i2}) = \exp\{-\theta_1 t_{i1}^{\alpha_1} - \theta_2 t_{i2}^{\alpha_2} - \theta_3 [\max(t_{i1}, t_{i2})]^{\alpha_3}\}, \ \theta_j, \alpha_j > 0, j = 1, 2, 3. \tag{7}$$

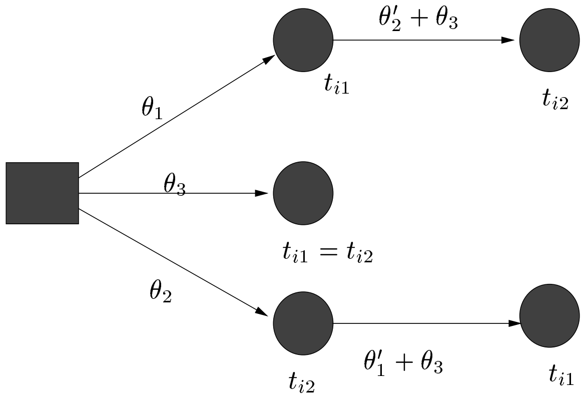

Figure 2: Diagram for the bivariate exponential distribution (6). The MO model (5) is obtained by setting $\theta_j' = \theta_j, j = 1, 2$.

Though the marginals of the above distribution are not in general Weibulls, it has meaningful interpretations in terms of the non-homogeneous Poisson processes. The distribution can be obtained by considering three independent Weibull random variables $u_{ij}, j = 1, 2, 3$ with parameters (α_j, θ_j) respectively and setting $t_{ij} = \min(u_{ij}, u_{i3}), j = 1, 2$. Note that this is distinct from the distribution obtained by means of power transformations of the marginals of the MO model (5) which is given by

$$S(t_{i1}, t_{i2}) = \exp\left\{-\theta_1 t_{i1}^{\alpha_1} - \theta_2 t_{i2}^{\alpha_2} - \theta_3 \max(t_{i1}^{\alpha_1}, t_{i2}^{\alpha_2})\right\},\ \theta_j > 0, j = 1, 2, 3.$$

The Weibull distribution similar to (6) can be found in Lu (1989) without the singularity part. We extend this distribution to accommodate the singularity part, i.e., failure at the same time. Let $\theta_j(t) = \theta_j \alpha_j t^{\alpha_j - 1}, j = 1, 2, 3$ and $\Theta_j(t) = \int_0^t \theta_j(u) du = \theta_j t^{\alpha_j}, j = 1, 2, 3$ and $\Theta(t) = \Theta_1(t) + \Theta_2(t) + \Theta_3(t)$. Further let,

$$\begin{aligned} \theta_j''(t_{ij}|t_{ij'}) &= \theta_j' \alpha_j' (t_{ij} - t_{ij'})^{\alpha_j' - 1} + \theta_3 \alpha_3 t_{ij}^{\alpha_3 - 1},\ \theta_j', \alpha_j' > 0, j \neq j' = 1, 2; \\ \Theta_j''(t_{ij}|t_{ij'}) &= \theta_j'(t_{ij} - t_{ij'})^{\alpha_j'} + \theta_3 t_{ij}^{\alpha_3},\ j \neq j' = 1, 2. \end{aligned}$$

Our extended distribution has the following pdf:

$$f(t_{i1}, t_{i2}) = \begin{cases} \theta_1(t_{i1}) \theta_2''(t_{i2}|t_{i1}) \exp\left\{-\Theta_2''(t_{i2}|t_{i1}) - \Theta_1(t_{i1}) - \Theta_2(t_{i1})\right\},\ t_{i1} < t_{i2} \\ \theta_2(t_{i2}) \theta_1''(t_{i1}|t_{i2}) \exp\left\{-\Theta_1''(t_{i1}|t_{i2}) - \Theta_1(t_{i2}) - \Theta_2(t_{i2})\right\},\ t_{i1} > t_{i2} \\ \theta_3(t) \exp\left\{-\Theta(t)\right\},\ t_{i1} = t_{i2} = t. \end{cases} \tag{8}$$

It is easily seen that the above is a valid bivariate probability density function which is non-negative and integrates to 1. This distribution reduces to the MF bivariate exponential case (6) if we set all the α's to unity. Unfortunately, if we set, $\theta_j' = \theta_j$ and $\alpha_j' = \alpha_j$ for $j = 1, 2$, we do not get the generalized MO model (7) except for the exponential case.

However, note a very general feature of the distribution (8). The distribution is essentially of the form:

intensity $\times$ exp($-$ cumulative intensity). In this sense, we can work with any form of the intensities $\theta(t)$ to accommodate different distributional assumptions. We have not seen the distribution (8) anywhere in the literature in this generality, though the intensities presented in Ghosh and Gelfand (1998) lead to this distribution in special cases.

3.3 Parametric Proportional Hazards Components

Parametric proportional hazards (PH) models provide a framework for incorporating covariate information. We extend the bivariate distributions presented above to include covariates, following, e.g., Klein *et al.* (1989).

The PH components for the MO models, (5) and (7), can be assumed as:

$$\theta_j = \exp\left\{\xi_j + \mathbf{z}_i^T \boldsymbol{\beta}_j\right\}, j = 1, 2, 3, i = 1, \ldots, n, \tag{9}$$

where $\exp(\xi_j)$ is the new baseline intensity. We use the notation $\lambda = \exp(\xi)$ to denote the intensity henceforth in this paper. We use the dual notations because computations are easier done in the log scale (ξ) and interpretations are easier with the intensities (λ).

For the MF models (6) and (8), note that θ'_j inferentially depends on the excess lifetime of component j. Hence it is reasonable to assume different PH models on them. This leads to assuming

$$\theta'_j = \exp\left\{\xi'_j + \mathbf{z}_i^T \boldsymbol{\beta}'_j\right\}, j = 1, 2,\ i = 1, \ldots, n. \tag{10}$$

The PH models (9) and (10) describe the behavior of the predictors on five different domains in Figure 2. Specifically, $\boldsymbol{\beta}_1$ shows the relationship of the predictors $\mathbf{z}$ with the marginal hazards for the first component in the domain where the first component is the first to fail. The other $\boldsymbol{\beta}_j$'s have similar interpretations. See Section 5.2.2 for further explanation using an example.

The Bayesian model in each case is completed by assuming suitable prior distributions for different parameters. All prior distributions are assumed to be diffuse for all the models considered in this paper so that inference is not driven by the prior. The components of $\boldsymbol{\beta}$ and the parameters ξ_j and ξ'_j are given independent flat normal prior distributions with mean 0 and large variance (10^4). In the Weibull cases we assume $Gamma(\kappa, \kappa)$ distribution for the α_j as in Section 2. The hyper-parameter κ is set at 0.1, although see the second example in Section 5 where we experiment with a prior distribution which restricts α's in a finite interval to study the sensitivity of the prior. We also note that an advantage of the MCMC implementation methods is the ability to incorporate other suitable prior distributions if it is so desired.

The full Bayesian MO models (5) and (7) with the PH components (9) are labelled by EMO and WMO respectively. The MF models (6) and (8) with the PH components (9) and (10) and the above prior assumptions are henceforth denoted by EMF and WMF respectively. See the appendix for the computational details needed for MCMC implementations of these models.

4 Model Choice

4.1 Bayesian Information Criterion

Often the Bayes factors (see e.g. Kass and Raftery, 1995) provide model comparison. Key *et al.*, (1999) discuss the relationships between the Bayes factor and other well-known model choice criteria, e.g. the Akaike information criteria (AIC) and the Bayesian information criteria (BIC), (Schwartz, 1978). They show that the approximate Bayes factors under different prior assumptions are the various model choice criteria.

Here we consider the BIC as our first model choice criterion since this provides a quick and easy method to compare different models. It is defined as

$$\text{BIC} = -2(\text{maximized log likelihood}) + p\log(n), \tag{11}$$

where n is the number of sample points and p is the number of estimated parameters. This takes sample size into account and support model parsimony. Its first term encourages a more complex model while the second term supports simpler models with fewer parameters.

Although the frailty model has one frailty parameter for each bivariate observation, we do not count these as parameters to make a fair comparison with the models described in Section 3. Also with the conjugate distributional assumption on the frailties, (made in Section 2) these parameters can be integrated out exactly. Indeed for our examples in Section 5 we take the marginal likelihood given by (4) to compute the BIC for the frailty models.

With an MCMC implementation we can not directly calculate the required maximized log likelihood to compute the BIC. However, the average value of the log likelihood function evaluated at each MCMC iteration can be used instead as an approximation. We adopt this technique to compute the BIC for all the models considered in this paper. Although this approach may not be totally satisfactory from a purist viewpoint, we use this to illustrate the above mentioned point that a more complex model will decrease the likelihood. This also helps us to understand more about the model choice criteria considered next.

4.2 Predictive Model Selection Using Loss Functions

Model diagnostic and comparison measures based on the posterior predictive densities are often easier to work with in MCMC model fitting settings. MCMC methods are able to produce these measures without much extra work, see e.g. Gelfand (1996). In the discussion below we develop one criterion to perform model choice.

Let $\mathbf{t}_{\text{obs}}$ with components $\mathbf{t}_{i,\text{obs}}, i = 1, \ldots, n$ denote the set of observed values of the survival times. Similarly we use the notation $\mathbf{t}_{\text{rep}}$ with components $\mathbf{t}_{i,\text{rep}}$ to denote a future set of observations under the assumed model. (Here obs and rep are the abbreviations for the observation and replicate respectively.) Let $\boldsymbol{\theta}$ denote the set of parameters of the current model.

The posterior predictive density, $\pi(\mathbf{t}_{\text{rep}}|\mathbf{t}_{\text{obs}})$, is the predictive density of a new independent set of observables, $\mathbf{t}_{\text{rep}}$ under the model, given the actual set of observables, $\mathbf{t}_{\text{obs}}$. By marginalizing $\pi(\mathbf{t}_{\text{rep}}|\mathbf{t}_{\text{obs}})$ we obtain the posterior predictive density of one observation $\mathbf{t}_{i,\text{rep}},\ i = 1, \ldots, n$, as follows,

$$\pi(\mathbf{t}_{i,\text{rep}}|\mathbf{t}_{\text{obs}}) = \int \pi(\mathbf{t}_{i,\text{rep}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{t}_{\text{obs}})d\boldsymbol{\theta}. \tag{12}$$

Let $\boldsymbol{\mu}_i$ and Σ_i denote the posterior predictive mean and covariance of $\mathbf{t}_{i,\text{rep}}$ under the density (12). We can easily estimate $\boldsymbol{\mu}_i$ and Σ_i by Monte Carlo integration as follows. Suppose that $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(B)}$ denote B Gibbs sampled values from $\pi(\boldsymbol{\theta}|\mathbf{t}_{\text{obs}})$. Then, a random sample $\mathbf{t}_i^{(r)}$ drawn from $\pi(\mathbf{t}_i|\boldsymbol{\theta}^{(r)})$, is a sample from the above predictive density, see e.g. Gelfand (1996).

To perform model choice, we consider the squared error loss function, following Laud and Ibrahim (1995),

$$L = (\mathbf{t}_{\text{rep}} - \mathbf{t}_{\text{obs}})^T(\mathbf{t}_{\text{rep}} - \mathbf{t}_{\text{obs}}). \tag{13}$$

The expected loss function, where expectation is taken with respect to the posterior predictive distribution $\pi(\mathbf{t}_{\text{rep}}|\mathbf{t}_{\text{obs}})$, serves as a model choice criterion. The criterion

$$D' = \sum_{i=1}^{n} tr(\Sigma_i) + \sum_{i=1}^{n} \left(\boldsymbol{\mu}_i - \mathbf{t}_{i,\text{obs}}\right)^T \left(\boldsymbol{\mu}_i - \mathbf{t}_{i,\text{obs}}\right), \tag{14}$$

emerges as a form partitioned into two parts. The first is a penalty term, denoted by P, which penalizes both under-fitted and over-fitted models, since the predictive variances in such cases will tend to be larger. The second term is a goodness-of-fit measure denoted by G. It may be expected that G should be smaller for a more complex model. However, it is not guaranteed that this should be the case as it is not the classical likelihood ratio statistic. It is the squared distance between the fitted and the observed survival times.

For censored data, the criterion (14) must be modified because $\mathbf{t}_{i,\text{obs}}$ is not available. Here the problem is solved by estimating the actual failure time $\mathbf{t}_{i,\text{obs}}$. If the ith observation is right censored at $\mathbf{v}_i$ then there are two possibilities depending on whether the predictive mean μ_i is smaller or greater than the censoring time $\mathbf{v}_i$. If $\boldsymbol{\mu}_i$ is smaller than $\mathbf{v}_i$ then we estimate the actual observation by the censoring time $\mathbf{v}_i$. On the other hand, if $\boldsymbol{\mu}_i$ is larger than $\mathbf{v}_i$ then we estimate the observation by $\boldsymbol{\mu}_i$. In this way it is also easy to estimate the actual observation if any other form of censoring were of interest. Estimating the observation in this way actually minimizes the goodness-of-fit in (14), see e.g. Gelfand and Ghosh (1998) for more discussion. Our modified criterion is as follows:

$$D = \sum_{i=1}^{n} tr(\Sigma_i) + \sum_{i=1}^{n} (\boldsymbol{\mu}_i - \mathbf{v}_i)^T (\boldsymbol{\mu}_i - \mathbf{v}_i), \tag{15}$$

where $\mathbf{v}_i$ is $\mathbf{t}_{i,\text{obs}}$ if the ith observation is a failure time, and it is maximum of the censoring time and the predictive mean $\boldsymbol{\mu}_i$ if the ith observation is censored.

Note that the loss function (13) penalizes equally for both under and over estimation by the replicated observations t_{rep}. This is meaningful if the assumed survival distribution is symmetric. However, we assume Weibull or exponential survival distributions (which are skewed) in the previous sections. Hence it is clear that the above loss function (13) should be modified to account for the skewness of the survival distributions. Here we suggest using the loss function (13) in the log-scale. That is, we adopt (13) where t_{rep} and t_{obs} are replaced by their natural logarithms. This dictates that the model choice criterion (15) is to be calculated using the logarithm of the observations rather than the observations themselves.

5 Examples

5.1 Simulated Example

We simulate n (=1000) observations from the WMF model with one covariate. The covariate takes the value 0 for the first 500 simulations and 1 for the last 500 observations. The WMF model has five λ parameters and five α parameters. With one covariate, the WMF model has also five β parameters. For our illustration we arbitrarily choose the parameter values as follows: $\boldsymbol{\lambda} = (0.018, 0.025, 0.0026, 0.022, 0.021)$ and $\boldsymbol{\beta} = (-0.856, 0.410, 0.545, -0.415, -1.37)$ and $\boldsymbol{\alpha} = (1.73, 1.80, 1.86, 1.99, 1.89)$. We censored 30% of the samples randomly as follows. An independent Bernoulli trial with success probability 0.30 decides whether to censor the ith observation. If the trial results in a success three situations arise depending on whether $t_{i1} < t_{i2}$, or $t_{i1} > t_{i2}$, or $t_{i1} = t_{i2}$. If the observation is such that $t_{i1} < t_{i2}$, an independent uniform random variable, u_i, is drawn in the interval (t_{i1}, t_{i2}). The second component of the ith observation is censored at $\min(u_i, t_{i2})$. The case when the original observation is such that $t_{i1} > t_{i2}$ is censored similarly and note that in this case the first component is censored. In the third situation, when $t_{i1} = t_{i2}$ an uniform random variable, u_i, in the interval $(0, t_{i1})$ is drawn and both t_{i1} and t_{i2} are censored at u_i. In the resulting sample, 320 observations become censored and the sample mean of the survival times are 8.035 and 6.052 respectively for the two components.

We fit all five parametric models to this data set. See the Appendix for the Gibbs sampling implementation details and notes on convergence. We only illustrate model choice using this example. (Our second example below illustrates posterior inference in detail.) Here the frailty model, FRL, has 4 parameters. The EMO, EMF, WMO and WMF models have 6, 10, 9 and 15 parameters respectively. Table 1 provides the model choice criteria for all five models. According to the BIC we get the model order: WMF, WMO, FRL, EMF, EMO with WMF being the best. However, according to the model choice criterion (15) we get the order WMO, WMF, FRL, EMO, EMF.

There are several explanations why WMF and WMO switch places in the two orderings. The value of the likelihood function evaluated at each MCMC iteration shows that the likelihood is generally much higher for the WMF than the WMO. Also the penalty term in BIC is much smaller compared to the likelihood term. This is why the BIC criterion says that WMF is better than WMO.

Although the WMF model is a more complex model than the WMO model, the goodness-of-fit, G, for the WMO model turns out to be smaller. As mentioned in Section 4 this is possible since G is not the classical likelihood-ratio statistic. Moreover, the WMO model is not a nested sub-model of the WMF model as discussed in Section 3. Therefore, it is generally not to be expected that WMF has smaller G than the WMO. Also WMF receives more penalty than the simpler WMO model. This is why the criterion (15) says that the WMO is better than the WMF. Similar comments apply to the reversal of EMO and EMF in the two orderings.

The frailty model, FRL, is not selected as the best model using any of the criteria. It is better than the simple EMO and EMF models since a flexible Weibull type hazards function is more suitable for the simulated data. However, as plotted in Figure 1 the frailty models generally induces more dependence than the models of Section 3 and this was not required for the simulated data from the WMF model.

5.2 Eye data Example

We consider a data set obtained from the Diabetic Retinopathy Study (Huster *et al.*, 1989) of time to blindness in each eye of 197 patients with diabetic Retinopathy. One eye of each patient was randomly selected for treatment (the effectiveness of laser photo coagulation in delaying the onset of blindness) and

	FRL	EMO	EMF	WM0	WMF
BIC	9989.57	10427.49	10294.218	9909.41	9709.616
G	1174.49	1115.78	1070.61	1017.02	1035.03
P	1793.28	3288.65	4895.94	1264.77	1534.68
$G+P$	2967.77	4404.43	5966.55	2281.79	2569.71

Table 1: Model selection criteria for the simulated data example. P is the penalty and G is the goodness-of-fit term in (15). BIC is the Bayesian information criterion given in (11).

the other eye was observed without treatment. A binary age covariate (0 for juvenile and 1 for adult) is available. The first component of the bivariate survival times is the time to blindness (measured from a suitable experimental starting time point) on the treated eye and the second component is the similar time for the untreated eye.

Out of the 197 patients, 159 patients experienced some form of censoring. Censoring was caused by death, dropout, or end of the study. Survival times of both eyes were censored for 80 patients. Both eyes of these patients were censored at the same time but the censoring times were different for different patients. The treated eyes alone were censored for 63 patients and the censoring times were greater than or equal to the failure times of the untreated eyes. The reverse happened for only 16 patients. The remaining 38 patients experienced failures in both eyes, of which simultaneous failures were observed in 6 patients.

The primary objectives of the study are to (1) see the effectiveness of treatment, (2) investigate whether there is any association between the treated and the untreated times to the event, and (3) investigate the effect of age of the patients on the event times. We first consider simple summary statistics (ignoring the censoring) to fix ideas about the data. The sample mean of the survival time for the treated eye is 38.87 months while the same for the untreated eye is 32.29. This may indicate that there may be treatment effect. Summary statistics measuring the association between the times to event on both the eyes are also calculated. The Kendall's τ is estimated to be 0.385 and the Pearson's product moment correlation between the log of the times is 0.395, indicating positive dependence between the survival times of the two eyes. The mean time to blindness for the juvenile patients is 15.38 months while the same for the adult patients is 20.2 months. Hence the age covariate may also have an effect on the time to blindness. A simple plot of the data is given in Figure 3.

For the frailty model, FRL, a binary covariate taking the value 1 for the treated eye and the value 0 for the untreated eye is introduced. Hence $\boldsymbol{\beta}$ in (1) has two components β_1 and β_2. Here β_1 is the regression

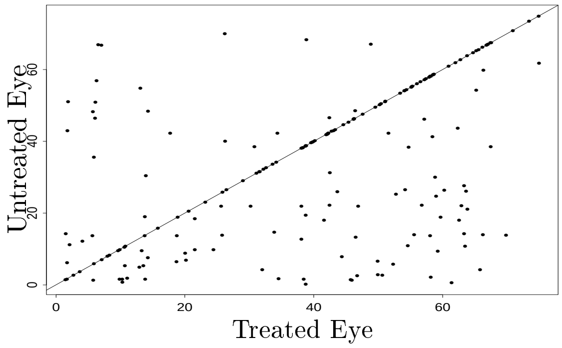

Figure 3: Scatter plot of the eye data. The line $y = x$ is super-imposed.

coefficient for the treatment effect and β_2 is the coefficient for the age effect.

However, the models presented in Section 3 do not need the binary covariate corresponding to the treatment. The treatment effect from these models can be retrieved from the intensity parameter $\lambda_j = \exp(\xi_j), j = 1, 2$ introduced in Section 3.3. Any difference between λ_1 and λ_2 can be accounted for by the treatment effect, under the very pragmatic assumption that both the eyes have the same baseline hazards. Consequently, the PH components in (9) and (10) are all one dimensional and the resulting β relates the age-related binary covariate to the hazards functions.

5.2.1 Model Choice

We fit five models, FRL, EMO, EMF, WMO and WMF to the data set. See the appendix for implementation notes for the Gibbs sampler. Note that the FRL model has 5 parameters and the EMO, EMF, WMO and WMF models have 6, 10, 9 and 15 parameters respectively. Table 2 shows the model choice criteria. According to the BIC we get the model order EMO, WMO, EMF, WMF, FRL. However, according to the (15) criterion we get the model order EMO, FRL, WMO, EMF, WMF.

The WMO and the WMF models receive much more penalty than the corresponding EMO and EMF models. The α parameters for the WMO and the WMF models had substantial posterior mass below unity. As a consequence the posterior predictive means and variances under the WMO and the WMF models became large on the average. Therefore the penalty terms for the Weibull models are higher than the corresponding exponential models. As mentioned in the previous example the goodness-of-fit term is not the classical likelihood ratio statistic. Hence there is no guarantee that the goodness-of-fit terms for the WMO and WMF models should be smaller than the EMO and EMF models respectively.

Now we discuss what happens when we change the prior assumptions for the shape parameters α's in the Weibull cases. We continue to use the same prior distributions for α's but restrict them in the interval between 1 and 5. The model choice criteria are given in Table 3. The BIC's are larger than the ones given in Table 2 because the criterion depends on the likelihood function and it is smaller for the restricted model. However, the values of the criterion (15) are smaller in Table 3 because the Weibull model with $\alpha > 1$ has much smaller mean and variance than the same for the model with $\alpha < 1$. Moreover, now the model ordering according to the criterion (15) is WMO, EMO, FRL, WMF and EMF.

	FRL	EMO	EMF	WM0	WMF
BIC	1768.10	1692.96	1715.11	1703.05	1729.27
G	1220.72	514.02	800.01	538.45	811.88
P	308.60	660.20	1211.83	1046.27	1811.33
$G+P$	1529.32	1174.22	2011.84	1584.72	2623.21

Table 2: Model selection criteria for the eye data example. P is the penalty and G is the goodness-of-fit term in (15). BIC is the Bayesian information criterion given in (11).

	FRL	WM0	WMF
BIC	1829.00	1713.29	1748.69
G	917.41	506.04	736.91
P	317.51	594.16	1002.50
$G+P$	1234.92	1100.20	1739.41

Table 3: Model selection criteria for the eye data example. Here α's are restricted to lie in $(1, 5)$.

5.2.2 Posterior Inference

The parameter estimates for the frailty model is given in Table 4. The posterior mean of β_1 is -0.9752 with standard deviation 0.187 and the 95% credible region is the interval $(-1.3363, -0.6210)$. This seems to indicate that the hazards for untreated eye is about 2.65 ($\approx \exp(0.9752)$) times the same for the treated eye on the average. (Note that the treatment covariate has value 1 for the treated eye and 0 for the untreated eye.) Also the treatment effect seems to be strong since the 95% credible region does not include the point zero.

To investigate the effect of age as estimated by the frailty model we consider the posterior distribution

	Mean	sd	2.5%	97.5%
η	1.036	0.367	0.558	1.959
μ	0.029	0.008	0.016	0.046
α	0.953	0.076	0.816	1.111
β_1	–0.975	0.186	–1.337	–0.621
β_2	0.031	0.231	–0.430	0.489

Table 4: Parameter estimates for the frailty model.

of β_2. The posterior mean of β_2 is 0.0312 with standard deviation 0.231 and the 95% credible region is (–0.429, 0.489). Recall that the age covariate has value 1 for adult and 0 for the juvenile patients. Since the posterior mean is positive, we can conclude that the hazards for the adult patients are higher than the same for the juveniles. However, the 95% credible region includes the point zero and the estimate of β_2 is not very far away from the point zero. Here we feel that the shape of the marginal posterior density of β_2 may give us more information. We plot this density in Figure 4. The marginal posterior distribution has a positive modal point. This shows that the effect of the age covariate is very little on the positive side.

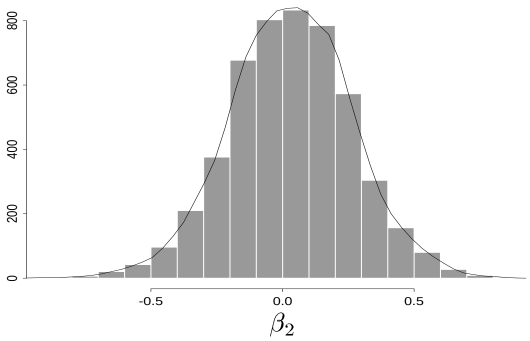

Figure 4: Marginal posterior density of β_2 under the frailty model.

Now we consider the EMO model. There are six parameters in this model and their estimates are given in Table 5. First we see the treatment effect by considering λ_1 and λ_2. Notice that 95% credible regions for these two parameters are almost non-overlapping and on the average λ_1 is smaller than λ_2. This is attributed to the treatment effect. The parameter $\lambda_1 + \lambda_3$ is the marginal baseline intensity of failure for the treated eyes and $\lambda_2 + \lambda_3$ is the same for the untreated eyes. Since the mean of the failure distribution is inversely related with the intensity, it is concluded that the untreated eye is likely to fail before the treated eye. It is also seen that this effect is very strong since the 95% credible regions are almost non-overlapping.

	Mean	sd	2.5%	97.5%
λ_1	0.0076	0.0014	0.0052	0.0106
λ_2	0.0125	0.0018	0.0092	0.0162
λ_3	0.0011	0.0006	0.0002	0.0026
β_1	–0.7845	0.3755	–1.5550	–0.0900
β_2	0.3800	0.2080	–0.0361	0.7894
β_3	0.3318	0.8358	–1.2950	1.9670

Table 5: Parameter estimates for the EMO model.

The parameter λ_3 relates to the intensity of failure of both eyes at the same time. We see that its posterior mean is 0.0011. This shows that there is positive probability of simultaneous failure. This is also confirmed by the data, see Figure 3. This also shows that the two survival times can not be treated independently as λ_3 is estimated to be positive. Hence there is association between the two survival times.

Now we consider the effect due to age. The parameters β_1, β_2 and β_3 relates the effect of this covariate to the hazards functions. Notice that β_j's are entered into the model through the equation (9). Therefore, β_1 and β_2 measure the effect of the covariate for the treated and the untreated eye respectively where these were first to fail. Since the posterior mean of β_1 is negative while the posterior mean of β_2 is positive, this shows the interaction between treatment and age. The treatment is highly effective for the juveniles and not so effective for the adults. If there were no interaction effect, the marginal posterior distributions of β_1 and β_2 would have been the same. In our case the two distributions are very different and their effective supports are almost non-overlapping, see Figure 5. These conclusions agree with the Kaplan-Meyer survival curves and the conclusions reported by Huster *et al.* (1989). That is, the treatment is more effective for the juvenile patients.

The parameter β_3 measures the effect of age for the simultaneous failure of both eyes. Since its estimate is positive, it shows that if the eyes were to fail simultaneously then the hazards for the adult patients increases exp(0.3318) times the hazards for the juvenile patients.

Before we consider the EMF model we provide the estimates of the posterior correlations between the parameters in the EMO model in Table 6. The parameter pair λ_j and β_j for each $j = 1, 2, 3$ is negatively correlated as expected. The maximal negative correlation is between λ_2 and β_2 and this is not alarmingly high.

Table 7 gives the summary statistics for the parameters of the EMF model. The six parameters λ_j and

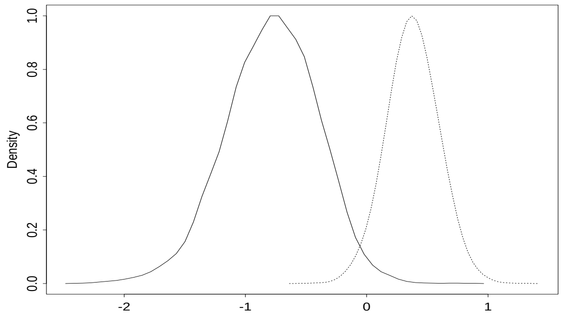

Figure 5: Marginal posterior densities of β_1 and β_2 under the EMO model. The figure shows the interaction effect between treatment and the age covariate. The solid curve on the left is the density of β_1 (coefficient of age for the treated eye) and the dotted curve is the density of β_2 (coefficient of age for the untreated eye).

	λ_1	λ_2	λ_3	β_1	β_2	β_3
λ_1	1.000	0.023	–0.118	–0.479	–0.008	0.078
λ_2	0.023	1.000	–0.054	–0.007	–0.695	0.053
λ_3	–0.118	–0.054	1.000	0.048	0.034	–0.664
β_1	–0.479	–0.007	0.048	1.000	0.002	–0.182
β_2	–0.008	–0.695	0.034	0.002	1.000	–0.043
β_3	0.078	0.053	–0.664	–0.182	–0.043	1.000

Table 6: Correlation estimates between the parameters of the EMO model.

β_j for $j = 1, 2$ and 3 have similar interpretations as in the EMO model. Hence this is not repeated here. We have also computed the correlations between the parameters. As in the EMO model the maximal negative correlation (–0.72) was observed between λ_2 and β_2. Again this is not alarmingly high. Comparison of λ_j and $\lambda'_j, j = 1, 2$ from this model shows how the intensity of failure changes after the failure of one eye. It is observed from the table that on the average λ'_j is higher than the λ_j for each $j = 1, 2$. Hence it can be concluded that there is positive association between the failure times. This also shows that the other eye is more susceptible to fail after the failure of one eye. In other words the hazards for the surviving eye increases on the average after the failure of one eye according to this model.

Note that β'_1 and β'_2 are not very efficiently estimated. Their standard deviations are little higher than the standard deviations for the remaining parameters. This is because, in this data set there were not enough observations to estimate these effectively. In our data set there were only 7 observations (for which the untreated eye failed after the treated eye) which can provide information about β_2'. Nevertheless, we can learn more about the interaction effect between the treatment and age from here. The parameter β_1' measures the effect of age on the hazards function after the failure of the untreated eye. Now the effect of age is much more pronounced here. After the failure of the untreated eye the hazards for the adult patients increases approximately 3.31 times ($\approx \exp(1.1998)$) the same for the juvenile patients. The estimate of β_2' also shows that the adult patients have higher hazards than the juvenile patients.

	Mean	sd	2.5%	97.5%
λ_1	0.0065	0.0014	0.0040	0.0096
λ_2	0.0120	0.0019	0.0086	0.0161
λ_3	0.0010	0.0006	0.0002	0.0023
λ_1'	0.0118	0.0038	0.0053	0.0200
λ_2'	0.0147	0.0053	0.0060	0.0268
β_1	−0.8587	0.4487	−1.8000	−0.0252
β_2	0.4166	0.2210	−0.0150	0.8467
β_3	0.4442	0.9084	−1.3181	2.2650
β_1'	−1.1998	1.6421	−7.2292	0.3108
β_2'	−0.2638	1.6520	−6.1076	1.5026

Table 7: Parameter estimates for the EMF model.

The models based on the Weibull distributions, WMO and WMF, provide similar conclusions. We do not consider these in detail, since these have not been selected by the model choice criteria discussed above.

6 Discussion

The frailty models incorporating individual level effects are easy to posit for multivariate survival data. The bivariate exponential and Weibull models require simultaneous failures while the frailty models do not. However, the frailty models can still be used for the type of data sets which include data points with simultaneous failures. Huster *et. al.* (1989) use such models for the eye data set analyzed here. Also the the bivariate exponential and Weibull models can be modified by removing the singularity part in each case to model data sets where simultaneous failure is not required.

This paper illustrates that the frailty models may fail to achieve better fits than the models based on bivariate exponential and Weibull distributions. However, this conclusion may well be reversed in the general multivariate setup, i.e., when the failure times are multivariate. The discussion here is closely related to its parallel in the generalized linear mixed models literature. A largely unresolved question is 'is it better to model correlated data points using the repeated measures modeling approach than to model directly using multivariate models?'. We hope that this paper has gone some way in resolving this. However, further investigation in this area is necessary.

The two model choice criteria adopted in this paper illustrates model comparison techniques using summary measures which can be calculated easily by little additional work after fitting a model by the MCMC methods. However, we note that in order to calculate the BIC criterion in (11) exactly, one needs to implement a maximization algorithm, e.g. the EM algorithm, for fitting the models. In this article we have calculated the BIC using average values of the likelihood functions evaluated at the posterior samples generated by Gibbs sampling. This allowed us to easily compare the models using a likelihood based criterion. Our second model choice criterion based on the loss function approach however is more appropriate and it can be used in practical problems easily.

Relevant questions on parsimony, i.e., whether to use EMO or WMO models can easily be resolved using Bayesian model selection criteria as is done here. In general, the Weibull distribution based models should give more adequate fits since those are more flexible. For our simulated example the Weibull models provided better fits, however, for the second example the exponential based models were better. This shows the importance of the model choice methods to deciding which model to adopt.

The bivariate models presented here are more suitable for large data sets. For the EMF and WMF models, the data set has to fit five PH models corresponding to five different domains as indicated through the ellipses in Figure 2. A large number of data points in each domain is needed to efficiently estimate the regression coefficients from each PH model. For small data sets, however, convenient fix-ups can be performed, e.g. some $\boldsymbol{\beta}$ can be constrained to lie in a neighborhood of zero.

The computational methods developed in this paper are easy to use and are potential tools for further analysis and exploration of bivariate survival data. Engineering applications involving reliability and incorporating covariate information can be undertaken easily using these methods. It is however, noted that the models discussed here are heavily parametric. Fully non-parametric Bayesian models in the multivariate setup are difficult to consider because of identifiability issues, see e.g. Crowder (1994). However, a middle ground e.g. semi-parametric models can be adopted, see Gustafson (1997).

Acknowledgements

We thank Professor Alan Gelfand for providing a preprint of his paper and for many helpful conversations and Steven Self for providing the data set. We also thank two anonymous referees for their helpful comments and suggestions which improved the presentation of the paper.

Appendix: Computational Details

Likelihood Contributions

One of the following six censoring situations can happen for each data point (t_{i1}, t_{i2}).

1. $t_{i1} < t_{i2}$: the first component fails before the second and they both fail.
2. $t_{i1} > t_{i2}$: the second component fails first and they both fail.
3. $t_{i1} = t_{i2} = t$: they both fail at the same time.
4. $t_{i1} \leq C_{i2} < t_{i2}$: the second is censored at C_{i2} and the first fails before C_{i2}.
5. $t_{i2} \leq C_{i1} < t_{i1}$: the first is censored at C_{i1} the second fails before C_{i1}.
6. $C_i < \min(t_{i1}, t_{i2})$: they both are censored at the same time.

The likelihood contributions for any of the four models (EMO, EMF, WMO and WMF) are straightforward from the first three cases (1, 2 and 3 from the above list). Routine integrations yield the likelihood contributions for the last three censored cases. One simple rule applies to these three cases. If a component has been censored then replace it by its censoring value and the likelihood contribution in this case is the density in the corresponding un-censored case without the intensity multiplier. For the sake of completeness we write down the likelihood contributions below for all four models EMO, EMF, WMO and WMF. In each of the four models the expression in the enumerated list correspond to the corresponding censoring mechanism enumerated above. For the EMF model the likelihood contributions are:

1. $\theta_1(\theta_2' + \theta_3)\exp\left\{-(\theta_2' + \theta_3)(t_{i2} - t_{i1}) - \theta t_{i1}\right\}$.
2. $\theta_2(\theta_1' + \theta_3)\exp\left\{-(\theta_1' + \theta_3)(t_{i1} - t_{i2}) - \theta t_{i2}\right\}$

3. $\theta_3 \exp\{-\theta t\}$ where $\theta = \theta_1 + \theta_2 + \theta_3$.

4. $\theta_1 \exp\{-(\theta_2' + \theta_3)(C_{i2} - t_{i1}) - \theta t_{i1}\}$.

5. $\theta_2 \exp\{-(\theta_1' + \theta_3)(C_{i1} - t_{i2}) - \theta t_{i2}\}$.

6. $\exp\{-\theta C_i\}$.

For the EMO model we set $\theta_j' = \theta_j$ for $j = 1, 2$ in the above. Before considering the WMO and the WMF models we recall the following definitions given in Section 3.2. We have assumed $\theta_j(t) = \theta_j \alpha_j t^{\alpha_j - 1}, j = 1, 2, 3$ and $\Theta_j(t) = \theta_j t^{\alpha_j}, j = 1, 2, 3$; $\Theta(t) = \Theta_1(t) + \Theta_2(t) + \Theta_3(t)$ and

$$\begin{aligned}
\theta_j''(t_{ij}|t_{ij'}) &= \theta_j'\alpha_j'(t_{ij} - t_{ij'})^{\alpha_j' - 1} + \theta_3\alpha_3 t_{ij}^{\alpha_3 - 1},\ \theta_j', \alpha_j' > 0, j \neq j' = 1, 2; \\
\Theta_j''(t_{ij}|t_{ij'}) &= \theta_j'(t_{ij} - t_{ij'})^{\alpha_j'} + \theta_3 t_{ij}^{\alpha_3},\ j \neq j' = 1, 2.
\end{aligned}$$

The likelihood contributions for the WMO models are as follows:

1. $\theta_1(t_{i1})\{\theta_2(t_{i2}) + \theta_3(t_{i2})\}\exp\{-\Theta_1(t_{i1}) - \Theta_2(t_{i2}) - \Theta_3(t_{i2})\}$.

2. $\theta_2(t_{i2})\{\theta_1(t_{i1}) + \theta_3(t_{i1})\}\exp\{-\Theta_1(t_{i1}) - \Theta_2(t_{i2}) - \Theta_3(t_{i1})\}$.

3. $\theta_3(t)\exp\{-\Theta(t)\}$ where $t_{i1} = t_{i2} = t$.

4. $\theta_1(t_{i1})\exp\{-\Theta_1(t_{i1}) - \Theta_2(C_{i2}) - \Theta_3(C_{i2})\}$.

5. $\theta_2(t_{i2})\exp\{-\Theta_1(C_{i1}) - \Theta_2(t_{i2}) - \Theta_3(C_{i1})\}$.

6. $\exp\{-\Theta(C_i)\}$.

The likelihood contributions for the WMF models are as follows:

1. $\theta_1(t_{i1})\theta_2''(t_{i2}|t_{i1})\exp\{-\Theta_2''(t_{i2}|t_{i1}) - \Theta_1(t_{i1}) - \Theta_2(t_{i1})\}$.

2. $\theta_2(t_{i2})\theta_1''(t_{i1}|t_{i2})\exp\{-\Theta_1''(t_{i1}|t_{i2}) - \Theta_1(t_{i2}) - \Theta_2(t_{i2})\}$.

3. $\theta_3(t)\exp\{-\Theta(t)\}$ where $t_{i1} = t_{i2} = t$.

4. $\theta_1(t_{i1})\exp\{-\Theta_2''(C_{i2}|t_{i1}) - \Theta_1(t_{i1}) - \Theta_2(t_{i1})\}$.

5. $\theta_2(t_{i2})\exp\{-\Theta_1''(C_{i1}|t_{i2}) - \Theta_1(t_{i2}) - \Theta_2(t_{i2})\}$.

6. $\exp\{-\Theta(C_i)\}$.

Steps for the Gibbs Sampler

Here the Gibbs sampler is implemented in two main steps. First we calculate the joint posterior density of the parameters and then we use the adaptive rejection Metropolis sampling scheme of Gilks *et al.* (1995) to sample the individual parameters. (These routines are publicly available from `http://www.mrc-bsu.cam.ac.uk/`.) Although the Gibbs sampler worked for the two examples we consider in this paper, there are other MCMC algorithms, e.g. the hybrid Monte Carlo method, which can be used, see Gustafson (1997, 1998).

The joint posterior density is calculated as follows. For the EMO and the WMO models we impose the proportional hazards components given by (9) in the likelihood contributions. That is, we calculate the likelihood by replacing θ_j by $\exp\left\{\xi_j + \mathbf{z}_i^T \boldsymbol{\beta}_j\right\}$ for the ith data point. Then the likelihood is multiplied by the prior densities $N(0, 10^4)$ for ξ_j, $j = 1, 2, 3$ and for each component of $\boldsymbol{\beta}$. In addition for the WMO model the prior density $Gamma(\kappa, \kappa)$ for each α_j are also multiplied. The resulting product is the full posterior density for $\boldsymbol{\beta}, \xi_1, \xi_2, \xi_3$ in the EMO case and for $\boldsymbol{\beta}, \xi_1, \xi_2, \xi_3, \alpha_1, \alpha_2, \alpha_3$ in the WMO case.

Now we turn to the EMF and WMF case. For these two models we replace θ_j by $\exp\left\{\xi_j + \mathbf{z}_i^T \boldsymbol{\beta}_j\right\}$ as in the EMO and the WMO models and in addition we replace θ_j' by $\exp\left\{\xi_j' + \mathbf{z}_i^T \boldsymbol{\beta}_j'\right\}$ as given in (10). Then the prior densities for all the parameters ξ_j, ξ_j', α_j, α_j' and $\boldsymbol{\beta}_j$ and $\boldsymbol{\beta}_j'$ are multiplied. This results in the joint posterior density.

Convergence Assessment

Several visual convergence diagnostics, e.g. the time series plots and the auto-correlation plots of different parameters have been checked. For illustration, Figure 6 gives the plots for the parameter β_1 in all five models for the eye data example. These did not show any problem in convergence. Although these diagnostics have value, these can not be used to provide definitive proofs of convergence. Usually many more checks using different starting values, random number seeds, and prior-parameter specifications are necessary to see that the coded Gibbs sampler is giving stable estimates of the parameters. All of these have been performed for our examples. These are omitted for brevity. In the end we decided to use 5000 iterates for performing posterior inference and model choice analysis after discarding 1000 initial estimates.

REFERENCES

Block, H. W. and Basu, A. P. (1974) A Continuous Bivariate Exponential Extension. *J. Amer. Statist. Assoc.*, **69**, 1031–1037.

Clayton, D. (1978) A model for association in bivariate life tables and its application in epidemiological studies of familiar tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.

Clayton, D. (1991) A Monte Carlo Method for Bayesian inference in frailty model. *Biometrics*, **47**, 467–485.

Clayton, D. and Cuzick, J. (1985) Multivariate generalizations of the proportional hazards model (with discussion). *J. Roy. Statist. Soc.*, A **148**, 82–117.

Crowder, M. (1994) Identifiability Crises in Competing Risks. *Int. Statist. Rev.*, **62**, 379–391.

Freund, J. E. (1961) A Bivariate Extension of the Exponential Distribution. *J. Amer. Statist. Assoc.,* **56**, 971–977.

Gelfand, A. E. (1996) Model determination using sampling based methods. In *Markov Chain Monte Carlo in Practice* (Eds. W. R. Gilks, S. Richardson and D. J. Spiegelhalter). London: Chapman and Hall, pp 145–161.

Gelfand, A. E. and Ghosh, S. K. (1998) Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika*, **85**, 1–11.

Ghosh, S.K. and Gelfand, A, E. (1998) Latent Waiting Time Models for Bivariate Event Times with Censoring. *Sankhya*, B, **60**, 34–47.

Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995) Adaptive Rejection Metropoils Sampling within Gibbs Sampling. *Appl. Statist.*, 455–472.

Gustafson, P. (1997) Large hierarchical Bayesian analysis of multivariate survival data. *Biometrics*, **53**, 230–242.

Gustafson, P. (1998) Flexible Bayesian modelling for survival data *Lifetime Data Analysis*, **4**, 281–299.

Hanagal, D. D. (1992) Some Inference Results in Modified Freund's Bivariate Exponential Distribution. *Biom. J.*, **34**, 745–756.

Huster, W. J., Brookmeyer, R. and Self, S. G. (1989) Modelling Paired Survival Data with Covariates. *Biometrics*, **45**, 145–156.

Joe, H. (1997) *Multivariate Models and Dependence Concepts.* London: Chapman & Hall.

Kass, R. E. and Raftery, A. E. (1995) Bayes factors. *J. Amer. Statist. Assoc.*, **90**, 773–795.

Key, J. T., Pericchi, L. R. and Smith, A. F. M. (1999) Bayesian Model Choice: What and Why? In *Bayesian Statistics 6* (Eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith). Oxford University Press, to appear.

Klein, J. P., Keiding, N. and Kamby, C. (1989) Semiparametric Marshall-Olkin Models Applied to the Occurrence of Metastases at Multiple Sites After Breast Cancer. *Biometrics*, **45**, 1073–1086.

Laud, P. W. and Ibrahim, J. G. (1995) Predictive Model Selection. *J. R. Statist. Soc.* B, **57**, 247–262.

Lee, L. (1979) Multivariate distributions having Weibull properties. *J. Mult. Anal.*, **9**, 267–277.

Lu, J. C. (1989) Weibull Extensions of the Freund and the Marshall-Olkin Bivariate Exponential Models. *IEEE Transactions on Reliability*, **38**, 615–619.

Lu, J. C. and Bhattacharyya, G. K. (1990) Some new constructions of bivariate Weibull models. *Ann. Inst. Statist. Math.*, **42**, 543–549.

Marshall, A. W. and Olkin, I. (1967) A Multivariate Exponential Distribution. *J. Amer. Statist. Assoc.*, , **62**, 30–44.

Oakes, D. (1989) Bivariate Survival Models Induced by Frailties. *J. Amer. Statist. Assoc.*, **84**, 487–493.

Proschan, F. and Sullo, P. (1974) Estimating the parameters of a bivariate exponential distribution in several sampling situations. In *Reliability and Biometry*, (Eds. F. Proschan and R. J. Serfling). Soc. Indust. Appl. Math., Philadelphia, Pa., pp. 423–440.

Sahu, S. K., Dey, D. K., Aslanidou, H. and Sinha, D. (1997) A Weibull Regression Model with Gamma Frailties for Multivariate Survival Data. *Lifetime Data Analysis*, **3**, 123–137.

Schwartz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Sinha, D. and Dey, D. K. (1996) Semiparametric Bayesian Analysis of Survival Data. *J. Amer. Statist. Assoc.*, **92**, 1195–1212.

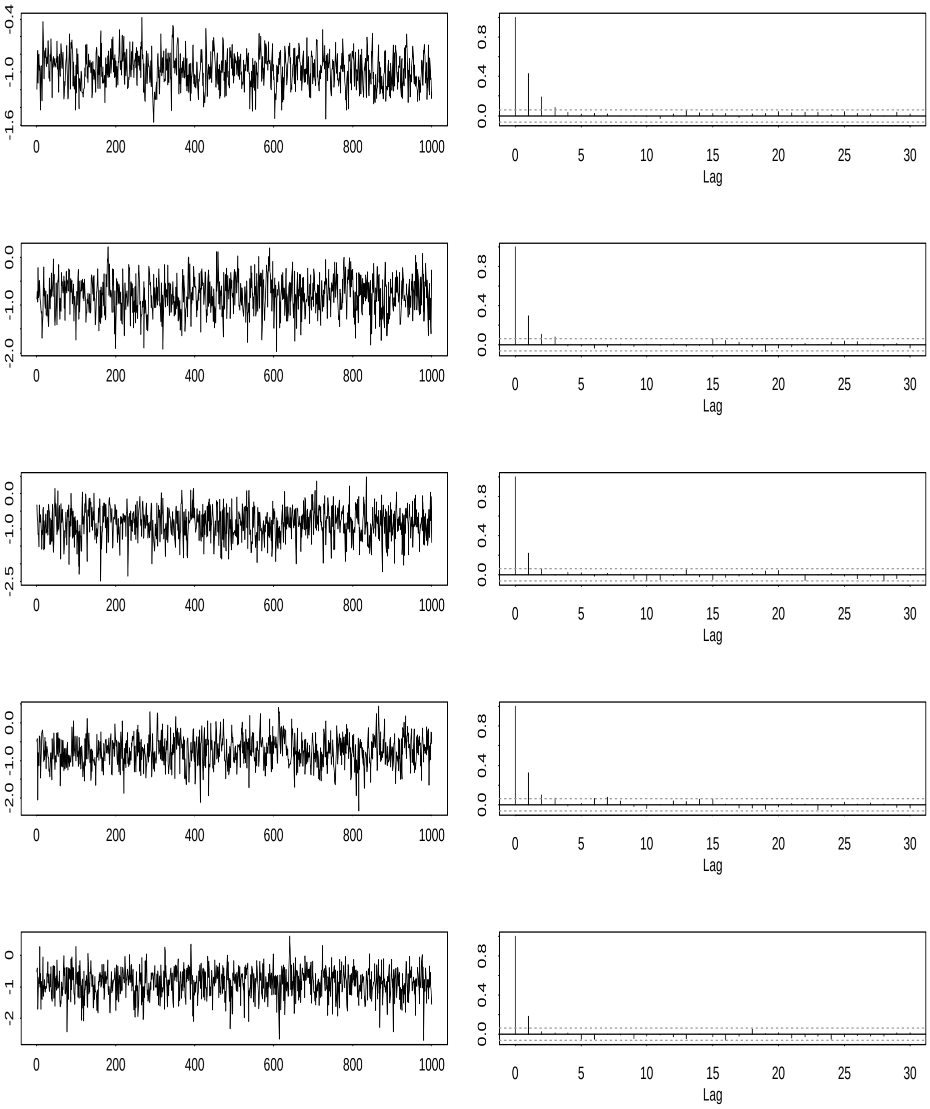

Figure 6: Time series and auto-correlation plot of β_1 from all five models for the eye data example. The plots are given in the order FRL, EMO, EMF, WMO and WMF. This plot shows rapid mixing of the Gibbs sampler for all five models for the parameter β_1.