Simultaneous Bayesian modeling of longitudinal and survival data in breast cancer patients

DOI: 10.1080/03610926.2019.1635701

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Communications in Statistics - Theory and Methods

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Simultaneous Bayesian modeling of longitudinal and survival data in breast cancer patients

Ali Azarbar
Department of Statistics, Amirkabir University of Technology, Tehran, Iran
Yu Wang, Saralees Nadarajah
School of Mathematics, University of Manchester, Manchester, UK
email: mbbsssn2@manchester.ac.uk
August 13, 2019

Abstract

Using simultaneous Bayesian modeling, an attempt is made to analyze data on the size of lymphedema occurring in the arms of breast cancer patients after breast cancer surgery (as the longitudinal data) and the time interval for disease progression (as the time-to-event occurrence). A model based on a multivariate skew t distribution is shown to provide the best fit. This outcome was confirmed by simulation studies too.

Keywords: Longitudinal and survival data; MCMC method; Simultaneous Bayesian modeling; Time-to-event

2000 Mathematics Subject Classification: 49K05, 49K15, 49S05.

1 Introduction

The simultaneous modeling of longitudinal and survival data plays a crucial role in the analysis of medical, social, and economic data. For example, in a clinical study aiming at analyzing the data on breast cancer, a biomarker such as the size of a lymph node in the arm of the patient is conceived of as the survival predictor. The putative biomarker is collected for each individual over time until the occurrence of certain events such as patients’ refusal, death, or disease recurrence. In most of previously-conducted studies on simultaneous modeling, symmetric distributions, especially the normal distribution, are deployed to model longitudinal responses; however, if the data are not symmetric, this can lead to invalid inferences. In recent years, simultaneous modeling of longitudinal data and time-to-event occurrence has received unprecedented attention. At first, a two-step approach was proposed by Tsiatis et al. (1995). Ye et al. (2008a, 2008b) put forward a two-step regression model to estimate the relationship between longitudinal data and time-to-event occurrence data. Albert and Shih (2010) demonstrated that the bias of the two-step model might be inherent and accordingly, showed how the bias can be decreased using another model. Wulfsohn and Tsiatis (1997) proved that although two-step models have the advantage of computational simplicity, they are plagued with some fundamental problems. Faucett and Thomas (1996) showed that two-step models use survival data in the modeling of longitudinal processes and, accordingly, might lead to biased results and decrease efficiency of the estimators. Furthermore, the estimated longitudinal measures in the first step are used as constants in the next stage. Hence,
this method is characterized by not transferring the uncertainty of the first stage to the second stage. Another method developed to obtain estimates and to make statistical inferences is based on the simultaneous likelihood function for longitudinal and survival data. Deploying this method leads to more exact estimates. Moreover, in this method, the relationship between longitudinal data and time-to-event occurrence is accentuated better. In this method, the parameters of the model are estimated simultaneously and accordingly, in the modeling, uncertainty is taken into account. Simultaneous modeling for the longitudinal and survival data was improved after 1990. Henderson et al. (2000) and Hashemi et al. (2003) conducted a simultaneous modeling using the hidden variable method. Ibrahim et al. (2004), Tsiatis and Davidian (2004), and Wu and Cook (2012), using their own proposed model, attempted to generalize this model. This modeling encompasses innovations in the structure of both longitudinal and survival models. These discussions, especially for longitudinal and survival submodels, involving Bayesian methods, were investigated by Chi and Ibrahim (2007), Bakar et al. (2009), Rizopoulos et al. (2008a, 2008b), Rizopoulos and Ghosh (2011), Rizopoulos (2010, 2013), Baghfalaki et al. (2014) and Baghfalaki and Ganjali (2015). Non-parametric issues to avoid the assumption that random effects are normal were formalized using the EM algorithm method in Song et al. (2002), and Tsiatis and Davidian (2001). Baghfalaki et al. (2014) and Baghfalaki and Ganjali (2015), using heavy-tailed distributions and avoiding commonly-held assumptions such as normality, fitted longitudinal and survival data with common random effects. Issues concerning robust modeling, using the $t$ distribution, were scrutinized in Brown and Ibrahim (2003), Li et al. (2009), and Huang and Dagne (2011). Lange et al. (1989) demonstrated that the inaccuracy of the assumption of normality for random effects exerts an adverse effect on the variance of estimates. Given this problem, fitting random effects model in the presence of skewed distributions has attracted researchers’ attention. The skew normal family of distributions constitutes an important class of asymmetric distributions. Azzalini (1985) introduced the first version of this family of distributions. Generalizations of this distribution to the multivariate case were provided in Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Branco and Dey (2001), and Sahu et al. (2003). The applicability of the multivariate normal distribution in mixed-effects models was studied by Arellano-Valle et al. (2007), and Lin and Lee (2008). Furthermore, applicability of this model in the presence of missing data was examined by Lin et al. (2009), Baghfalaki et al. (2014) and Baghfalaki and Ganjali (2015). Choudhary et al. (2014) introduced a general skew-t mixed model that allows different degrees of freedom for random effects. Lu et al. (2017) introduced partially linear mixed effects joint models for skewed and missing longitudinal competing risks outcomes.

In this paper, we analyze longitudinal and survival data with a family of skewed-elliptical multivariate distributions using Bayesian approach. The family is more efficient than the multivariate normal distribution.

The contents of this paper are organized as follows. In Section 2, we introduce skewed-elliptical multivariate distributions. Section 3 introduces longitudinal and survival models and their notations. Section 4 introduces Bayesian construction for simultaneous modeling of longitudinal and survival data. Section 5 performs a simulation study for checking the proposed model. Section 6 fits the proposed model to a data from breast cancer patients. Discussion and conclusions are given in Section 7.
2 The family of skewed-elliptical multivariate distributions

The elliptical distribution family refers to a set of distributions whose density is an ellipsoidal equation. Suppose $\mu \in R^k$ is the location parameter and $\Sigma$ is a positive-definite matrix with the dimension $k$ and $X$ is also a $k$ dimensional random vector with the following probability density function:

$$f(x; \mu, \Sigma; g^{(k)}) = |\Sigma|^{-\frac{1}{2}} g^{(k)} \left\{ (x - \mu)' \Sigma^{-1} (x - \mu) \right\}, \ x \in R^k,$$

where $g^{(k)}$ is the probability density generator function of the random variable of $X$ defined as

$$g^{(k)}(\cdot) = \frac{1}{\pi^{k/2}} g(\cdot; k) \left( \int_0^\infty u^k g(u; k) \, du \right)^{-1}.$$

In the case $g(x; k) = \exp(-x^2/2)$,

$$g^{(k)}(x) = \frac{1}{\pi^{k/2}} \exp\left(-\frac{x^2}{2}\right) \left( \int_0^\infty u^k \exp\left(-\frac{u^2}{2}\right) \, du \right)^{-1} = \frac{1}{\Gamma\left(\frac{k}{2}\right)} \exp\left(-\frac{x^2}{2}\right),$$

so (1) reduces to

$$f(x; \mu, \Sigma; g^{(k)}) = |\Sigma|^{-\frac{1}{2}} \frac{1}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right\}.$$

Other terms in $g^{(k)}$ generate the density in $f$ based on the structure of the combined function.

The skew-elliptical distribution generalizes the elliptical distributions to a family with double flexibility through a skewness parameter. Assume that $\epsilon$ and $Z$ are $k$ dimensional random vectors and $\mu$ is a $m$ dimensional vector and $\Sigma$ is a positive-definite matrix $k \times k$. In the following, we consider a random vector $X$ from the elliptical distribution family

$$X = \begin{pmatrix} \epsilon \\ Z \end{pmatrix} \sim EL\left(\theta, \Omega; g^{(2k)}\right),$$

where $\theta = (\theta_0)'$, $\Omega = \begin{pmatrix} \Sigma & 0 \\ 0 & I \end{pmatrix}$, $0$ is the zero matrix and $I$ is an identity matrix. Set $Y = DZ + \epsilon$, where $D = \text{diag}(d_1, d_2, \ldots, d_k)$. $Y|Z > 0$ extends the class of elliptical distributions to the skew-elliptical class. If all the elements of the diagonal matrix $d_i$ are positive, the distribution $Y|Z > 0$ is right-skewed. However, if $d_i$ are negative, the distribution is left-skewed. Drawing on the definition of conditional density and Bayesian rule, it can be demonstrated that

$$f(y|Z > 0) = 2^k f(y; \mu, \Sigma + D^2; g^{(k)}) p(Z > 0|y),$$

where $Z \sim EL(0, I; g^{(k)})$ and $p(Z > 0) = 2^{-k}$. One writes $Y \sim SE(\mu, \Sigma, D; g^{(k)})$. The skew normal distribution is the particular case of (2) for $g(u; m) = \exp(-u^2/2)$. The skew-t distribution
is the particular case of (2) for \( g(u; 2m, \nu) = (1 + u/\nu)^{-m - \nu/2} \). For more details, refer to Branco and Dey (2001), and Sahu et al. (2003).

Figure 1 shows density plots of the bivariate skew normal distribution.

![Figure 1: Densities of the bivariate skew normal distribution for \( \rho = 0.8, d_1 = 0.7 \) and \( d_2 = 0.9 \) (top left); \( \rho = -0.8, d_1 = -0.7 \) and \( d_2 = -0.9 \) (top right); \( \rho = 0.4, d_1 = 0.7 \) and \( d_2 = 0.9 \) (bottom left); \( \rho = -0.4, d_1 = -0.7 \) and \( d_2 = -0.9 \) (bottom right).]

3 Model and notation

The initial modeling of the response variables is that besides the error components in the longitudinal sub-model, the random effects in both the longitudinal and survival sub-models have probability distributions. We first assume that \( \mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i}) \) is a vector of length \( n_i \) to evaluate the
response variable of the $i$th individual. The longitudinal model based on the linear model with mixed effects is

$$
Y_i = X_{1i}'\beta_1 + Z_{1i}'b_{1i} + \epsilon_i
$$

for $i = 1, 2, \ldots, n$. The Cox proportional hazards model with survival time of $t_i$ and censor marker $\delta_{it}$ for the $i$th individual is

$$
h(t_i|X_{2i}, Z_{2i}, b_{2i}) = h_0(t_i) \exp \{X_{2i}'\beta_2 + Z_{2i}'b_{2i}\}.
$$

Replacing non-normal proportional distributions with the deviate nature of the distribution of random effects paves the way for improving the results of estimating models. Accordingly, in the simultaneous modeling based on deploying skew-elliptical distributions for the random-effects model, in the longitudinal and survival submodels, we assume that

$$
Y_i|\beta_1, b_{1i}, \sigma_\epsilon^2 \sim N_{n_i}(X_{1i}'\beta_1 + Z_{1i}'b_{1i}, \sigma_\epsilon^2 I_{n_i}),
$$

$$
b_{1i}|\Psi_1, D_{b1}; g^{(n_i)}_{\theta b1} \sim SE_{q1}(0, \Psi_1, D_{b1}; g^{(n_i)}_{\theta b1})
$$

and

$$
b_{2i}|\Psi_2, D_{b2}; g^{(n_i)}_{\theta b2} \sim SE_{q2}(0, \Psi_2, D_{b2}; g^{(n_i)}_{\theta b2}),
$$

where $SE_q(\mu, \Sigma; g\theta)$ denotes the $q$-variable skew elliptical distribution with location parameter of $\mu$, scale matrix of $\Sigma$, skew matrix $D$ and probability density generator function $g\theta$.

Taking the distubtional assumptions introduced in (3), the likelihood function of observations for the simultaneous model is

$$
L(\theta|y, b_1, b_2, t) = f(y|b_1, \sigma_\epsilon^2, D_{b1}, \Psi_1, g_{\theta b1}) f(t, b_2, \delta_{it}|\beta_2, D_{b2}, \Psi_2, g_{\theta b2})
$$

$$
= \prod_{i=1}^m 2^{n_i} f(y_i|X_{1i}'\beta_1 + Z_{1i}'b_{1i}, \sigma_\epsilon^2 I_{n_i}) 2^{q_1} f(b_{1i}|\mu_{b1}, \Psi_1, D_{b1}, g_{\theta b1})
$$

$$
\times \int_0^\infty f(\nu_{1i}|\mu_{\nu1i}, \Sigma_{\nu1i}; g_{\theta b1}) d\nu_{1i} \prod_{i=1}^n \left\{ h_{\delta_{it}}(t_i|X_{2i}, Z_{2i}, b_{2i}) \right\}
$$

$$
\times \exp \left\{ -H_0(t_i) \exp \{X_{2i}'\beta_2 + Z_{2i}'b_{2i}\} \right\} 2^{q_2} f(b_{2i}|\mu_{b2}, \Psi_2, D_{b2}, g_{\theta b2})
$$

$$
\times \int_0^\infty f(\nu_{2i}|\mu_{\nu2i}, \Sigma_{\nu2i}; g_{\theta b2}) d\nu_{2i},
$$

where $\theta = (\beta_1, \beta_2, \sigma_\epsilon^2, \Psi_1, \Psi_2, D_{b1}, D_{b2}, \nu_{b1}, \nu_{b2})$ are the unknown parameters.

### 4 The Bayesian structure

For Bayesian modeling, suitable prior distributions need to be selected for all of the parameters in the model. For the fixed effects model, a multivariate normal distribution is selected. Accordingly, the parameters $\beta_1$ and $\beta_2$ which are $p_1$ and $p_2$ variates are considered to follow the $N_{p_l}(\beta_{\delta l}, \Sigma_{\beta_l})$ for $l = 1, 2$. For the scale matrix of random effects $\Psi_l$, the inverse Wishart prior distribution
IW \((\tau_0, S_{b_i})\) is used for \(l = 1, 2\). The prior distribution of the error-scale parameter \(\sigma^2_\epsilon\) is taken as the inverse gamma distribution with parameters \(\alpha_\epsilon\) and \(\tau_\epsilon\). The skewness parameter \(d_b\) is taken to follow the truncated multivariate normal distribution \(N_{p_l} (\mu_{d_l}, \Sigma_{d_l}) I (d_b > 0)\) for \(l = 1, 2\). Furthermore, the uniform or truncated exponential distributions are used as the prior distribution for the degree of freedom parameter of the skew-t distribution.

To obtain parameter estimates, the simultaneous posterior density should be calculated, which is achieved by multiplying the simultaneous likelihood function and simultaneous prior distribution of parameters. By calculating the simultaneous posterior density and obtaining the full conditional posterior densities for individual parameters, we can conclude that these distributions do not have closed form. Accordingly, the direct calculation of the posterior mean of these distributions proves to be difficult and impossible. Hence, Markov Chain Monte Carlo (MCMC) methods are used. According to the observed and prior distributions considered, the following posterior distribution can be obtained

\[
\pi (\beta_1, \beta_2, b_{1i}, b_{2i}, d_{b1}, d_{b2}, \nu_1, \nu_2, \Psi_1, \Psi_2 | y, t) \propto \prod_{i=1}^{n} \left\{ \phi (y_i | X_{1i}' \beta_1 + Z_{1i}' b_{1i}, \sigma^2_\epsilon I_{n_i}) \right. \\
\cdot \phi_{d1_i} (b_{1i} | \mu_{d1i}, D_{b1i}, S^{-2}_{b1i}, \Psi_1) \exp \left\{ -\frac{1}{2} (\beta_1 - \beta_{01})' S^{-1}_{\beta_1} (\beta_1 - \beta_{01}) \right\} \\
\cdot \exp \left\{ -\frac{1}{2} (d_{b1i} - \mu_{d1i})' \Sigma^{-1}_{d1i} (d_{b1i} - \mu_{d1i}) \right\} I (d_{b1i} > 0) \exp \left\{ -\frac{\nu_{b1i}}{\lambda_{b1i}} \right\} I (\nu_{b1i} > 2) \\
\left. \prod_{i=1}^{n} h_i \delta_{t_i} (t_i | X_{2i}, Z_{2i}, b_{2i}) \exp \left\{ -H_0 (t_i) \exp (X_{2i}' \beta_2 + Z_{2i}' b_{2i}) \right\} \right\}
\]

\[
\phi_{d2_i} (b_{2i} | \mu_{d2i} + D_{b2i} w_{b2i}, S^{-2}_{b2i}, \Psi_2) \exp \left\{ -\frac{1}{2} (\beta_2 - \beta_{02})' S^{-1}_{\beta_2} (\beta_2 - \beta_{02}) \right\} \\
\cdot \exp \left\{ -\frac{1}{2} (d_{b2i} - \mu_{d2i})' \Sigma^{-1}_{d2i} (d_{b2i} - \mu_{d2i}) \right\} I (d_{b2i} > 0) \exp \left\{ -\frac{\nu_{b2i}}{\lambda_{b2i}} \right\} I (\nu_{b2i} > 2).
\]

The use of the MCMC methods is contingent upon obtaining full conditional posterior distributions for the parameters of the model. If we denote by \(\Theta_{(-\theta)}\) a vector parameters with \(\theta\) omitted then

\[
\beta_1 | b_{1i}, d_{b1}, \nu_{b1i}, \sigma^2_\epsilon, S^2_{b1i}, y \sim N_{p1} \left( \mu^{-1}_{\beta_1}, \alpha_{\beta_1} \right),
\]

where

\[
\alpha_{\beta_1} = S^{-1}_{\beta_1} \beta_{01} + \sum_{i=1}^{n} \sigma^2_\epsilon X_{1i}' (y_i - Z_{1i}' b_{1i});
\]

and

\[
\mu^{-1}_{\beta_1} = \sum_{i=1}^{n} \frac{1}{\sigma^2_\epsilon} X_{1i}' I^{-1}_{n_i} X_{1i} + S^{-1}_{\beta_1}.
\]

Furthermore,

\[
b_{1i} | \Theta_{(-b_{1i})}, y, t \sim N_{q1} \left( \Omega^{-1}_{b1i} \alpha_{b1i}, \Omega^{-1}_{b1i} \right),
\]

where

\[
\Omega^{-1}_{b1i} = S^2_{b1i} \Psi_1 + \sigma^{-2}_\epsilon Z_{1i}' I^{-1}_{n_i} Z_{1i}
\]
and
\[ \alpha_{b_{1i}} = \sigma_e^{-2} Z_{i}^T I_n^{-1} (y_i - X_i^T \beta_1) + \Psi_1^{-1} (D_{b_{1i}} \nu_{b_{1i}} - \mu_{b_{1i}}). \]
Furthermore,
\[ \nu_{b_{1i}} \mid \Theta(-\nu_{b_{1i}}) \sim N_{q_1} \left( \Omega_{\nu_{b_{1i}}}^{-1} \alpha_{b_{1i},}, \Omega_{\nu_{b_{1i}}}^{-1} \right) I \left( \nu_{b_{1i}} > 0 \right), \]
where
\[ \Omega_{\nu_{b_{1i}}}^{-1} = S_{b_{1i}}^2 (D_{b_{1i}} \Psi_1 D_{b_{1i}}) + I_{q_1}. \]
Furthermore,
\[ \Psi_1 \mid \Theta(-\Psi_1), y, t \sim I W_{q_1} (n + \alpha_{b_{1i}}, \Omega_{\Psi_1}), \]
where
\[ \Omega_{\Psi_1} = S_{b_{1i}}^{-1} + \sum_{i=1}^{n} S_{b_{1i}}^2 (b_{1i} - D_{b_{1i}} \nu_{b_{1i}} + \mu_{b_{1i}}) (b_{1i} - D_{b_{1i}} \nu_{b_{1i}} + \mu_{b_{1i}})^T. \]
The conditional posterior distribution for \( \Psi_2 \) can be obtained similarly. The conditional distributions for other parameters are not in closed-form or well-known.

5 Simulation

This section performs simulation for scrutinizing the performance of the proposed model in the presence and absence of skewness. A sample with the size of \( n \) with 10000 repetitions was generated and the mixed effect model was considered as
\[ y_{ij} = \beta_{11} + \beta_{12} t_j + \beta_{13} x_i + b_{1i} + b_{2i} t_j + \epsilon_{ij}, \]
where \( i = 1, 2, \ldots, n, j = 1, 2, 3 \) and \( t = (3, 6, 9) \). Moreover, \( \epsilon_{ij} \)'s were generated from the \( N(0, 1) \) distribution. The \( x_i \)'s were generated from the Bernoulli (0.3) distribution. The following Weibull proportional hazards model was considered:
\[ h(t_i) = \lambda_i^\lambda \exp \{ \beta_{21} + \beta_{22} x_i + r_1 b_{1i} + r_2 b_{2i} \}. \]
Moreover, for the random effects, one of \( N(0, 1) \), \( t \) distribution, or skew-t distribution was considered. The simulation results are provided in Figures 2 and 3. \( N - N \) denotes the model where residuals are \( N(0, 1) \) and random effects are also \( N(0, 1) \). \( N - T \) denotes the model where residuals are \( N(0, 1) \) and random effects are \( t \) distributed with degree of freedom \( \nu \). \( N - ST \) denotes the model where residuals are \( N(0, 1) \) and random effects are skew-t distributed with skewness parameters \( (\omega_1, \omega_2) \) and degree of freedom parameter \( \nu \). \( d_{11}, d_{12}, d_{21} \) and \( d_{22} \) denote elements of the covariance matrix.

Figures 2 and 3 plot the relative biases and relative mean squared errors of the estimators of the parameters for the \( N - N \), \( N - T \) and \( N - ST \) models. Note that the parameters for the \( N - N \) model are \( \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, d_{11}, d_{12}, d_{22}, \lambda, r_1 \) and \( r_2 \). The parameters for the \( N - T \) model are \( \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, d_{11}, d_{12}, d_{22}, \lambda, r_1, r_2 \) and \( \nu \). The parameters for the \( N - ST \) model
are $\beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, d_{11}, d_{12}, d_{22}, \lambda, r_1, r_2, \nu, \omega_1$ and $\omega_2$. In the simulations, the parameter values were set as $\beta_{11} = 4$, $\beta_{21} = \beta_{22} = 1$, $\beta_{12} = \beta_{13} = 1$, $r_1 = r_2 = -1$, $\lambda = 2$, $\omega_1 = -5$, $\omega_2 = 4$, $\nu = 5$, $d_{11} = 3$, $d_{12} = 0$ and $d_{22} = 4$. The $x$ and $y$ axes in Figure 3 are in logarithmic scale.

The $N - N$, $N - T$ and $N - ST$ models were fitted to each of the 10000 samples of size $n$. This yields 10000 estimates of the parameters in the $N - N$, $N - T$ and $N - ST$ models. For example, for the $N - N$ model, we obtained the estimates $\hat{\beta}_{11,n,i}$, $\hat{\beta}_{12,n,i}$, $\hat{\beta}_{13,n,i}$, $\hat{\beta}_{21,n,i}$, $\hat{\beta}_{22,n,i}$, $\hat{d}_{11,n,i}$, $\hat{d}_{12,n,i}$, $\hat{\lambda}_{n,i}$, $\hat{\tau}_{1,n,i}$ and $\hat{\tau}_{2,n,i}$ for $i = 1, \ldots, 10000$ and each $n$. These estimates were used to compute the relative biases and relative mean squared errors of the estimators for each $n$ and for the three models. For example, for the $N - N$ model, we computed the relative biases and relative mean squared errors as

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{11,n,i} - \beta_{11}}{\beta_{11}} \right), \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{12,n,i} - \beta_{12}}{\beta_{12}} \right),$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{13,n,i} - \beta_{13}}{\beta_{13}} \right), \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{21,n,i} - \beta_{21}}{\beta_{21}} \right),$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{22,n,i} - \beta_{22}}{\beta_{22}} \right), \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{d}_{11,n,i} - d_{11}}{d_{11}} \right),$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{d}_{12,n,i} - d_{12}}{d_{12}} \right), \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{d}_{22,n,i} - d_{22}}{d_{22}} \right),$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\lambda}_{n,i} - \lambda}{\lambda} \right), \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\tau}_{1,n,i} - r_1}{r_1} \right),$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\tau}_{2,n,i} - r_2}{r_2} \right),$$

and

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{11,n,i} - \beta_{11}}{\beta_{11}} \right)^2, \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{12,n,i} - \beta_{12}}{\beta_{12}} \right)^2,$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{13,n,i} - \beta_{13}}{\beta_{13}} \right)^2, \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{21,n,i} - \beta_{21}}{\beta_{21}} \right)^2,$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\beta}_{22,n,i} - \beta_{22}}{\beta_{22}} \right)^2, \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{d}_{11,n,i} - d_{11}}{d_{11}} \right)^2,$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{d}_{12,n,i} - d_{12}}{d_{12}} \right)^2, \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{d}_{22,n,i} - d_{22}}{d_{22}} \right)^2,$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\lambda}_{n,i} - \lambda}{\lambda} \right)^2, \quad \frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\tau}_{1,n,i} - r_1}{r_1} \right)^2,$$

$$\frac{1}{10000} \sum_{i=1}^{10000} \left( \frac{\hat{\tau}_{2,n,i} - r_2}{r_2} \right)^2,$$

respectively, for each $n$.  

8
Figure 2: Relative biases versus $n$ of the parameter estimates under the $N - N$ (blue), $N - T$ (red) and $N - ST$ (black) models. The parameters are in the following order: $\beta_{11}$ (first row, left), $\beta_{12}$ (first row, right), $\beta_{13}$ (second row, left), $\beta_{21}$ (second row, right), $\beta_{22}$ (third row, left), $d_{11}$ (third row, right), $d_{12}$ (fourth row, left), $d_{22}$ (fourth row, right), $\lambda$ (fifth row, left), $r_1$ (fifth row, right), $r_2$ (sixth row, left), $\nu$ (sixth row, right), $\omega_1$ (seventh row, left) and $\omega_2$ (seventh row, right).

Figure 3: Relative mean squared errors versus $n$ of the parameter estimates under the $N - N$ (blue), $N - T$ (red) and $N - ST$ (black) models. The parameters are in the following order: $\beta_{11}$ (first row, left), $\beta_{12}$ (first row, right), $\beta_{13}$ (second row, left), $\beta_{21}$ (second row, right), $\beta_{22}$ (third row, left), $d_{11}$ (third row, right), $d_{12}$ (fourth row, left), $d_{22}$ (fourth row, right), $\lambda$ (fifth row, left), $r_1$ (fifth row, right), $r_2$ (sixth row, left), $\nu$ (sixth row, right), $\omega_1$ (seventh row, left) and $\omega_2$ (seventh row, right).
The relative biases and the relative mean squared errors generally decrease with \( n \). The relative biases appear generally positive. They appear largest for the \( N - N \) model, followed by the \( N - T \) model and then the \( N - ST \) model. The relative mean squared errors appear largest for the \( N - N \) model, followed by the \( N - T \) model and then the \( N - ST \) model. Hence, the \( N - ST \) is the best model and the \( N - N \) model is the worst model. For the \( N - ST \) model, the relative biases appear reasonably small for all \( n \geq 400 \) and the relative mean squared errors appear reasonably small for all \( n \geq 500 \).

In our simulations, we have chosen specific values for \( t, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, d_{11}, d_{12}, d_{22}, \lambda, r_1, r_2, \nu, \omega_1 \) and \( \omega_2 \). But the results of simulation were similar for a wide range of values of these parameters. In particular, the relative biases and the relative mean squared errors generally decreased with \( n \); they always appeared largest for the \( N - N \) model; they always appeared smallest for the \( N - ST \) model; the relative biases for the \( N - ST \) model always appeared reasonably small for all \( n \geq 400 \); the relative mean squared errors for the \( N - ST \) model always appeared reasonably small for all \( n \geq 500 \).

6 Application

This section analyses data from a longitudinal study conducted on 236 breast-cancer patients. After the surgery, the participants of the study went to a hospital in Tehran. The size of the lymphedema occurring in their arms were recorded at different time intervals after cancer surgery.

First, we tested the data for normality. We used the following tests: Kolmogorov-Smirnov, Shapiro-Wilk’s, Jarque-Bera, Anderson-Darling, Cramer-von Mises, Pearson chi-square and Shapiro-Francia tests. All of them returned a \( p \) value less than 0.001, giving strong evidence that the data are not normally distributed.

Next, we performed tests for symmetry (or zero skewness) of the data. We used the following tests: Cabilio-Masaro test (Cabilio and Masaro, 1996), Mira test (Mira, 1999) and the MGG test (Miao et al., 2006). All of them returned a \( p \) value less than 0.0001, giving strong evidence that the data are skewed.

The following model for the size of the lymphedema is considered:

\[
y_{ij} = \beta_{11} + \beta_{12} t_j + \beta_{13} t_i \text{meta}_i + \beta_{14} \text{marital}_i + \beta_{15} \text{age}_i + \beta_{16} \text{grad}_i + b_{1i} + b_{2i} t_j + \epsilon_{ij},
\]

where the \text{meta} variable denotes the Metastasis in the patient, \text{marital} denotes the marital state of the patient (single, married, widow, divorced), age refers to the age of the patient, \text{grad} denotes tumor’s degree of deterioration (low, average, high), and \( t_{ij} \) denotes the time of patient visiting the hospital. For the time-to-event occurrence, a Weibull distribution is considered: \( T_i | b_i \sim \text{Weibull}(\eta_i^T, \lambda) \) with

\[
\eta_i^T = \beta_{21} + \beta_{22} \text{meta}_i + \beta_{23} \text{marital}_i + \beta_{24} \text{age}_i + \beta_{25} \text{grad}_i + r_1 b_{1i} + r_2 b_{2i},
\]

where \( b_i = (b_{1i}, b_{2i}) \) denotes the shared random effect between the two models and assumed to have a skew-t distribution. To run the Bayesian model by considering the distributions introduced in Sections 3 and 4, two parallel Markov Chain Monte Carlo (MCMC) were conducted with different primary values and 100000 repetitions for each run. The first 30000 repetitions were eliminated and posterior inference was based on the remaining 70000 samples. In all of the models \( \beta_k \sim N_{p_k}(0, 10000 \mathbf{I}_{p_k}) \), where \( \mathbf{I}_{p-k} \) is a \( p_k \) dimensional identity matrix, \( r_k \sim N(0, 100) \), \( p_1 = 6 \), \( p_2 = 5 \).
and \( k = 1, 2 \). The prior distribution for the skew-t distribution degree of freedom parameter is \( \nu \sim U(1, 9) \). Moreover, the prior distribution for the skewness parameters is \( N(2, 1)I(\delta > 0) \).

Table 1 compares the Bayesian models, AIC and BIC indexes are given.

<table>
<thead>
<tr>
<th>Criterion value</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N-N</td>
</tr>
<tr>
<td>AIC</td>
<td>10453</td>
</tr>
<tr>
<td>BIC</td>
<td>10601</td>
</tr>
</tbody>
</table>

According to both AIC and BIC, the ST-ST model is the most appropriate one. Table 2 gives parameter estimates (standard deviation in brackets) and 95 percent confidence intervals for the ST-ST model.

Table 2: The Bayesian estimates of parameters (standard deviation) and 95 percent confidence intervals for the analysis of the breast-cancer data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard deviation</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{11} )</td>
<td>10.290</td>
<td>0.725</td>
<td>(8.933, 11.791)</td>
</tr>
<tr>
<td>( \beta_{12} )</td>
<td>-0.306</td>
<td>0.048</td>
<td>(-0.444, -0.244)</td>
</tr>
<tr>
<td>( \beta_{13} )</td>
<td>0.001</td>
<td>0.071</td>
<td>(-0.118, 0.163)</td>
</tr>
<tr>
<td>( \beta_{14} )</td>
<td>0.036</td>
<td>0.666</td>
<td>(-1.267, 1.269)</td>
</tr>
<tr>
<td>( \beta_{15} )</td>
<td>-0.326</td>
<td>0.046</td>
<td>(-4.240, -0.242)</td>
</tr>
<tr>
<td>( \beta_{16} )</td>
<td>0.186</td>
<td>0.265</td>
<td>(-0.318, 0.713)</td>
</tr>
<tr>
<td>( \beta_{21} )</td>
<td>-0.234</td>
<td>0.388</td>
<td>(-0.982, 0.542)</td>
</tr>
<tr>
<td>( \beta_{22} )</td>
<td>-5.207</td>
<td>0.636</td>
<td>(-6.221, -3.738)</td>
</tr>
<tr>
<td>( \beta_{23} )</td>
<td>0.453</td>
<td>0.292</td>
<td>(-0.147, 0.983)</td>
</tr>
<tr>
<td>( \beta_{24} )</td>
<td>0.028</td>
<td>0.657</td>
<td>(-1.279, 1.281)</td>
</tr>
<tr>
<td>( \beta_{25} )</td>
<td>2.321</td>
<td>0.371</td>
<td>(1.594, 3.075)</td>
</tr>
<tr>
<td>( r_1 )</td>
<td>-0.313</td>
<td>0.047</td>
<td>(-0.411, -0.225)</td>
</tr>
<tr>
<td>( r_2 )</td>
<td>-3.538</td>
<td>0.481</td>
<td>(-4.507, -2.544)</td>
</tr>
<tr>
<td>( \nu_1 )</td>
<td>4.487</td>
<td>0.461</td>
<td>(7.296, 8.987)</td>
</tr>
<tr>
<td>( \nu_2 )</td>
<td>3.987</td>
<td>0.652</td>
<td>(2.915, 5.441)</td>
</tr>
<tr>
<td>( \omega_1 )</td>
<td>0.452</td>
<td>0.281</td>
<td>(-1.145, 0.985)</td>
</tr>
<tr>
<td>( \omega_2 )</td>
<td>0.861</td>
<td>0.321</td>
<td>(0.161, 1.618)</td>
</tr>
</tbody>
</table>

The estimates in Table 2 can be interpreted as follows: a unit change in marital status increases the size of lymphedema by 0.036 and increases the time to event by 0.453; a unit increase in age decreases the size of lymphedema by 0.326 and increases the time to event by 0.028; a unit change in tumor’s degree of deterioration increases the size of lymphedema by 0.186 and increases the time to event by 2.321; and so on. The estimates of the following parameters do not appear significantly different from zero: \( \beta_{13}, \beta_{16}, \beta_{21}, \beta_{23}, \beta_{24} \) and \( \omega_1 \). The estimates of the following parameters appear significantly negative: \( \beta_{12}, \beta_{15}, \beta_{22}, r_1 \) and \( r_2 \). The estimates of the following parameters appear
significantly positive: $\beta_{11}, \beta_{25}, \nu_1, \nu_2$ and $\omega_2$.

The prior distributions were selected as $U(1,9)$ and $N(2,1)I(\delta > 0)$ for convenience. But other choices gave similar results in that the $ST - ST$ model always showed the best performance.

7 Conclusions

In the previous sections, simultaneous modeling of the longitudinal and survival data for the breast-cancer data was investigated under the assumptions of skewed and symmetric distributions. For the longitudinal data modeling, a linear-mixed model was used. For the survival data modeling, the Weibull proportional hazards model was used. To estimate the parameters, because of the form of the likelihood function, the maximum likelihood method could not be used directly. Furthermore, to increase the exactness of the estimates, the Bayesian method was used. By comparing different distributional assumptions for residuals and random effects, a model with the skew-t (ST) distribution assumption was proved to be the most efficient. This conclusion was also obtained in the simulation part.

Some future are work to apply other multivariate skew models to the data discussed in Section 6. Possible choices may include mixtures of multivariate skew-t distributions due to Lee and McLachlan (2014); the multivariate skew-normal generalized hyperbolic distribution due to Vilca et al. (2014); the multivariate Birnbaum-Saunders distribution due to Jamalizadeh and Kundu (2015); the multivariate skew-normal-Cauchy distribution due to Kahrari et al. (2016); the multivariate geometric skew-normal distribution due to Kundu (2017); scale and shape mixtures of multivariate skew-normal distributions due to Arellano-Valle et al. (2018); the skew mixed effects model due to Eftekhari Mahabadi and Rahimi Jafari (2018); the multivariate skew slash distribution due to Tian et al. (2018).

Acknowledgments

The authors would like to thank the Editor and the two referees for careful reading and comments which greatly improved the paper.

References


