

Bayesian Inference on Bivariate Semi-continuous Mixed-effects Models with Application to Longitudinal Substance Use Data

Dongyuan Xing, Yangxin Huang, Henian Chen, Yiliang Zhu, Getachew A. Dagne and Julie Baldwin

College of Public Health, University of South Florida, Tampa, Florida, United States
Email: yhuang@health.usf.edu

Abstract Multivariate (bivariate) correlated data encountered frequently in longitudinal studies are often analyzed using a multivariate linear mixed-effects model with normality assumption. Semi-continuous data in the form of a mixture of high proportion of zeros and right-skewed positive values bring special challenges to the field of multivariate modeling. In this paper, we propose a Bayesian approach to analyze bivariate semi-continuous outcomes by jointly modeling a generalized logistic mixed-effects model on zero-inflation in either response and a bivariate linear mixed-effects model (BLMM) on the positive values given both responses occurred through a correlated random-effects structure. Multivariate skew distributions including skew- t and skew-normal distributions are used to relax the normality assumption in BLMM. The proposed models are illustrated with an application to the correlated alcohol and drug uses data from a longitudinal observational study. A simulation study is conducted to evaluate the performance of the proposed models.

Keywords: Bayesian analysis, semi-continuous data, joint modeling, bivariate mixed-effects model, skew distributions.

1 Introduction

Semi-continuous data in the form of a mixture of zeros and skewed positive values arise frequently in economic, biological, and social science studies. Multivariate (bivariate) semi-continuous variables that are usually correlated are the primary outcomes in many cross-sectional and longitudinal studies. In a longitudinal study on substance abuse, for instance, the alcohol and illicit drug uses are often collected on the same subject simultaneously at each follow-up. The traditional approach for analyzing bivariate correlated continuous outcomes is via multivariate linear mixed-effect model (LMM) with normality assumption for the random-effects and within-subject errors. However, those analyses may not provide robust inference when data exhibit skewness particularly in the situation of semi-continuous data where the extra zeros and skewed nonzero values may generate an extreme non-normality in data. One common approach for analyzing data with skewness is to apply data transformation such as log-transformation to make the transformed response approximately Gaussian. Data transformation may not be appropriate in multivariate analysis because component-wise transformation does not lead to joint normality [1]. Besides, the transformations may not be universal, i.e. transformation used for one variable may not be adapted for the others. Moreover, results based on transformed data are typically hard to interpret and retransformation may not be straightforward [2]. This motivates the researchers to explore the alternative parametric distributions that have flexibility in distributional assumptions of random-effects and/or model errors to produce robust parameter estimates. Ghosh *et al.* [3] developed a bivariate linear mixed-effects model (BLMM) and relaxed the normality assumption by using multivariate skew-normal distributions and applied the method to HIV data, followed by more studies in the literature [4–6].

Two-part models originated in econometrics [7, 8] have been developed extensively in the last three decades to analyze zero-inflated count and semi-continuous data. In two-part modeling, we view the zero-inflated outcome variable as the result of two processes: one binomial process determining whether the positive value occurs and one continuous process determining the actual value if it is nonzero. Zero-inflated Poisson (ZIP) model and zero-inflated Negative Binomial (ZINB) model are generally used for zero-inflated count data. Duan *et al.* [8] developed a two-part model for semi-continuous outcome with a pair of

equations where a probit model was used for the probability of cost being positive and a linear model was used for the level of the positive values (log-transformed) and applied the model to a cross-sectional medical cost data. Olsen and Schafer [10], Tooze *et al.* [11] and Xing *et al.* [9] extended the model to longitudinal setting by introducing the correlated random-effects into each component of two-part models. Log-normal distribution has been the most commonly used distributional assumption for Part II of two-part semi-continuous model.

In the analysis of multivariate zero-inflated data, bivariate ZIP and bivariate ZINB models have been developed on the bivariate zero-inflated count data [12,13]. Gurmu and Dagne [14] developed a zero-inflated bivariate ordered probit regression model for the zero-inflated ordinal data in a cross-sectional setting. However, there is very little literature on bivariate semi-continuous models, particularly in longitudinal setting. Duan *et al.* [8] discussed a four-part model to distinguish inpatient and outpatient costs in the analysis of cross-section medical cost data. Liu *et al.* [15] extended the four-part model to the repeated measures of medical cost data by introducing a correlated random-effects structure and used likelihood approach for parameter estimation. Hatfield *et al.* [16,17] studied joint modeling of the survival time and multivariate zero-inflated PRO data where each PRO item assumed a zero-augmented beta (ZAB) distribution and estimated via a Bayesian approach. The multivariate semi-continuous outcomes were assumed to be independent conditional on the random-effects in these research [15–17].

Our research is motivated by a substance use data from a longitudinal observational study, in which alcohol abuse/dependence symptom (AADS) scales and marijuana abuse/dependence symptom (MADS) scales were measured simultaneously at each of four follow-up visits. Both AADS and MADS are zero-inflated followed by positive continuous values. We are interested in the temporal correlation between the trajectories of the alcohol and drug uses and their associations with the shared underlying demographic and social-economical characteristics. Theoretically, the joint distribution of two correlated semi-continuous variables could have the following categories: i) both responses are zero (i.e., neither symptoms), named as zero-zero state; ii) only AADS is zero (i.e., no alcohol abuse/dependence symptom but has marijuana abuse/dependence symptom) named as zero-one state; iii) only MADS is zero (i.e., no marijuana abuse/dependence symptom but has alcohol abuse/dependence symptom), named as one-zero state; and iv) both AADS and MADS are positive continuous (i.e., both symptoms occurred), named as one-one state. In our real data, zero-inflation occurred at the three states that contain zero in a total of 42.2% data. Consequently, we create an indicator variable for concurrence of both symptoms versus others (i.e., zero in either response). We consider the bivariate correlated data come from two data processes with one determining whether both responses are positive and one determining the intensity of the positive responses. It is possible that there are common underlying factors that drive both data processes and lead to the correlation between the probability of the concurrence of positive values in both responses and the intensity of the responses given they are positive. The correlation between the two data processes can be jointly modeled with a random-effects structure. In addition, the correlation among repeated measurements on the same subject and the correlation among different response variables should be considered in longitudinal multivariate analysis.

In this paper, we propose a flexible joint modeling approach to the following two types of models: i) a generalized logistic mixed-effects model on zero-inflation in either response; and ii) a BLMM given both responses are positive; while the two models were linked through a correlated random-effects structure. We relax the normality assumption in BLMM for the model errors with skew distributions including skew- t (ST) and skew-normal (SN) distributions in the family of skew-elliptical (SE) distribution [1,18]. A detailed review and related properties on different versions of multivariate skew distributions can be found in the literature [1,19–22]. In particular, Sahu *et al.* [21] developed Bayesian inference for skewed data using SN and ST distributions followed by Jara *et al.* [1]. The skew distribution family, which accommodates asymmetry in a more flexible manner, can model both positively and negatively skewed data (depending on the sign of the estimated skewness parameter from data), and includes normal and t distributions as special cases. We consider a multivariate ST distribution introduced by Sahu *et al.* [21] and Muthen and Asparouhov [22], which is suitable for a Bayesian inference. It is noted that the ST distribution is approximate to the SN distribution when its degrees of freedom approach infinity, the ST distribution reduces to a t distribution if skewness parameter is zero and the SN distribution reduces to a normal (N) distribution if skewness parameter is zero. Thus, we use an ST distribution to develop the

two-part model and associated statistical methodologies, as it can be easily reverts to the other cases such as the SN and N distributions. Three (N, SN and ST) distributions will be considered in this paper.

The rest of the paper proceeds as follows. In Section 2, we describe the data set that motivated this research, introduce BLMM with ST distributions and present the associated Bayesian inferential method. In Section 3, we apply the proposed method to the real data set described in Section 2 and report the analysis results. Section 4 conducts a simulation study aimed to evaluate the performance of the proposed models and method. We conclude the paper with discussion in Section 5.

2 Data Description and Joint Model with ST and SN Distributions

2.1 Motivating Data Set

The data set that motivated this research is from the Children in the Community (CIC) study (<http://nyspi.org/childcom>). The CIC study was a longitudinal study to evaluate the prevalence of various mental health problems, the risk factors that make such problems more likely, and the outcomes in adulthood. The study started in 1975 with one child per family randomly selected from 2 upper state New York counties. The sample's racial distribution (91% Caucasian, 8% African-American) and socioeconomic status (SES) (21% with family income below the poverty line and 25% with upper middle class educational and income) is similar to that of the entire USA in 1975. The study is one of the few studies that have conducted systematic, interview-based assessments of psychopathology over 30 years beginning with childhood interviews. AADS and MADS were assessed in 1983 (mean age of 13), 1986 (mean age of 16), 1993 (mean age of 23) and 2003 (mean age 33) consistently with Diagnosis and Statistical Manual of Mental Disorders (DSM) criteria [23]. Cohen *et al.* [24] studied the associations of AADS and MADS and baseline variables including gender, race, family SES and adolescent personality disorder in 1983 with LMMs on log-transformed scales. The analysis of this paper included 1266 repeated measures of AADS and MADS scales from 318 subjects (41% male).

The AADS scales ranged from 0 to 17 with a mean of 2.3 (SD=2.6) and a median of 1.0 (inter-quartiles: 0 - 3.5). The data contain 26.2% of zeros and the positive values are right-skewed even after taking log-transformation (Figure 1a, 1b). Similarly, MADS scales ranged from 0 to 9 with a mean of 0.7 (SD=0.9) and a median of 0.5 (inter-quartiles: 0.5 - 0.5) and 21.2% of zeros (Figure 1c, 1d). In the joint distribution of AADS and MADS, both AADS and MADS are zero in 5.3% of data, AADS being zero and MADS being > zero in 20.9%, MADS being zero and AADS being > zero in 16.0%, and both responses being positive in 57.8% of data (Table 1).

Table 1. Bivariate frequency distribution of AADS and MADS

AADS	MADS		Total (N)
	0	> 0	
0	67 (5.3%)	265 (20.9%)	332 (26.2%)
> 0	202 (16.0%)	732 (57.8%)	934 (73.8%)
Total (N)	269 (21.2%)	997 (78.8%)	1266 (100.0%)

2.2 Joint Models of a Logistic Mixed-effects Model on Zero-inflation and a Bivariate Linear Mixed-effects Models with ST Distribution

Let $y_{ij}^{(a)}$ be the measurement of AADS from the i th ($i = 1, \dots, n$) subject at j th time point ($j = 1, \dots, n_i$) and $y_{ij}^{(m)}$ be the measurement of MADS defined similarly. An indicator variable denoting both responses being positive versus at least one being zero is defined by

$$r_{ij} = \begin{cases} 1 & \text{if } y_{ij}^{(a)} > 0 \text{ and } y_{ij}^{(m)} > 0 \\ 0 & \text{otherwise} \end{cases}$$

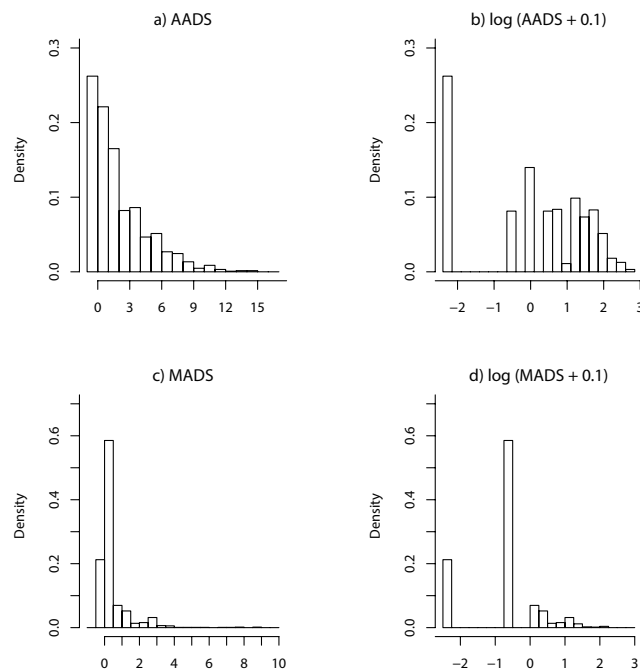


Figure 1. Histograms of AADS (a,b) and MADS (c, d)

We model probability with $r_{ij} = 1$ using the following generalized (logistic) mixed-effects model.

$$\text{logit}(r_{ij} = 1) = \log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \mathbf{x}_{ij}\boldsymbol{\alpha} + \mathbf{z}_{ij}\mathbf{a}_i, \tag{1}$$

where $P_{ij} = Pr(r_{ij} = 1|\boldsymbol{\alpha}, \mathbf{a}_i)$ is the probability of both responses being positive conditional on random-effects \mathbf{a}_i , $\mathbf{x}_{ij}(1 \times p)$ is the vector of covariates including time which may be associated with the probability of occurrence of $r_{ij} = 1$ and $\mathbf{z}_{ij}(1 \times q)$ is the design matrix corresponding to random-effects \mathbf{a}_i which may include random intercept, slope and so on to vary by subject.

Let $\mathbf{x}_{ij}^{(a)}$ and $\mathbf{x}_{ij}^{(m)}$ be the $1 \times p^*$ vectors of covariates associated with the fixed-effects $\boldsymbol{\beta}^{(a)}$ and $\boldsymbol{\beta}^{(m)}$ of the two responses, respectively, and $\mathbf{z}_{ij}^{(a)}$ and $\mathbf{z}_{ij}^{(m)}$ be the $1 \times q^*$ vectors of covariates associated with the random-effects $\mathbf{b}_i^{(a)}$ and $\mathbf{b}_i^{(m)}$ of the two responses, respectively; and the fixed and random effects may or may not be identical to those effects in model (1). To make notations more compact, given $r_{ij} = 1$, let $\mathbf{y}_{ij} = (y_{ij}^{(a)}, y_{ij}^{(m)})^T$, $\mathbf{x}_{ij}^* = \text{diag}(\mathbf{x}_{ij}^{(a)}, \mathbf{x}_{ij}^{(m)})$, $\mathbf{z}_{ij}^* = \text{diag}(\mathbf{z}_{ij}^{(a)}, \mathbf{z}_{ij}^{(m)})$, $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(a)T}, \boldsymbol{\beta}^{(m)T})^T$, $\mathbf{b}_i = (\mathbf{b}_i^{(a)T}, \mathbf{b}_i^{(m)T})^T$ and $\mathbf{e}_{ij} = (e_{ij}^{(a)}, e_{ij}^{(m)})^T$ where $e_{ij}^{(a)}$ and $e_{ij}^{(m)}$ are the within-subject residuals for two responses, respectively and $\text{diag}(\mathbf{A}, \mathbf{B})$ denotes a block diagonal matrix. The bivariate continuous responses given $r_{ij} = 1$ can be modeled by a BLMM with model error following multivariate ST distribution defined as below.

$$\mathbf{y}_{ij}|r_{ij}=1 = \mathbf{x}_{ij}^*\boldsymbol{\beta} + \mathbf{z}_{ij}^*\mathbf{b}_i + \mathbf{e}_{ij}, \quad \mathbf{e}_{ij} \sim ST_{2,\nu}(-J(\nu)\boldsymbol{\delta}, \boldsymbol{\Sigma}, \boldsymbol{\delta}), \tag{2}$$

where $J(\nu) = (\nu/\pi)^{1/2} \{\Gamma[(\nu - 1)/2]/\Gamma(\nu/2)\}$ where $\Gamma(\cdot)$ is a gamma function. The model errors \mathbf{e}_{ij} follow a multivariate ST distribution with degrees of freedom ν and a variance-covariance structure $\boldsymbol{\Sigma} = (\sigma_{ll'}^2)_{2 \times 2}$, ($l, l' = 1, 2$); and $\boldsymbol{\delta} = (\delta^{(a)}, \delta^{(m)})^T$ with elements denoting the skewness parameters of AADS and MADS, respectively.

The random-effects in equations (1) and (2) follow a multivariate normal distribution with unknown variance-covariance matrix of $\boldsymbol{\Psi}$:

$$\mathbf{c}_i = (\mathbf{a}_i^T, \mathbf{b}_i^T)^T \sim N(\mathbf{0}, \boldsymbol{\Psi}).$$

In this paper, we include random intercept in each component of equations (1) and (2) such as a_{i0} in equation (1) and $b_{i0}^{(a)}$ and $b_{i0}^{(m)}$ for AADS and MADS, respectively, in equation (2). The random-effects

follow a 3-dimensional multivariate normal distribution with unknown variance-covariance matrix Ψ as follows.

$$\mathbf{c}_i = \begin{pmatrix} a_{i0} \\ b_{i0}^{(a)} \\ b_{i0}^{(m)} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Psi = \begin{pmatrix} \sigma_{a_0}^2 & \sigma_{a_0 b_0^{(a)}}^2 & \sigma_{a_0 b_0^{(m)}}^2 \\ \sigma_{a_0 b_0^{(a)}}^2 & \sigma_{b_0^{(a)}}^2 & \sigma_{b_0^{(a)} b_0^{(m)}}^2 \\ \sigma_{a_0 b_0^{(m)}}^2 & \sigma_{b_0^{(a)} b_0^{(m)}}^2 & \sigma_{b_0^{(m)}}^2 \end{pmatrix} \right). \quad (3)$$

It is easy to extend the model to include more random-effects, for example, random slope in each component of BLMM. The conditional probability density function (pdf) for bivariate responses \mathbf{y}_{ij} is expressed as

$$\begin{aligned} f(\mathbf{y}_{ij} | \mathbf{a}_i, \mathbf{b}_i) &= [1 - P_{ij}]^{1-r_{ij}} \times [P_{ij} \times ST_{2,\nu}(\mathbf{y}_{ij} | \mathbf{b}_i)]^{r_{ij}} \\ &= \{[1 - P_{ij}]^{1-r_{ij}} P_{ij}^{r_{ij}}\} \times [ST_{2,\nu}(\mathbf{y}_{ij} | \mathbf{b}_i)]^{r_{ij}}. \end{aligned} \quad (4)$$

2.3 Simultaneous Bayesian Inferential Approach for Joint Modeling

We propose a Bayesian inferential method via Markov chain Monte Carlo (MCMC) procedure to estimate the parameters in equations (1) and (2). Following the studies by Sahu et al. [21] and Muthen and Asparouhov [22], it can be shown that, by introducing a random vector $\mathbf{w}_{ij} = (w_{ij}^{(a)}, w_{ij}^{(m)})^T$ with $w_{ij}^{(a)}$ and $w_{ij}^{(m)}$ being two random variables based on the stochastic representation for the ST distribution, equation (2) can be hierarchically formulated as

$$\begin{aligned} \mathbf{y}_{ij} | \mathbf{b}_i, \mathbf{w}_{ij} &\sim T_{2,\nu}(\mathbf{x}_{ij}^* \boldsymbol{\beta} + \mathbf{z}_{ij}^* \mathbf{b}_i + \boldsymbol{\delta}[\mathbf{w}_{ij} - J(\nu) \mathbf{1}_2], \boldsymbol{\Sigma}), \\ \mathbf{w}_{ij} &\sim T_{2,\nu}(\mathbf{0}, \mathbf{I}_2) I(\mathbf{w}_{ij} > \mathbf{0}), \end{aligned} \quad (5)$$

where $\mathbf{1}_2 = (1, 1)^T$, $I(\mathbf{w}_{ij} > \mathbf{0})$ is an indicator function and \mathbf{w}_{ij} is truncated in the space $\mathbf{w}_{ij} > \mathbf{0}$ (i.e., standard half- t distribution).

Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \Psi, \nu, \boldsymbol{\delta})$ be the collection of unknown population parameters in equations (1), (3) and (5). To complete the Bayesian formulation, we specify prior distributions for all the unknown population parameters as follows.

$$\begin{aligned} \boldsymbol{\alpha} &\sim N_p(\boldsymbol{\alpha}_0, \boldsymbol{\Lambda}_1), \boldsymbol{\beta} \sim N_{2p^*}(\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_2), \boldsymbol{\Sigma} \sim IW(\boldsymbol{\Omega}_1, \omega_1), \\ \boldsymbol{\Psi} &\sim IW(\boldsymbol{\Omega}_2, \omega_2), \boldsymbol{\delta} \sim N_2(\boldsymbol{\delta}_0, \boldsymbol{\Lambda}_3), \nu \sim Exp(\omega_3) I(\omega_3 > 2), \end{aligned} \quad (6)$$

where the mutually independent Normal (N), Inverse Gamma (IG), Inverse Wishart (IW) and exponential (Exp) distributions are chosen to facilitate computations.

We assume the same design matrices in equations (1) and (2) (i.e., $\mathbf{x}_{ij} = \mathbf{x}_{ij}^*$, $\mathbf{z}_{ij} = \mathbf{z}_{ij}^*$) and let the observed data $\mathcal{D} = \{(r_{ij}, \mathbf{y}_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}); i = 1, \dots, n; j = 1, \dots, n_i\}$, $f(\cdot | \cdot)$ be a conditional density function and $h(\cdot)$ be a prior density function. We assume that $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \nu, \boldsymbol{\Psi}$ and $\boldsymbol{\delta}$ are independent of each other; in other words, $h(\boldsymbol{\theta}) = h(\boldsymbol{\alpha})h(\boldsymbol{\beta})h(\boldsymbol{\Sigma})h(\nu)h(\boldsymbol{\delta})h(\boldsymbol{\Psi})$. After we specify the models for the observed data and prior distributions of the unknown model parameters, we can draw samples for the parameters based on their posterior distributions under the Bayesian framework. Thus, the joint posterior density of $\boldsymbol{\theta}$, conditional on \mathcal{D} , can be given by

$$f(\boldsymbol{\theta} | \mathcal{D}) \propto \left\{ \prod_{i=1}^n \int \prod_{j=1}^{n_i} p(r_{ij} | \mathbf{a}_i) [f(\mathbf{y}_{ij} | \mathbf{b}_i, \mathbf{w}_{ij}) f(\mathbf{w}_{ij} | \mathbf{w}_{ij} > \mathbf{0})]^{r_{ij}} f(\mathbf{a}_i, \mathbf{b}_i) d\mathbf{a}_i d\mathbf{b}_i \right\} h(\boldsymbol{\theta}), \quad (7)$$

where $p(r_{ij} | \mathbf{a}_i) = (1 - P_{ij})^{1-r_{ij}} P_{ij}^{r_{ij}}$.

In general, the integral in equation (7) is of high dimension and does not have a closed form. Analytic approximations to the integrals may not be sufficiently accurate. Therefore, it is prohibitive to directly calculate the posterior distribution of $\boldsymbol{\theta}$ based on the observed data \mathcal{D} . As an alternative, the MCMC procedure can be used to sample population parameters, $\boldsymbol{\theta}$, and random-effects $(\mathbf{a}_i^T, \mathbf{b}_i^T)^T$, from conditional posterior distributions, based on equation (7) using the Gibbs sampler along with Metropolis-Hastings (M-H) algorithm. In particular, Gibbs sampling is used to iteratively generate updates for parameters whose full conditional distributions are available [3, 21, 25, 26] and the M-H algorithm is used to update

parameters where the full conditional distributions can not be expressed explicitly given the remaining parameters and data [26]. Convergence is monitored by running multiple chains with dispersed initial values and standard tools within WinBUGS software [27] such as Gelman-Rubin (GR) diagnostics [28], autocorrelation and trace plots. The advantage of the above representation based on hierarchical models is that they can be very easily implemented using the freely available WinBUGS software.

3 Substance Use Data Analysis

3.1 Specific Model and Implementation

We illustrate proposed models and method by applying them to the AADS and MADS data described in Section 2. To evaluate the impact of skewness on model estimation, the models with ST and SN distributions are compared with the model with normal distribution.

- **ST Model:** Model with ST distribution for model errors.
- **SN Model:** Model with SN distribution for model errors.
- **N Model:** Model with normal distribution for model errors.

As we described in Section 2, zero-inflation in our real data occurred in 42.2% of data at three states (zero-zero: 5.3%, zero-one: 20.9% and one-zero: 16.0%) (Table 1). We are interested in evaluating the association of AADS/MADS and the baseline variables including linear and quadratic age, gender, and the interaction between gender and age, family SES and race based on a previous publication on the same data [24]. The same set of baseline variables were used in different parts of joint modeling so that we can assess whether there are differentiated effects of covariates on the probability of concurrence of AADS and MADS and severity of the responses given the symptoms occurred. To compare the performance of models with different distributional assumptions, the same set of covariates are included in all models. The centered age variable (i.e., age-mean age) is used in models such that the intercept coefficient estimate can be interpreted as AADS/MADS scale at the mean age. The corresponding parameters are α_0 , $\beta_0^{(a)}$ and $\beta_0^{(m)}$ for intercept, α_1 , $\beta_1^{(a)}$ and $\beta_1^{(m)}$ for age, α_2 , $\beta_2^{(a)}$ and $\beta_2^{(m)}$ for quadratic age, α_3 , $\beta_3^{(a)}$ and $\beta_3^{(m)}$ for gender, α_4 , $\beta_4^{(a)}$ and $\beta_4^{(m)}$ for the interaction between age and gender, α_5 , $\beta_5^{(a)}$ and $\beta_5^{(m)}$ for race, and α_6 , $\beta_6^{(a)}$ and $\beta_6^{(m)}$ for family SES.

In the absence of historical data/experiment, we specify practical weakly informative priors for all model parameters. In particular, (i) The fixed-effects are taken to be independent normal distribution $N(0, 100)$ for each component of the population parameter vectors α and β . (ii) The prior for the variance-covariance matrices of the model errors Σ is taken to be inverse Wishart distribution with covariance matrix $\Omega_1 = \text{Diag}(1, 1)$ and degrees of freedom $\omega_1 = 3$. (iii) The prior for the variance-covariance matrix of the random-effects Ψ is taken to be inverse Wishart distribution with covariance matrix $\Omega_2 = \text{Diag}(1, 1, 1)$ and degrees of freedom $\omega_2 = 4$. (iv) For skewness parameters $\delta^{(a)}$ and $\delta^{(m)}$, independent normal distribution $N(0, 100)$ is used to accommodate either positive or negative skewness and allows the data to determine it. (v) The degrees of freedom parameter ν follows a truncated exponential distribution, $\nu \sim \text{Exp}(0.1)I(\nu > 3)$.

The MCMC sampler was implemented using WinBUGS software [27] and the program code is available from the corresponding author upon request. When the MCMC procedure was applied to the substance use data, convergence of the generated samples was assessed using standard tools within WinBUGS software such as Gelman-Rubin (GR) diagnostics [28]. Figure 2 shows the dynamic version of GR diagnostic plots, the trace plots and autocorrelation plots based on Model ST for the representative parameters α_2 , $\beta_2^{(a)}$ and $\delta^{(a)}$. For the plots of GR diagnostics (top panel) three curves are given: the middle and bottom curves below the dashed horizontal line (indicating the value one) represent the pooled posterior variance (\hat{V}) and average within-sample variance (\hat{W}), respectively, and the top curve represents their ratio (\hat{R}). It is seen that \hat{R} is generally expected to be higher than one at the initial stage of the algorithms, but \hat{R} tends to 1, and \hat{V} and \hat{W} stabilize as the number of iterations increases, indicating that the algorithm has approached convergence. The trace plots (middle panel) show that the lines of three different chains mix or cross in trace, implying that convergence is reached. We further monitor convergence using autocorrelation plots (bottom panel) and find that autocorrelations are very low with a lag being 50, implying that

convergence is obtained. When these criteria suggested the convergence of chains, we proposed that the three chains were run with the following considerations. For each model, we ran three, initially dispersed chains for 100,000 iterations each, discarding the first 50,000 as a burn-in period, and retained every 50th sample for a total of 3,000 samples of targeted posterior distributions of the unknown parameters to make inference. Note that in comparison with maximum likelihood methods, our Bayesian approach via MCMC procedure caused less convergence problems and the computational burden was reasonable. For example, to fit ST Model, it took about 20 minutes on a Windows workstation with Intel Core 2 Quad CPU @ 2.83GHz and RAM of 8.0GB.

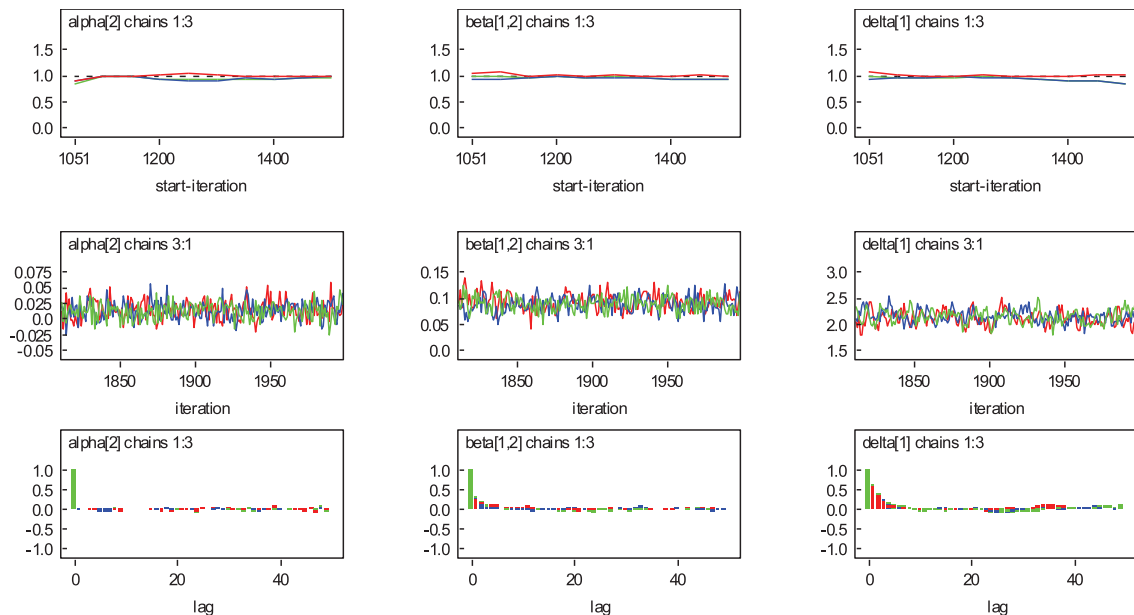


Figure 2. Convergence diagnostics with three Markov chains as obtained from the WinBUGS software for representative parameters based on Model ST: Gelman-Rubin (GR) diagnostic plots (top panel), trace plots (middle panel), and autocorrelation plots (bottom panel).

3.2 Comparison of Model Results

We conducted the following scenarios. First, we investigated how asymmetric (ST and SN) distributions for model errors (Models ST and SN) impact parameter estimation in comparison with a symmetric (normal) distribution for model error (Model N). Second, for the best model obtained above, we will report the results in detail. For selecting the best model that fits the data adequately, the Deviance Index Criteria (DIC) proposed by Spiegelhalter *et al.* [29] was used. The DIC is defined as $DIC = \bar{D}(\theta) + p_D = 2\bar{D}(\theta) - D(\hat{\theta})$, where $\bar{D}(\theta) = E[D(\theta)|y]$ is the posterior mean of the deviance with a smaller value indicating better fit, and $p_D = \bar{D}(\theta) - D(\hat{\theta}) = E[D(\theta)|y] - D[E(\theta|y)]$ is the effective number of parameters, defined as the difference in the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters. The structure of DIC allows for automatic computation in WinBUGS software. As with other model selection criteria, we caution that DIC is not intended for identification of the "correct" model, but rather merely as a method of comparing a collection of alternative formulations.

Bayesian modeling approach was used to fit the data. DIC, the posterior mean (PM), the corresponding standard deviation (SD) and 95% credible interval (CI) for fixed-effects parameters, skewness parameters, df and variance-covariance matrices based on Models N, SN and ST are presented in Table 2. The following findings are observed based on the modeling results.

Table 2. Summary of DIC, the estimated posterior mean (*PM*) of parameters of population (fixed-effects), skewness, degrees of freedom, variance-covariance matrix lower limit (*L_{CI}*) and upper limit (*U_{CI}*) of 95% equal-tail credible interval (CI) based on Model N, SN and ST.

DIC	Model N			Model SN			Model ST		
	6540			4764			4233		
	<i>PM</i>	<i>L_{CI}</i>	<i>U_{CI}</i>	<i>PM</i>	<i>L_{CI}</i>	<i>U_{CI}</i>	<i>PM</i>	<i>L_{CI}</i>	<i>U_{CI}</i>
Pr(both symptoms)									
α_0	1.136	0.227	2.027	1.167	0.303	2.082	1.296	0.401	2.214
α_1	0.015	-0.012	0.044	0.014	-0.014	0.041	0.014	-0.013	0.041
α_2	-0.023	-0.027	-0.020	-0.023	-0.027	-0.020	-0.023	-0.027	-0.020
α_3	0.734	0.314	1.170	0.752	0.311	1.177	0.762	0.333	1.196
α_4	0.034	-0.005	0.074	0.035	-0.002	0.073	0.034	-0.002	0.071
α_5	0.511	-0.351	1.399	0.476	-0.423	1.330	0.336	-0.535	1.226
α_6	0.003	-0.204	0.206	-0.028	-0.241	0.182	-0.037	-0.242	0.168
AADS									
$\beta_0^{(a)}$	2.878	2.029	3.757	3.167	2.535	3.727	3.114	2.525	3.696
$\beta_1^{(a)}$	0.142	0.104	0.181	0.061	0.024	0.102	0.090	0.059	0.123
$\beta_2^{(a)}$	-0.020	-0.023	-0.016	-0.010	-0.014	-0.006	-0.013	-0.017	-0.010
$\beta_3^{(a)}$	1.160	0.738	1.566	0.636	0.336	0.947	0.691	0.388	1.006
$\beta_4^{(a)}$	0.055	0.003	0.107	0.040	-0.001	0.084	0.035	-0.004	0.075
$\beta_5^{(a)}$	0.631	-0.204	1.474	0.308	-0.241	0.932	0.409	-0.143	0.960
$\beta_6^{(a)}$	0.003	-0.187	0.193	-0.052	-0.185	0.079	-0.054	-0.192	0.078
$\delta_1^{(a)}$.	.	.	3.565	3.292	3.834	2.126	1.812	2.419
MADS									
$\beta_0^{(m)}$	0.665	0.286	1.043	1.138	0.871	1.332	0.810	0.684	0.925
$\beta_1^{(m)}$	0.007	-0.010	0.026	0.000	-0.009	0.009	-0.001	-0.007	0.004
$\beta_2^{(m)}$	-0.003	-0.004	-0.001	0.000	-0.001	0.000	0.000	-0.001	0.000
$\beta_3^{(m)}$	0.131	-0.057	0.318	0.038	-0.041	0.122	0.026	-0.026	0.078
$\beta_4^{(m)}$	0.015	-0.010	0.039	0.005	-0.008	0.017	0.003	-0.004	0.010
$\beta_5^{(m)}$	0.290	-0.074	0.657	0.091	-0.084	0.350	0.034	-0.067	0.143
$\beta_6^{(m)}$	0.026	-0.062	0.116	0.002	-0.037	0.042	0.003	-0.020	0.028
$\delta_2^{(m)}$.	.	.	1.313	1.243	1.387	0.476	0.426	0.532
ν (df)	3.036	3.001	3.136
Σ (model errors)									
σ_{11}^2	4.165	3.625	4.782	0.207	0.086	0.411	0.280	0.131	0.505
σ_{22}^2	0.895	0.793	1.009	0.020	0.015	0.028	0.011	0.009	0.014
σ_{12}^2	0.667	0.482	0.860	0.005	-0.012	0.025	0.007	-0.004	0.020

The estimates of the vector α are similar across three models. The estimates of the vector β are similar between Models SN and ST and are smaller than that from Model N. The estimates of the skewness parameters $\delta^{(a)}$ and $\delta^{(m)}$ are significantly positive from Models SN and ST providing evidence of high to moderate right-skewness of our data. The estimates (95% CI) of $\delta^{(a)}$ are 2.126 (1.812, 2.419) and 3.565 (3.292, 3.834) from Models ST and SN, respectively, and the estimates (95% CI) of $\delta^{(m)}$ are 0.476 (0.426, 0.532) and 1.313 (1.243, 1.387) from Models ST and SN, respectively. Figure 3 shows the boxplots for the skewness parameters, $\delta^{(a)}$ and $\delta^{(m)}$ based on 3,000 posterior samples of Models ST and SN. Note that the 95% CI does not include zero for both models, confirming that positive asymmetry. Thus, incorporating a skewness parameter in the modeling of the data is recommended. We also find that the skewness for AADS is significantly higher than MADS. The posterior mean estimates of ν for ST model is 3.036 with 95% CI (3.001, 3.136), which indicates a certain level of heavy tail in data.

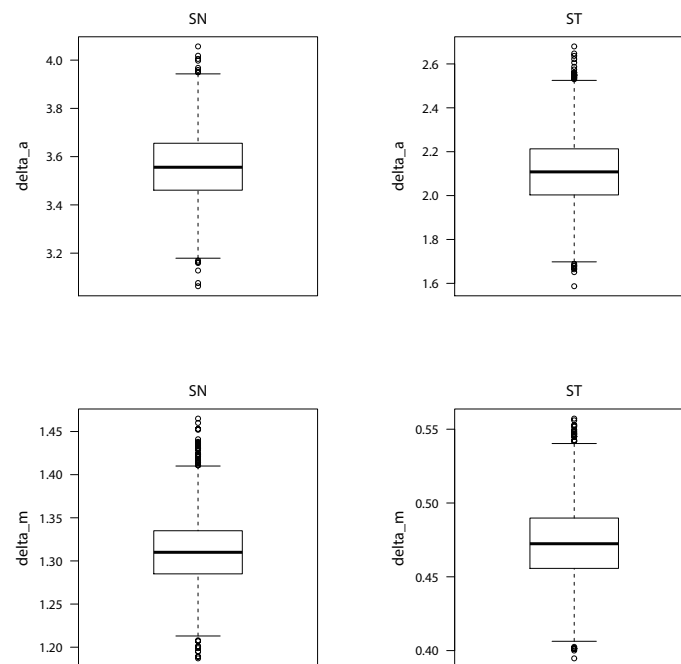


Figure 3. Boxplots of asymmetry parameters for Models ST and SN. The upper and lower panels are boxplots for $\delta^{(a)}$ and $\delta^{(m)}$, respectively.

The estimates of within-subject variances σ_{11}^2 , σ_{22}^2 and covariance σ_{12}^2 for the two asymmetric Models (SN and ST) are smaller than those of Model N. This is expected because high variability, heaviness of the tails and skewness are interrelated to a certain extent.

The DIC values are 6540, 4764 and 4233 for Models N, SN and ST, respectively (Table 2). Based on DIC, Model ST is the best-fitting model, followed by Model SN, supporting the contention of a departure from normality.

3.3 Estimation Results Based on ST Model

Based on DIC, the best fit model is the ST model, in which model errors are assumed to have ST distribution. In Model ST, the estimated skewness parameters for AADS and MADS are 2.126 and 0.476, respectively (Table 2). Because the 95% CIs for both skewness parameters do not include zero, this confirms the positive skewness observed from the data. As indicated in Table 2, there is a positive association between linear age and the odds of concurrence of AADS and MADS, although not significant ($\alpha_1^{(a)}=0.014$ with 95% CI (-0.013, 0.041)). There is a significant negative association between quadratic age and the odds, indicating that the odds of concurrence of both symptoms decrease after subject reaches certain age ($\alpha_2 = -0.023$ with 95% CI (-0.027, -0.020)). Male gender is associated with higher odds of concurrence of AADS and MADS determined by $\alpha_3 = 0.762$ with 95% CI (0.333, 1.196), which is equivalent to an odds ratio of 2.147 for male versus female with 95% CI (1.40, 3.307). The odds of concurrence of AADS and MADS are not significantly associated with family SES or race.

The severity of AADS given both symptoms occurred increases with growing age ($\beta_1^{(a)}=0.090$ with 95% CI (0.059, 0.123)) and decreases after subjects reach certain age determined by the negative parameter estimate for quadratic age ($\beta_2^{(a)}=-0.013$ with 95% CI (-0.017, -0.010)). There is a significant association between male gender and the severity of AADS ($\beta_3^{(a)}=0.691$ with 95% CI (0.388, 1.006)). Male gender is associated with a larger rate of change in severity of AADS determined by $\beta_4^{(a)}=0.035$ with 95% CI (-0.004, 0.075) although the p-value may be slightly above 0.05.

There is a trend for the association between male gender and increased MADS, although it is not significant (0.026 with 95% CI (-0.026, 0.078)). The severity of MADS is not significantly associated with any other covariates in Model ST.

4 Simulation Study

In this section we conduct a simulation study under longitudinal setting to evaluate the performance of Models ST, SN and N and associated method. The simulated data were generated following the algorithm below.

$$\begin{aligned}
 \text{logit}[Pr(r_{ij} = 0)] &= \alpha_0 + \alpha_1 t_{ij} + \alpha_2 x_i + a_{i0}, \\
 y_{ij}^{(a)} | r_{ij} = 1 &= \beta_0^{(a)} + \beta_1^{(a)} t_{ij} + \beta_2^{(a)} x_i + b_{i0}^{(a)} + e_{ij}^{(a)}, \\
 y_{ij}^{(m)} | r_{ij} = 1 &= \beta_0^{(m)} + \beta_1^{(m)} t_{ij} + \beta_2^{(m)} x_i + b_{i0}^{(m)} + e_{ij}^{(m)}, \\
 (a_{i0}, b_{i0}^{(a)}, b_{i0}^{(m)})^T &\sim N(\mathbf{0}, \Psi_{3 \times 3}),
 \end{aligned}
 \tag{8}$$

where the measurement time points are given as $t_{ij} = 0, 0.1, 0.2, 0.3, 0.5, 0.75, 1, 2, 3, 4$ and 5 ; a binary covariate $x_i = 0$ or 1 ; the true values of parameter vector α are chosen as $(\alpha_0, \alpha_1, \alpha_2)^T = (-2.4, 1.0, 2.0)$ such that the percentage of zeros is approximately 50%; the true values of parameter vector $(\beta_0^{(a)}, \beta_1^{(a)}, \beta_2^{(a)})^T = (1.8, 2.0, 2.0)^T$ and $(\beta_0^{(m)}, \beta_1^{(m)}, \beta_2^{(m)})^T = (2.0, 1.0, 1.0)^T$, the true values are 0.09 on the diagonal of Ψ and 0.045 for off-diagonal which means that the variance of random-effects is 0.09 and correlation between random-effects is 0.5. We generated $e_{ij}^{(a)} = \varepsilon_{ij} - 1$, in which ε_{ij} follows a gamma distribution Gamma(1,1), yielding a skewed distribution with $E(e_{ij}) = 0$ and $var(e_{ij}) = 1$. The generated data $y_{ij}^{(a)}$ ranged approximately from 0 to 20. Similarly, we generated $e_{ij}^{(m)} = \varepsilon_{ij} - 0.5$, in which ε_{ij} follows a gamma distribution Gamma(0.5,1), yielding a skewed distribution with $E(e_{ij}) = 0$ and $var(e_{ij}) = 0.5$. The generated $y_{ij}^{(m)}$ ranged approximately from 0 to 10. We generated trajectories of 500 individuals with the first 250 subjects assigned to $x_i = 0$ and others to $x_i = 1$. The weakly informative prior distributions used in real data are adopted here.

Table 3. Summary of Monte Carlo simulation results for MC estimates of fixed-effects as well as bias and MSE for Models ST, SN and N based on 100 simulated datasets.

Model		α_0	α_1	α_2	$\beta_0^{(a)}$	$\beta_1^{(a)}$	$\beta_2^{(a)}$	$\beta_0^{(m)}$	$\beta_1^{(m)}$	$\beta_2^{(m)}$
	True value	-2.40	1.00	2.00	1.80	2.00	2.00	2.00	1.00	1.00
ST	MC mean	-2.4304	1.0098	2.0309	1.8211	1.9993	2.0020	1.9626	0.9994	1.0080
	Bias	-1.2659	0.9800	1.5450	1.1739	-0.0374	0.0987	-1.8705	-0.0572	0.8042
	MSE	3.9243	3.4941	4.7861	2.2695	0.2320	1.6435	2.3032	0.2106	3.3147
SN	MC mean	-2.4320	1.0109	2.0273	1.8762	1.9983	2.0159	2.0770	0.9994	1.0339
	Bias	-1.3315	1.0864	1.3651	4.2320	-0.0864	0.7931	3.8475	-0.0582	3.3860
	MSE	3.5492	3.4773	4.3018	4.6494	0.2520	1.6971	4.1581	0.2814	4.3940
N	MC mean	-2.4215	1.0082	2.0245	1.8126	1.9960	1.9932	2.0082	0.9988	0.9930
	Bias	-0.8966	0.8174	1.2258	0.7022	-0.1993	-0.3400	0.4100	-0.1195	-0.7000
	MSE	3.7588	3.4107	4.6070	2.8589	0.5629	2.4112	1.8727	0.8096	3.8939

Notes: MC mean=Monte Carlo mean
 Bias=Percent bias = $100 \times \text{bias}_i / |\beta_i|$.
 MSE=Percent $\sqrt{MSE} = 100 \times \sqrt{MSE}_i / |\beta_i|$.

We are interested in assessing the impact of distributional assumptions on the estimate of the vector β . Models ST, SN and N are fit to simulated data set with 100 repetitions each. In each repetition, we ran a chain of 20,000 iterations with 10,000 discarded as burn-in. Every 10th MCMC sample is retained for inference. Table 3 summarizes simulation results which include the MC mean of fixed-effects, as well as associated percent bias (defined by $100 \times \text{bias}_l / |\beta_l|$) and percent mean-square-errors (MSE), defined by $100 \times \sqrt{MSE_l} / |\beta_l|$ ($l = 0, 1, 2$). The parameter estimates on α_0, α_1 and α_2 are similar across models. The proposed model ST provides the most accurate estimate of $\beta_1^{(a)}$ (time measurements) with a percent bias of -0.0374%, followed by Model SN (-0.086%) and Model N (-0.1993%); similar pattern was observed for $\beta_1^{(m)}$. The three models provide similar accuracy on the estimates of $\beta_2^{(a)}$ and $\beta_2^{(m)}$ for the binary covariate. Our simulation results suggest that models with ST and SN distributions are more powerful in handling the zero-inflated continuous data with skewed feature compared with model with normal distribution.

5 Concluding Discussion

In this paper, we propose a joint modeling approach to analyze the correlated bivariate semi-continuous outcomes in a longitudinal setting under a Bayesian framework. In bivariate semi-continuous data, the zero-inflation may occur at any or all of three states: zero-zero, zero-one and one-zero. Theoretically, we can develop the specific model for each type of zero-inflation. We can create one categorical (or binomial) variable to describe the zero-inflation status and model it with an appropriate generalized linear model. For example, if zero-inflation occurs at the zero-zero state only, a binomial variable r_{ij} denoting either or both symptoms versus no symptoms can be generated and analyzed with logistic mixed model. In the situation that zero-inflation occurs at all three states, we can create an ordinal variable r_{ij} representing no symptoms, one symptom versus both symptoms and analyze r_{ij} with a proportional odds mixed model. Similarly, a categorical variable r_{ij} can be created to denote no symptoms, alcohol symptom only, marijuana symptom only, and both symptoms and analyzed with multinomial baseline-logit model. The bivariate continuous responses conditional on $r_{ij} = 1$ can be modeled with a BLMM in which we propose to use a multivariate skew distribution to account for the skewness in the joint distribution. The generalized mixed model for r_{ij} and the BLMM are then jointly modeled through a correlated random-effects structure. Therefore, the correlation between the correlated variables is accounted for by both the random-effects structure and the variance-covariance structure of model errors. The joint modeling of proportional odds mixed model or multinomial mixed model with BLMM is more complicated and involves extra equations and parameter estimations. This topic will be explored in our future work.

We carried out sensitivity analyses to examine the dependence of parameter estimates on the choices of hyper-parameters in the (non-informative) prior distributions and found that the parameter estimates from proposed models were reasonable and robust under different hyper-parameter values. To our best knowledge, there is very limited research on multivariate semi-continuous models. The four-part modeling developed by Duan *et al.* [8] and Liu *et al.* [15] used random-effects for the correlation between the two correlated bivariate responses and assumed independence between model errors. The independence in model errors of two response variables indicates that there is no correlation between the bivariate responses conditional on the random-effects. This is a very strong assumption. Fieuw and Verbeke [30] found that this random-effects approach may yield misleading results in evaluating the relationship of the correlated bivariate responses. Our proposed models utilized both the random-effects structure and variance-covariance of model errors to account for the bivariate correlation. Relaxing the normality assumption in BLMM with multivariate ST/SN distribution reduces bias and improves efficiency in parameter estimation.

Acknowledgments. The authors gratefully acknowledge the Editor and two anonymous referees for their insightful comments and helpful suggestions that led to an improvement of the article. This research was partially supported by University South Florida Creative Scholarship Grant 18324 to Y. Huang.

References

1. A. Jara, F. Quintana and E.S. Martin, "Linear mixed models with skew-elliptical distributions: A Bayesian approach," *Computational Statistics and Data Analysis*, vol. 52, pp. 5033-5045, 2008.

2. N. Duan, "Smearing estimate: a nonparametric retransformation method", *Journal of the American Statistical Association*, vol. 78, pp. 605-610, 1983.
3. P. Ghosh, M.D. Branco and H. Chakraborty, "Bivariate random effect model using skew-normal distribution with application to HIV-RNA," *Statistics in Medicine*, vol 26, pp. 1255-1267, 2007.
4. D. Bandyopadhyay, V. H. Lachos, C.A. Abanto-Valle and P. Ghosh, "Linear mixed models for skew-normal/independent bivariate responses with an application to periodontal disease," *Statistics in Medicine*, vol. 29, no. 25, pp. 2643-2655, 2010.
5. Y. Huang, R. Chen, G. Dagne, Y. Zhu and H. Chen, "Bayesian bivariate linear mixed-effects models with skew-normal/independent distributions, with application to AIDS clinical studies," *Journal of Biopharmaceutical Statistics*, vol. 25, no. 3, pp. 373-396, 2014.
6. S. Mahmud S, W. Lou and N.W. Johnston. "A probit-log-skew-normal mixture model for repeated measures data with excess zeros, with application to a cohort study of paediatric respiratory symptoms," *BMC Medical Research Methodology*, vol. 10, pp. 55, 2010.
7. J.G. Crag, "Some statistical models for limited dependent variables with application to the demand for durable goods," *Econometrica*, vol. 39, pp. 829-844, 1971.
8. N. Duan, W.G. Manning, C.N. Morris and J.P. Newhouse, "A comparison of alternative models for the demand of medical care," *Journal of Business and Economic Statistics*, vol. 1, pp. 115-126, 1983.
9. D. Xing, Y. Huang, H. Chen, Y. Zhu, G.A. Dagne and J. Baldwin, "Bayesian inference for two-part mixed-effects model using skew distributions, with application to longitudinal semi-continuous alcohol data," *Statistical Method in Medical Research*, DOI: 10.1177/0962280215590284,2015.
10. M. Olsen and J. Schafer, "A two-part random effects model for semicontinuous longitudinal data," *Journal of the American Statistical Association*, vol. 96, pp. 730-745, 2001.
11. J. Tooze, G. Grunwald and R. Jones, "Analysis of repeated measures data with clumping at zero," *Statistical Methods in Medical Research*, vol. 11, pp. 341-55, 2002.
12. C.S Li, J.C. Lu, J.H. Park, K. Kim, P.A. Brinkley and J.P. Peterson, "Multivariate zero-inflated Poisson models and their applications," *Technometrics*, vol. 41, pp. 29-38, 1999.
13. A.H. Lee, K. Wang, J.A. Scott, K.K. Yau and G.J. McLachlan, "Multilevel zero-inflated Poisson regression modelling of correlated count data with excess zeros," *Statistical Methods in Medical Research*, vol. 15, pp. 47-61, 2006.
14. S. Gurmu and G.A. Dagne, "Bayesian Approach to Zero-Inflated Bivariate Ordered Probit Regression Model, with an Application to Tobacco Use," *Journal of Probability and Statistics*, doi:10.1155/2012/617678, 2012.
15. L. Liu, M.R. Conaway, W.A. Knausb and J.D. Bergin, "A random effects four-part model, with application to correlated medical costs," *Computational Statistics and Data Analysis*, vol. 52, pp. 4458-4473, 2008.
16. L.A. Hatfield, M.E. Boye and B.P. Carlin, "Joint Modeling of Multiple Longitudinal Patient-Reported Outcomes and Survival," *Journal of Biopharmaceutical Statistics*, vol. 21, no. 5, pp. 971-991, 2001.
17. L.A. Hatfield, M.E. Boye, M.D. Hackshaw and B.P. Carlin, "Multilevel Bayesian Models for Survival Times and Longitudinal Patient-Reported Outcomes With Many Zeros," *Journal of the American Statistical Association*, vol. 107, no. 499, pp. 875-885, 2012.
18. A. Azzalini, "A class of distributions which includes the normal ones," *Scandinavian Journal of Statistics*, vol. 12, pp. 171-178, 1985.
19. R.B. Arellano-Valle and M. Genton, "On fundamental skew distributions," *Journal of Multivariate Analysis*, vol. 96, pp. 93-116, 2005.
20. M.G. Genton, *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapman and Hall: Florida, 2006.
21. S.K. Sahu, D.K. Dey and M.D. Branco, "A new class of multivariate skew distributions with applications to Bayesian regression models," *The Canadian Journal of Statistics*, vol. 31, pp. 129-150, 2003.
22. B. Muthen and T. Asparouhov, "Growth mixture modeling with non-normal distributions," *Statistics in Medicine*, DOI: 10.1002/sim.6388, 2014.
23. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition, Text Revision (DSM-IV-TR), 2000.
24. P. Cohen, H. Chen, T.N. Crawford, J.S. Brook and K. Gordon, "Personality disorders in early adolescence and the development of later substance use disorders in the general population," *Drug Alcohol Dependence*, vol. 88, Suppl 1, pp. S71-84, 2007.
25. Y. Huang and H. Wu, "A Bayesian approach for estimating antiviral efficacy in HIV dynamic models," *Journal of Applied Statistics*, vol. 33, pp. 155-174, 2006.
26. M. Davidian and D.M. Giltinan, *Nonlinear models for repeated measurement data*, London, Chapman and Hall, 1995.
27. D.J. Lunn, A. Thomas, N. Best and D.J. Spiegelhalter, "WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and Computing*, vol. 10, pp. 325-337, 2000.

28. A. Gelman and D.B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, pp. 457-511, 1992.
29. D.J. Spiegelhalter, N.G. Best, B.P. Carlin and A. van der Linde, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society, Series B*, vol. 64, pp. 583-639, 2002.
30. S. Fieuws and G. Verbeke, "Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach," *Statistics in Medicine*, vol. 23, pp. 3093-3104, 2004.

Appendix A. Multivariate skew- t and skew-normal distribution

Different versions of multivariate skew distributions have been proposed and used in the literature [1,19–22]. A new class of distributions by introducing skewness in multivariate elliptically distributions, referred to as skew-elliptical (SE) distributions, were developed in the literature [20,21]. The class, which is obtained by using transformation and conditioning, contains many standard families including the multivariate skew- t (ST) and skew-normal (SN) distributions as special cases.

A k -dimensional random vector \mathbf{Y} follows a k -variate SE distribution if its probability density function (pdf) is given by

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}; m_\nu^{(k)}) = 2^k f(\mathbf{y}|\boldsymbol{\mu}, \mathbf{A}; m_\nu^{(k)}) P(\mathbf{V} > \mathbf{0}), \quad (\text{A.1})$$

where $\mathbf{A} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}^2$, $\boldsymbol{\mu}$ is a location parameter vector, $\boldsymbol{\Sigma}$ is a $k \times k$ positive (diagonal) covariance matrix, $\boldsymbol{\Delta} = \text{diag}(\delta_1, \delta_2, \dots, \delta_k)$ is a $k \times k$ skewness matrix with the skewness parameter vector $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_k)^T$; \mathbf{V} follows the elliptical distribution $El(\boldsymbol{\Delta} \mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \mathbf{I}_k - \boldsymbol{\Delta} \mathbf{A}^{-1} \boldsymbol{\Delta}; m_\nu^{(k)})$ and the density generator function $m_\nu^{(k)}(u) = \frac{u^{k/2-1} m_\nu(u)}{\int_0^\infty u^{k/2-1} m_\nu(u) du}$, with $m_\nu(u)$ being a function such that $\int_0^\infty u^{k/2-1} m_\nu(u) du$ exists. The function $m_\nu(u)$ provides the kernel of the original elliptical density and may depend on the parameter ν . This SE distribution is denoted by $SE(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}; m_\nu^{(k)})$. Two examples of $m_\nu(u)$, leading to important special cases used throughout the paper, are $m_\nu(u) = (1 + u/\nu)^{-(\nu+k)/2}$ and $m_\nu(u) = \exp(-u/2)$, where $\nu > 0$. These expressions lead to the multivariate ST and SN distributions, respectively. In the former case, ν corresponds to the degrees of freedom parameter.

For completeness, this Appendix briefly summarizes the multivariate ST and SN distributions introduced by Sahu *et al.* [21] and Muthen and Asparouhov [22] to be suitable for a Bayesian inference since it is built using the conditional method; see references by Sahu *et al.* [21] and Muthen and Asparouhov [22] for detailed discussions on properties of ST and SN distributions.

Assume a k -dimensional random vector \mathbf{Y} follows a k variate ST or SN distribution with location vector $\boldsymbol{\mu}$, $k \times k$ positive (diagonal) covariance matrix $\boldsymbol{\Sigma}$, $k \times k$ skewness diagonal matrix $\boldsymbol{\Delta} = \text{diag}(\delta_1, \delta_2, \dots, \delta_k)$ with the skewness parameter vector $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_k)^T$ and degrees of freedom ν (for ST).

A.1 Multivariate skew- t distribution

A k -dimensional random vector \mathbf{Y} follows a k variate ST distribution, if its pdf is given by

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \nu) = 2^k t_{k,\nu}(\mathbf{y}; \boldsymbol{\mu}, \mathbf{A}) T_{1,\nu+k}(c^{-1} \mathbf{B}; 0, 1), \quad (\text{A.2})$$

where $\mathbf{A} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}^2$, $\mathbf{B} = \boldsymbol{\delta}^T \mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \sqrt{(\nu + k)/(1 - \boldsymbol{\delta}^T \mathbf{A}^{-1} \boldsymbol{\delta})}$ and $c = \nu + (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu})$; $t_{k,\nu}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is pdf of k -dimensional t -distribution with location vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν ; $T_{1,\nu}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is cumulative distribution function (cdf) of standard (univariate) t -distribution with degrees of freedom ν . We denote the above distribution by $ST_{k,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$. It can be shown that the mean and covariance matrix of \mathbf{Y} for the ST distribution $ST_{k,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ are given by

$$\begin{aligned} E(\mathbf{Y}) &= \boldsymbol{\mu} + (\nu/\pi)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \boldsymbol{\delta}, \\ \text{cov}(\mathbf{Y}) &= [\boldsymbol{\Sigma} + \boldsymbol{\Delta}^2] \frac{\nu}{\nu-2} - \frac{\nu}{\pi} \left[\frac{\Gamma\{(\nu-1)/2\}}{\Gamma(\nu/2)} \right]^2 \boldsymbol{\Delta}^2. \end{aligned} \quad (\text{A.3})$$

It is noted that when $\boldsymbol{\Delta} = \mathbf{0}$, the ST distribution reduces to the usual t -distribution. In order to have a zero mean vector, we should assume the location parameter $\boldsymbol{\mu} = -(\nu/\pi)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \boldsymbol{\delta}$. According to

the study by cite Muthén and Asparouhov [22], if \mathbf{Y} follows $ST_{k,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$, it can be expressed by a convenient stochastic representation as follows.

$$\mathbf{Y} = \boldsymbol{\mu} + \delta|X_0| + \boldsymbol{\Sigma}^{1/2}\mathbf{X}_1, \quad (\text{A.4})$$

where \mathbf{X}_1 is k -dimensional vector with a multivariate t -distribution with zero mean, identity covariance matrix \mathbf{I}_k and degrees of freedom parameter ν ; X_0 is an univariate variable with a standard t -distribution with mean 0, variance 1 and degrees of freedom parameter ν . The term $\delta|X_0|$ can be thought of as a univariate skewness factor with factor loadings represented by the skew parameters of $\boldsymbol{\delta}$, where the skewness is identified as that part of the \mathbf{Y} distribution not captured by the symmetric part \mathbf{X}_1 . Let $w = |X_0|$; then, w follows a standard t -distribution truncated in the space $w > 0$ (i.e., a standard half t -distribution). Thus, a two-level hierarchical representation of (A.4) is given by

$$\mathbf{Y}|w \sim t_{k,\nu}(\boldsymbol{\mu} + \boldsymbol{\delta}w, \boldsymbol{\Sigma}), \quad w \sim t_{1,\nu}(0,1)I(w > 0), \quad (\text{A.5})$$

where $I(\cdot)$ denotes an indicator function. Note that when $\boldsymbol{\delta} = \mathbf{0}$, the ST distribution $ST_{k,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ presented by the hierarchical expression (A.5) reduces to the multivariate t -distribution.

A.2. Multivariate SN distribution

We briefly discuss a multivariate SN distribution introduced by Sahu *et al.* [21] here. An k -dimensional random vector \mathbf{Y} follows an k - variate SN distribution, if its probability density function (pdf) is given by

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}) = 2^k |\mathbf{A}|^{-1/2} \phi_k\{\mathbf{A}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\}P(\mathbf{V} > \mathbf{0}), \quad (\text{A.6})$$

where $\mathbf{A} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}^2$, $\mathbf{V} \sim N_m\{\boldsymbol{\Delta}\mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \mathbf{I}_k - \boldsymbol{\Delta}\mathbf{A}^{-1}\boldsymbol{\Delta}\}$, and $\phi_k(\cdot)$ is the pdf of $N_k(\mathbf{0}, \mathbf{I}_k)$. We denote the above distribution by $SN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$. An appealing feature of equation (A.6) is that it gives independent marginal when $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$. The pdf (A.6) thus simplifies to

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}) = \prod_{i=1}^k \left[\frac{2}{\sqrt{\sigma_i^2 + \delta_i^2}} \phi \left\{ \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \Phi \left\{ \frac{\delta_i}{\sigma_i} \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \right], \quad (\text{A.7})$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively.

The mean and covariance matrix are given by $E(\mathbf{Y}) = \boldsymbol{\mu} + \sqrt{2/\pi}\boldsymbol{\delta}$, $Cov(\mathbf{Y}) = \boldsymbol{\Sigma} + (1 - 2/\pi)\boldsymbol{\Delta}^2$. It is noted that when $\boldsymbol{\delta} = \mathbf{0}$, the SN distribution reduces to usual normal distribution. In order to have a zero mean vector, we should assume the location parameter $\boldsymbol{\mu} = -\sqrt{2/\pi}\boldsymbol{\delta}$.

According to the study by Arellano-Valle and Genton [19], if \mathbf{Y} follows $SN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$, it can be expressed by a convenient stochastic representation as follows.

$$\mathbf{Y} = \boldsymbol{\mu} + \delta|X_0| + \boldsymbol{\Sigma}^{1/2}\mathbf{X}_1, \quad (\text{A.8})$$

where \mathbf{X}_1 is k -dimensional vector with a multivariate normal distribution $N_k(\mathbf{0}, \mathbf{I}_k)$; X_0 is an univariate variable with a standard normal distribution $N(0, 1)$. Let $w = |X_0|$; then, w follows a standard normal distribution truncated in the space $w > 0$ (i.e., a standard half normal distribution). Thus, a two-level hierarchical representation of (A.8) is given by

$$\mathbf{Y}|w \sim N_k(\boldsymbol{\mu} + \boldsymbol{\delta}w, \boldsymbol{\Sigma}), \quad w \sim N(0, 1)I(w > 0). \quad (\text{A.9})$$

Note that when $\boldsymbol{\delta} = \mathbf{0}$, the hierarchical expression (A.9) presented for the SN distribution $SN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ reduces to a special case which is the multivariate normal distribution $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.