Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2015; 1(3-4): 112-120.

Original Article

Bayesian Analysis of Non-normal and Non-independent Mixed Model Using Skew-Normal/Independent Distributions

Mohammad Gholami-Fesharaki[1], Anoshiravan Kazemnejad[2*]

[1] Assistant Professor, Department of Biostatistics, School of Medical Sciences, Tarbiat Modarres University, Tehran, Iran
[2] Professor, Department of Biostatistics, School of Medical Sciences, Tarbiat Modarres University, Tehran, Iran

ARTICLE INFO

Received 23.02.2015
Revised 18.05.2015
Accepted 07.06.2015
Published 20.07.2015

Available online at:
http://jbe.tums.ac.ir

Key words:
multilevel modeling,
bayesian analysis,
normal/independent
distributions,
triglycerides

ABSTRACT

The main assumptions in liner mixed model are normality and independency of random effect component. Unfortunately, these two assumptions might be unrealistic in some situations. Therefore, in this paper, we will discuss about the analysis of Bayesian analysis of non-normal and non-independent mixed model using skew-normal/independent distributions, and finally, this methodology is illustrated through an application to a triglyceride data from Isfahan's Mobarakeh Steel Company Cohort Study.

Introduction

Longitudinal studies are common and reliable surveys in the medical field (1) that involve repeated observations of the same subjects over time. The most common statistical tool for analyzing longitudinal and repeated measurements data is linear mixed model (LMM) (2, 3). The two basic assumptions in LMM are normality and independency of random effect component which are chosen substantially for mathematical convenience. However, these two assumptions might be unsuitable in some situations. Inference on fixed effects without considering non-independency of the random effects causes a lower estimate of standard errors and increase Type I error consequently (4, 5). Although previous studies shown asymptotically robust estimation to non-normality of the random effects (6, 7), it is so important to select appropriate random effects distribution for efficient estimation and unbiased model-based standard errors (8).

In order to overcome non-normality, different solutions have been proposed by different people. The simplest way, especially in the presence of severe skewness of distribution is the use of data transform specially Box – Cox transformation. Although transformation is generally used, interpreting of parameter under transformation is difficult and some alternative ways are more desirable (9-11). Another solution for overcoming this problem the is use of models which are theoretically able to explain the observed changes without the use of normal

* Corresponding Author: Kazemnejad Anoshiravan, Postal Address: Department of Biostatistics, School of Medical Sciences, Tarbiat Modarres University, Tehran, Iran.
Email: kazem_an@modares.ac.ir

Please cite this article in press as: Gholami-Fesharaki M, Kazemnejad A. Bayesian Analysis of Non-normal and Non-independent Mixed Model Using Skew-Normal/Independent Distributions. *J Biostat Epidemiol. 2015; 1(3-4): 112-120*

distribution. In this context, two approaches have been proposed.

The first approach is the use of semiparametric LMM. Many researchers have studied the use of this method, including the works done with Davidian and Gallant (12), Magder and Zeger (13), Verbeke and Lesaffre (14), Kleinman and Ibrahim (15), Aitkin (16), Jiang (17), Tao et al. (18), Zhang and Davidian (19), and Ghidey et al. (20).

The second approach is the use of the family of asymmetric distributions that have the ability to explain the skewness and kurtosis in data. In this context, we can referred to work be done with Pinheiro et al. (21), Zhou and He (22), Rosa et al. (23), Lin and Lee (24, 25), Lange and Sinsheimer (26), Ma and Genton (27), Arellano-Valle et al. (28), Lachos et al. (29), Jara et al. (11) and Bandyopadhyay et al. (30). The literature review in this area shows that most models that use skew distribution in longitudinal data are two levels and fewer study fit skew distribution in three and more levels like nested longitudinal data.

Hence, in this paper, we discuss about new approach to analysis non-normal and non-independent LMM. The remainder of this paper is organized as follows.

After short introduction about skew normal (SN) and skew-normal/independent (SNI) distribution in Section 2, the statistical models and likelihood function are presented in Section 3 and then priors and joint posterior distributions and practical implementation are discussed, and then in Section 4, the advantage of the proposed methodology is illustrated with triglycerides (TG) data, and finally, some concluding remarks are presented in Section 5.

SN and SNI Distribution

SN distribution

The SN density that introduced with Azzalini (31) is distribution that its density function is given by:

$$f\lambda(x) = 2\emptyset(x)\Phi(\lambda x) \qquad \lambda \in R \& x \in R \qquad (1)$$

Where φ(x) and Φ(x) are probability density function (PDF) and cumulative density function (CDF) of the normal distribution, respectively. In this density, if = 0, SN reduces to the standard normal density and if $\lambda \rightarrow \pm\infty$ SN tends to the half-normal distribution. The important point of SN is, it accommodates skewness but it also includes as a special case of normal density, and it has the best normal distribution properties. The range of possible skewness values of SN is (-0.995, 0.995) (32). The important properties of SN are:

Property 1:

if Z_1, Z_2 ~ N(0,1), then $X = \delta|Z_1|+\sqrt{1-\delta^2 Z_2} \sim SN(\lambda)$

$$\lambda = \frac{\delta}{\sqrt{1-\delta^2}} \quad \delta \in [-1, +1]$$

Property 2:

If X~SN(λ), then Y = μ + σX~SN(μ, σ, λ) with following density function.

$$f(y, \mu, \sigma, \lambda) = \frac{2}{\sigma}\emptyset\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right) \ \lambda \& \mu \in R \& \sigma < 0 \& x \in R$$

SNI distribution

SNI distributions define like equation 2

$$X = \mu + \frac{Z}{\sqrt{U}} \qquad (2)$$

In equation 2, μ is location parameter, U is positive random distribution with CDF H(u|v) and PDF h(u|v), v is a scalar or vector indexing the distribution of U. Z is SN distribution with location, dispersion, and skewness parameters, μ σ^2, λ, respectively. Given U = μ the distribution of X is $X|U = u \sim SN\left(\mu, \frac{\sigma^2}{u}\right), \lambda$ with density function like equation 3

$$f(x) = 2\int \frac{\sqrt{u}}{\sigma} \times \varphi\left(\frac{\sqrt{u}(x-\mu)}{\sigma}\right)\Phi\left(\sqrt{u}\lambda\frac{x-\mu}{\sigma}\right)dH(u|\,v) \qquad (3)$$

If U taking distribution $\frac{X^2(v)}{v}$, $X^2(1)$, $X^2(v)v \rightarrow \infty$, and beta(v,1) then the distribution of X reduce to Skew T (ST) with v degree of freedom, Skew-Cauchy, SN and Skew Slash (SS) respectively (33). And also if λ = 0 then the

SNI reduce to the normal-independent distribution (26).

Statistical Models and Likelihood Function

Simple LMM data with non-independent random effect can be written like a simple 3 level modeling (34, 35). Consider

Level 1: $y_{i(jk)} = \beta_{0(jk)} + \beta_1 x_{i(jk)} + \varepsilon_{i(jk)}$
Level 2: $\beta_{0(jk)} = \beta_{0(k)} + u_{(jk)}$
Level 3: $\beta_{0(k)} = \beta_0 + v_k$ (4)

With assumptions

$$v_k \sim N(0,\sigma_V^2),\ u_{j(k)} \sim N(0,\sigma_u^2),\ \varepsilon_{i(jk)} \sim N(0,\sigma_\varepsilon^2) \quad (5)$$

$$\forall i: 1..M_{ik} \text{ and } j: 1..M_k \text{ and } k: 1..M \quad (6)$$

$$Cov(v_k, u_{j(k)} = Cov(v_k, \varepsilon_{i(jk)} = Cov(u_{j(k)}, \varepsilon_{i(jk)}) = 0$$

Where M is the number of cluster in the total dataset, M_k is the number of subjects in cluster k, and M_{ik} is the repetition subject i in cluster k.

In equation 4, random effect v causes non-independency of random effect u, because u nested in v. According to equations 4 and 6 $Cov(u_{(jk)}, u_{(jk)}) = \sigma_v^2$. After substituting the level 3 in level 2 and then level 2 in level 1 and rearranging the terms, we got the model like equation 7.

$$y_{i(jk)} = \beta_0 + \beta_1 x_{i(jk)} + v_k + u_{j(k)} + \varepsilon_{i(jk)} \quad (7)$$

i: $1..M_{ik}$, j: $1..M_k$, k: 1..M

In this paper, we want to use SNI distribution instead of normal distribution for $u_{j(k)}$

$$u_{j(k)} \sim SNI(0, \lambda_u, \sigma_u^2, w_u) \quad (8)$$

With use of property 1 and 2, we can write $y_{i(jk)}$ with use of equation 9 like equation 10

$$u_{j(k)} = \lambda_u \frac{|t^{u^1}_{j(k)}|}{\sqrt{w^u_{j(k)}}} + \frac{t^{u^2}_{j(k)}}{w^u_{j(k)}}$$ that

$$t^{u^1}_{j(k)} \& t^{u^2}_{j(k)} \sim N(0,\sigma_u^2) \& w^u_{j(k)} \sim f(w^u|w^u_0) \quad (9)$$

$$y_{i(jk)} = \beta_0 + \beta_1 x_{i(jk)} + \frac{t^{u^2}_{j(k)}}{w^u_{j(k)}} + \lambda_u \frac{|t^{u^1}_{j(k)}|}{w^u_{j(k)}} + v_k + \varepsilon_{i(jk)} \quad (10)$$

Consider

$$X = \begin{pmatrix} 1 & x_{111} \\ 1 & x_{111} \\ . & . \\ . & . \\ 1 & x_{MM_MMM_M} \end{pmatrix},\ Y = \begin{pmatrix} y_{111} \\ y_{112} \\ \cdot \\ \cdot \\ y_{MM_MMM_M} \end{pmatrix},\ \beta = (\beta_0, \beta_1),\ U = \begin{pmatrix} u_{11} \\ u_{12} \\ \cdot \\ \cdot \\ u_{MM_M} \end{pmatrix},\ W^u = \begin{pmatrix} w^u_{11} \\ w^u_{12} \\ \cdot \\ \cdot \\ w^u_{MM_M} \end{pmatrix}\ T^{u_1} = \begin{pmatrix} t^{u_1}_{11} \\ t^{u_1}_{12} \\ \cdot \\ \cdot \\ t^{u_1}_{MM_M} \end{pmatrix}\ V = \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_M \end{pmatrix} \quad (11)$$

Because of nested structure of data and independency of random vector in multilevel modeling, we can write PDF of f(Y, U, W^u, T^{u1}, V) like equation 12

$$f(Y, U, W^u, T^{u_1}, V) = f(Y|U, W^u, T^{u_1}, V) \times f(U|W^u, T^{u_1}, V) \times f(T^{u_1}|W^u, V) \times (W^u|V) \times f(V) \quad (12)$$

Consider $n_0 = \sum_{(k=1)}^{M} \sum M_k$ and $n_{00} = \sum_{k=1}^{M} \sum_{j=1}^{M_k} M_{kj}$ according to assumption we have

$$Y|U, W^u, T^{u_1}, V \sim N_{n..}(X\beta + V + U, \sigma_\varepsilon^2 I) \quad (13)$$

$$f(Y|U, W^u, T^{u_1}, V) = \prod_{k=1}^{M} \prod_{j=1}^{M_k} \prod_{i=1}^{M_{kj}} \sqrt{\frac{1}{2\pi\sigma_\varepsilon^2}} \exp(-\frac{1}{2\sigma_\varepsilon^2}(y_{i(jk)} - \beta_0 - \beta_1 x_{i(jk)} - v_k - u_{j(k)})^2) \quad (14)$$

$$U|\,W^u, T^{u_1}, V \sim N_n(\lambda_u (T^{u_1})^t (W^u)^{-0.5}, \sigma_u^2 W^{u-1} I) \quad (15)$$

$$f(U|W^u, T^{u_1}, V) = \prod_{k=1}^{M} \prod_{j=1}^{M_k} \sqrt{\frac{w^u_{j(k)}}{2\pi\sigma_u^2}} \exp(-\frac{w^u_{j(k)}\left(u_{j(k)} - \frac{\lambda_u |t^{u^1}_{j(k)}|}{\sqrt{w^u_{j(k)}}}\right)^2}{2\sigma_u^2}) \quad (16)$$

$$T^{u_1}|W^u, V \sim N_n(0, \sigma_u^2 I) \quad (17)$$

$$f(T^{u_1}|, W^u, V) = \prod_{k=1}^{M} \prod_{j=1}^{M_k} \sqrt{\frac{1}{2\pi\sigma_u^2}} \times \exp(-\frac{t^{u^1\,2}_{j(k)}}{2\sigma_u^2}) \quad (18)$$

$$f(W^u|V) = \prod_{k=1}^{M} \prod_{j=1}^{M_k} f_{w^u}(w^u_{j(k)}) \quad (19)$$

$$V \sim N_M(0, \sigma_V^2 I) \quad (20)$$

$$f(V) = \prod_{k=1}^{M} \sqrt{\frac{1}{2\pi\sigma_v^2}} \times \exp(-\frac{v_k^2}{2\sigma_v^2}) \quad (21)$$

Substitute equations 14, 16, 18, 19, and 21 in equation 12 the joint CDF $f(Y, U, W^u, T^{u_1}, V)$ is like equation 22. Multiplying this CDF in prior distribution, we can achieve posterior distribution. With the use of hierarchical Bayesian approach, we can estimate parameters like, beta, sigma, and lambda.

$$f(Y|U, W^u, T^{u_1}, V) = \prod_{k=1}^{M} \prod_{j=1}^{M_k} \prod_{i=1}^{M_{kj}} \sqrt{\frac{1}{2\pi\sigma_\varepsilon^2}} \exp(-\frac{1}{2\sigma_\varepsilon^2}(y_{i(jk)} - \beta_0 - \beta_1 x_{i(jk)} - v_k - u_{j(k)})^2) \times \prod_{k=1}^{M} \prod_{j=1}^{M_k} \sqrt{\frac{w_{j(k)}^u}{2\pi\sigma_u^2}} \exp\left(-\frac{w_{j(k)}^u \left(u_{j(k)} - \frac{\lambda_u |t_{j(k)}^{u^1}|}{\sqrt{w_{j(k)}^u}}\right)^2}{2\sigma_u^2}\right) \times \sqrt{\frac{1}{2\pi\sigma_u^2}} \exp(-\frac{t_{j(k)}^{u^1\,2}}{2\sigma_u^2}) \times f_{W^u}(w_{j(k)}^u)) \times \prod_{k=1}^{M} \sqrt{\frac{1}{2\pi\sigma_v^2}} \times \exp(-\frac{v_k^2}{2\sigma_v^2}) \quad (22)$$

Priors and Joint Posterior Distributions and Practical Implementation

Priors and joint posterior distributions

In this paper normal, inverse gamma, and exponential distributions were considered as prior distributions, respectively, for beta coefficient and skewness parameter, scale parameter, and degree of freedom in ST and SS distribution. These distributions were popular choice in Bayesian LMM (36, 37). With considering $\theta = (\beta_0, \beta_1, \sigma_u^2, \sigma_v^2, \sigma_e^2, \lambda_u, v_u$ the joint priors distribution is like equation 23.

$$\beta_0 \sim N(\mu_{\beta 0}, \sigma_{\beta 0}^2)$$
$$\beta_1 \sim N(\mu_{\beta 1}, \sigma_{\beta 1}^2)$$
$$\lambda_u \sim N(\mu_{\lambda u}, \sigma_{\lambda u}^2)$$
$$\sigma_v^2 \sim IG(\alpha_v, \gamma_v)$$
$$\sigma_u^2 \sim IG(\alpha_u, \gamma_v)$$
$$\sigma_\varepsilon^2 \sim IG(\alpha_{\varepsilon u}, \gamma_\varepsilon)$$
$$v_u \sim exp(v^u)$$

$$\pi(\theta) = \pi(\beta_0) \times \pi(\beta_1) \times \pi(\sigma_u^2) \times \pi(\sigma_v^2) \times \pi(\sigma_e^2) \times \pi(\lambda_u) \times \pi(\lambda_\varepsilon) \times \pi(v_u) \quad (23)$$

Combining the likelihood function (equation 22) and the prior distributions (equation 23), the joint posterior distribution for θ is now

$$\pi(\theta u, v, t^{u_1}, w^u|y) = \pi(\theta) \times \prod_{k=1}^{M} \prod_{j=1}^{M_k} \prod_{i=1}^{M_{kj}} \sqrt{\frac{1}{2\pi\sigma_\varepsilon^2}} \exp(-\frac{1}{2\sigma_\varepsilon^2}(y_{i(jk)} - \beta_0 - \beta_1 x_{i(jk)} - v_k - u_{j(k)})^2) \times \prod_{k=1}^{M} \prod_{j=1}^{M_k} \sqrt{\frac{w_{j(k)}^u}{2\pi\sigma_u^2}} \exp\left(-\frac{w_{j(k)}^u \left(u_{j(k)} - \lambda_u |t_{j(k)}^{u^1}| / \sqrt{w_{j(k)}^u}\right)^2}{2\sigma_u^2}\right) \times \sqrt{\frac{1}{2\pi\sigma_u^2}} \exp(-\frac{t_{j(k)}^{u^1\,2}}{2\sigma_u^2}) \times f_{W^u}(w_{j(k)}^u)) \times \prod_{k=1}^{M} \sqrt{\frac{1}{2\pi\sigma_v^2}} \times \exp(-\frac{v_k^2}{2\sigma_v^2}) \quad (24)$$

Distribution (24) is analytically intractable, but MCMC methods such as the Gibbs sampler and Metropolis–Hastings algorithm can be used to draw samples, from which features of the marginal posterior distribution of interest can be inferred. An outline of the conditional posteriors of all model parameters is given in Appendix A.

Practical implementation

In our situation, vague prior distributions (equation 25) are utilized; then we used WinBUGS software for Bayesian analysis. Results are based on every 100 draw from an MCMC chain of length 11,000 with a burn-in of 1000. This proved more than enough for convergence, and much shorter runs led to virtually identical results. For investigating sensitivity analysis, we change the prior of parameters and monitor the posterior distributions. In this study, we used graphical tools like density plot, trace plot, and Gelman-Rubin convergence diagnostic test for model checking.

$$\beta_k \text{ and } \lambda_{\acute{k}} \sim N(0,100) \ \forall \ k \ \& \ \acute{K}$$
$$\sigma_v^2, \sigma_u^2 \text{ and } \sigma_\varepsilon^2 \sim IG(0.01,0.01)$$
$$v_u \sim \exp(0.1) \quad (25)$$

Model selection and goodness-of-fit

For model selection, we use deviance information criterion (DIC) that it is defined in as equation 26 (38).

$$\mathrm{DIC} = \mathrm{D}(\bar{\theta}) + 2 + \mathrm{P_D}$$
$$\mathrm{D}(\theta) = -2 \times \log(\mathrm{f}(\mathrm{y}|\theta))$$
$$\mathrm{P_D} = \overline{\mathrm{D}(\theta)} - \mathrm{D}(\bar{\theta})$$

Where $D(\theta)$ is the usual deviance measure, $D(\bar{\theta})$ is its posterior mean and P_D, can be interpreted as the number of "effective" parameters for model considered. Smaller DIC values indicate a better-fitting model (39).

Medical Example

Shift work (SW) and TG

SW is an essential part of today's business reality. SW is often defined as work outside the hours of around 7 a.m. and 6 p.m. (40, 41). Few studies have investigated the relationship between TG and SW. TG is a factor affecting overall cardiovascular health (42, 43). Thus, the current study aimed to test the association between the SW and TG with the use of non-normal and non-independent LMM.

The data used in this study were from a longitudinal historical study that conducted on all employed workers of Isfahan's Mobarakeh Steel Company in Iran between 1997 and 2011. A total of 574 workers participated in this study and 4600 records of data were derived from their medical records using the stratified random sampling method. The variable of SW was categorized as Routine Rotating Shifts (RRS) (2 morning shifts, 2 evening shifts, 2 night shifts, and 2 days off) and Weekly Rotating Shifts (WRS) (3 morning shifts, 3 evening shifts, and one day off every two weeks, Fridays always off). Regular Day Workers (RDY) worked from morning to evening on weekdays and had Thursdays and Fridays off. In this study, TG was considered as the dependent variable, and SW, age, and body mass index (BMI) were considered as independent variables.

Data analysis and finding

The statistical model that fit in this paper was like equation 27.

$$\begin{aligned}\text{Level 1: } \mathrm{TG}_{\mathrm{i(jk)}} &= \beta_{0(\mathrm{jk})} + \beta_1 \mathrm{Age}_{\mathrm{i(jk)}} \\ &\quad + \beta_2 \mathrm{BMI}_{\mathrm{i(jk)}} + \beta_3 \mathrm{Shiftr}_{\mathrm{i(jk)}} \\ &\quad + \beta_4 \mathrm{Shiftw}_{\mathrm{i(jk)}} + \varepsilon_{\mathrm{i(jk)}}\end{aligned}$$

$$\text{Level2: } \beta_{0(\mathrm{jk})} = \beta_{0(\mathrm{k})} + \mathrm{u}_{(\mathrm{jk})}$$

$$\begin{aligned}&\text{Level3: } \beta_{0(\mathrm{k})} = \beta_0 + \mathrm{v}_{(\mathrm{k})} \\ &\varepsilon_{\mathrm{i(jk)}} \sim N(0, \sigma_\varepsilon^2) \\ &\mathrm{u}_{\mathrm{j(k)}} \sim \mathrm{SNI}(0, \lambda_\mathrm{u}, \sigma_\mathrm{u}^2, \mathrm{w_u}) \\ &\mathrm{v_k} N(0, \sigma_\mathrm{v}^2)\end{aligned} \tag{27}$$

In this equation, Shift r and Shift w stand for the effect of work in RRS and WRS rather RDY, respectively, and BMI stand for BMI variable.

We apply 5 models in TG data. Model 1 is a simple LMM (normal-independent), Model 2: Simple LMM with non-independent and normal random effect (normal-non independent), Model 3: Simple LMM with non-independent and SN random effect (SN-non independent), Model 4: Simple LMM with non-independent and ST random effect (ST-non independent) and finally Model 5: Simple LMM data with non-independent and SS random effect (SS-non independent). In table 1 and figure 1, summary statistics of TG and density plot are shown, respectively. Also table 2 represents the comparison among the 5 competing models using Bayesian model choice criterion. Note that all independent and skew models produced lower DIC and Dbar rather than the normal model. In particular, ST-non independent model produces the best fit among the competing skew models.

Table 3 provides posterior estimates of beta coefficients, asymmetry parameters, the variance components of random errors, and degree of freedom of ST and SS distributions. In particular, we provide estimates of posterior mean, standard deviation (SD), and 95% credible intervals (CI).

Table 1. Summary statistics of TG

Mean	Standard division	Median	First quantile	Third quantile	Skewness	Kurtosis
162.15	150.22	138	95.0	162.1	2.38	9.35

TG: Triglycerides

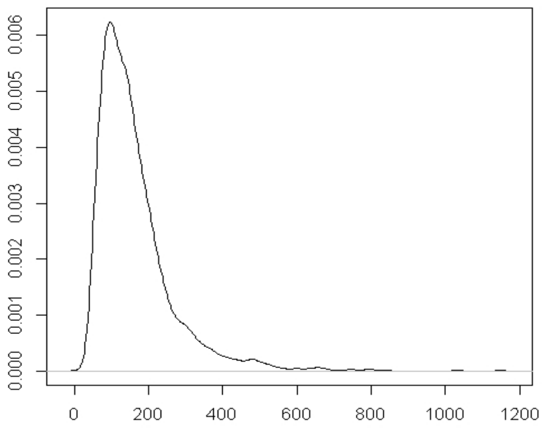

Figure 1. Density plot of triglycerides

As shown in table 3, skewness parameter is significant and positive for all three fitted models providing evidence of right-skewness for our data. And DF parameter in ST and SS model is significant and confirms its sufficient disparity from the normal framework. The one point on table 3 is that estimate of the within-subject variances σ_u^2, is smaller in the skewed class of models as compared with the normal model, it is because of inter relation between high variability, heavy tails as well as skewness (30). In table 3, we provide the posterior mean, SD, and 95% CIs for the posterior estimates of parameters. RRS and WRS are not significant in all 5 models and BMI is significant in all models, and finally Age is not significant in the skew model and significant in the normal model.

Table 2. Model comparison using DIC and Dbar

Number	Model	DIC	pD	Dbar
1	Normal-independent	53290	491.6	53781.6
2	Normal-non independent	53250	503.0	53753
3	Skew normal-non independent	53210	443.2	53653.2
4	Skew T-non independent	53180	402.1	53582.1
5	Skew Slash-non independent	53190	394.9	53584.9

DIC: Deviance information criterion

Table 3. Posterior estimates of fixed effect parameters

		Normal-independent	Normal-non independent	Skew normal-non independent	Skew T-non independent	Skew Slash-non independent
β_0	Mean	-20.44	-18.92	-41.94	-35.39	-35.78
	SD	8.41	8.54	8.78	8.17	8.17
	95% CI	(-36.37, -3.06)	(-35.88, -2.66)	(-58.66, -24.67)	(-51.27, -20.04)	(-50.03, -18.33)
Age	Mean	0.54	0.57	0.169	0.31	0.31
	SD	0.243	0.23	0.22	0.20	0.20
	95% CI	(0.09, 1.03)	(0.13, 1.04)	(-0.25, 0.62)	(-0.06, 0.72)	(-9.11, 0.70)
BMI	Mean	6.56	6.57	5.24	5.29	5.26
	SD	0.41	0.40	0.42	0.03	0.39
	95% CI	(5.66, 7.35)	(5.77, 7.29)	(4.36, 6.03)	(4.49, 5.29)	(4.51, 6.04)
Shift r	Mean	-2.94	-1.57	-2.63	-0.72	-1.49
	SD	3.65	3.92	3.63	3.47	3.64
	95% CI	(-10.10, 4.18)	(-8.98, 6.25)	(-9.63, 4.31)	(-7.55, 6.05)	(-8.49, 5.65)
Shift w	Mean	-4.72	-4.06	-5.30	-3.72	-4.61
	SD	7.91	6.60	6.36	6.23	6.03
	95% CI	(-16.78, 7.91)	(-16.24, 9.01)	(-18.01, 6.99)	(-15.73, 8.56)	(-1.60, 8.01)
σ_e^2	Mean	5670	5613.14	5627.34	5661	5658.43
	SD	126.6	128.81	126.02	126.1	121.87
	95% CI	(5429.30, 5924.60)	(5361.50, 5861.51)	(5377.86, 5881.79)	(5417.07, 5911.15)	(5423.50, 5910.06)
σ_u^2	Mean	4433.33	4320.44	132.94	29.91	181.86
	SD	311.10	312.16	58.02	32.67	76.06
	95% CI	(3862, 5074)	(3738.9, 4945.3)	(60.40, 262.45)	(7.55, 126.8)	(8.87, 374.65)
λ_u	Mean	-	-	9.04	7.53	2.96
	SD	-	-	2.05	2.73	0.64
	95% CI	-	-	(5.87, 12.64)	(2.42, 11.35)	(1.59, 4.11)
d_u	Mean	-	-	-	2.68	0.99
	SD	-	-	-	0.41	0.12
	95% CI	-	-	-	(1.98, 3.57)	(0.76,1.26)
σ_v^2	Mean	598.74	587.64	311.04	308.30	204.46
	SD	293.44	288.99	211.87	170.10	174.31
	95% CI	(33.76, 1302.45)	(40.42, 1239.43)	(2.89, 805.46)	(67.32, 722.9)	(0.02, 628.13)

SD, 2.5 and 97.5% represents, respectively, the standard deviation and percentiles from the posterior distributions of parameters. SD: Standard deviation, BMI: Body mass index, CI: Credible intervals

Conclusion

In this paper, we introduce a new version of LMM with non-normal and non-independent random effect. Using this method, the ST model provided the best fit to these data among other competing models. It means the data show some degree of skewness and kurtosis that it can violate traditional normality assumptions of the random effect. ST was shown best fits in another study like work done by Lachos et al. (29) and Bandyopadhyay et al. (30). Our methodology can be further extended to modeling LMM with non-normal error term and non-normal contextual random effect (v in equation 4) and also categorical and survival data analysis, which will be pursued in future research.

Acknowledgments

The support of this work by Tarbiat Modares University, Tehran, Iran is gratefully acknowledged.

References

1. Ahrens W, Pigeot I. Handbook of epidemiology. New York, NY: Springer Science & Business Media; 2005.
2. Diggle P, Heagerty P, Liang K, Zeger S. Analysis of longitudinal data. 2nd ed. Oxford, UK: OUP Oxford; 2002.
3. Fitzmaurice G, Laird N, Ware JH. Applied longitudinal analysis. Hoboken, NJ: John Wiley & Sons; 2004.
4. Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Anal 2006; 1(3): 473-514.
5. Luke DA. Multilevel modeling. 3rd ed. Washington DC: SAGE Publications; 2004.
6. Butler SM, Louis TA. Random effects models with non-parametric priors. Stat Med 1992; 11(14-15): 1981-2000.
7. Verbeke G, Lesaffre E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Computational Statistics & Data Analysis 1997; 23(4): 541-56.
8. Ghidey W, Lesaffre E, Verbeke G. A comparison of methods for estimating the random effects distribution of a linear mixed model. Stat Methods Med Res 2010; 19(6): 575-600.
9. Lin TI, Lee JC. On modelling data from degradation sample paths over time. Australian & New Zealand Journal of Statistics 2003; 45(3): 257-70.
10. Lee JC, Lin TI, Lee KJ, Hsu YL. Bayesian analysis of Box–Cox transformed linear mixed models with ARMA (p,qp,q) dependence. Journal of Statistical Planning and Inference 2005; 133(2): 435-51.
11. Jara A, Quintana F, San Martín E. Linear mixed models with skew-elliptical distributions: A Bayesian approach. Computational Statistics & Data Analysis 2008; 52(11): 5033-45.
12. Davidian M, Gallant AR. The nonlinear mixed effects model with a smooth random effects density. Biometrika 1993; 80(3): 475-88.
13. Magder LS, Zeger SL. A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. Journal of the American Statistical Association 1996; 91(435): 1141-51.
14. Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association 1996; 91(433): 217-21.
15. Kleinman KP, Ibrahim JG. A semi-parametric Bayesian approach to generalized linear mixed models. Stat Med 1998; 17(22): 2579-96.
16. Aitkin M. A general maximum likelihood analysis of variance components in generalized linear models. Biometrics 1999; 55(1): 117-28.
17. Jiang J. Conditional inference about generalized linear mixed models. Ann Statist 1999; 27(6): 1974-2007.
18. Tao H, Palta M, Yandell BS, Newton MA. An estimation method for the semiparametric mixed effects model. Biometrics 1999; 55(1): 102-10.
19. Zhang D, Davidian M. Linear mixed models with flexible distributions of random effects for longitudinal data. Biometrics 2001; 57(3):

795-802.

20. Ghidey W, Lesaffre E, Eilers P. Smooth random effects distribution in a linear mixed model. Biometrics 2004; 60(4): 945-53.
21. Pinheiro JC, Liu C, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. Journal of Computational and Graphical Statistics 2001; 10(2): 249-76.
22. Zhou T, He X. Three-step estimation in linear mixed models with skew-t distributions. Journal of Statistical Planning and Inference 2008; 138(6): 1542-55.
23. Rosa GJM, Padovani CR, Gianola D. Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. Biometrical Journal 2003; 45(5): 573-90.
24. Lin TI, Lee JC. A robust approach to t linear mixed models applied to multiple sclerosis data. Stat Med 2006; 25(8): 1397-412.
25. Lin TI, Lee JC. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. Journal of Statistical Planning and Inference 2007; 137(2): 484-95.
26. Lange K, Sinsheimer JS. Normal/independent distributions and their applications in robust regression. Journal of Computational and Graphical Statistics 1993; 2(2): 175-98.
27. Ma Y, Genton MG. A flexible class of skew-symmetric distributions. Scand J Statist 2004; 31: 459-68.
28. Arellano-Valle RB, Bolfarine H, Lachos VH. Bayesian inference for skew-normal linear mixed models. Journal of Applied Statistics 2007; 34(6): 663-82.
29. Lachos VH, Dey DK, Cancho VG. Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective. Journal of Statistical Planning and Inference 2009; 139(12): 4098-110.
30. Bandyopadhyay D, Lachos VH, Abanto-Valle CA, Ghosh P. Linear mixed models for skew-normal/independent bivariate responses with an application to periodontal disease. Stat Med 2010; 29(25): 2643-55.
31. Azzalini A. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 1985; 12(2): 171-8.
32. Nadarajah S, Kotz S. Skewed distributions generated by the normal kernel. Statistics & Probability Letters 2003; 65(3): 269-77.
33. Bandyopadhyay D, Lachos VH, Abanto-Valle CA, Ghosh P. Labra, Linear mixed models for skew-normal/independent bivariate responses with an application to periodontal disease. Stat Med 2010; 29(25): 2643-55.
34. de Deleeuw J, Goldstein H, Meijer E. Handbook of multilevel analysis. New York, NY: Springer; 2007.
35. Hox J. Multilevel Analysis: Techniques and applications. 2nd ed. London, UK: Routledge; 2010.
36. Hobert JP, Casella G. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. Journal of the American Statistical Association 1996; 91(436): 1461-73.
37. Zhao Y, Staudenmayer J, Coull BA, Wand MP. General design Bayesian generalized linear mixed models. Statistical Science 2006; 21(1): 35-51.
38. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. J R Statist Soc B 2002; 64(4): 583-639.
39. Ntzoufras I. Bayesian modeling using win BUGS. Hoboken, NJ: Wiley; 2009.
40. Pati AK, Chandrawanshi A, Reinberg A. Shift work: Consequences and management. Current Science 2001; 81(1): 32-52.
41. Sharifian A, Farahani S, Pasalar P, Gharavi M, Aminian O. Shift work as an oxidative stressor. J Circadian Rhythms 2005; 3: 15.
42. Poss J, Custodis F, Werner C, Weingartner O, Bohm M, Laufs U. Cardiovascular disease and dyslipidemia: beyond LDL. Curr Pharm Des 2011; 17(9): 861-70.
43. Cziraky MJ, Watson KE, Talbert RL. Targeting low HDL-cholesterol to decrease residual cardiovascular risk in the managed care setting. J Manag Care Pharm 2008; 14(8 Suppl): S3-28.

Appendix A: Outline of conditional posterior distributions

Under the full model as described in (24), the full conditional distribution of parameter is given as

$$(\beta_0|\beta_1,\lambda_u,\sigma_v^2,\sigma_u^2,\sigma_\varepsilon^2,Y,U,W^u,T^{u_1},V)\sim N\left(\frac{\frac{\mu_{\beta_0}}{\sigma_{\beta_0}^2}+\frac{1}{\sigma_\varepsilon^2}\sum_{k=1}^{M}\sum_{j=1}^{M_k}\sum_{i=1}^{M_{kj}}(y_{i(jk)}-\beta_1 x_{i(jk)}-v_k-u_{j(k)})}{\frac{1}{\sigma_{\beta_0}^2}+\frac{n_{00}}{\sigma_\varepsilon^2}},\frac{1}{\frac{1}{\sigma_{\beta_0}^2}+\frac{n_{00}}{\sigma_\varepsilon^2}}\right) \quad \text{(A1)}$$

$$(\beta_1|\beta_0,\lambda_u,\sigma_v^2,\sigma_u^2,\sigma_\varepsilon^2,Y,U,W^u,T^{u_1},V)\sim N\left(\frac{\frac{\mu_{\beta_1}}{\sigma_{\beta_1}^2}+\frac{1}{\sigma_\varepsilon^2}\sum_{k=1}^{M}\sum_{j=1}^{M_k}\sum_{i=1}^{M_{kj}}(y_{i(jk)}-\beta_0-v_k-u_{j(k)})x_{i(jk)}}{\frac{1}{\sigma_{\beta_1}^2}+\frac{1}{\sigma_\varepsilon^2}\sum_{j=1}^{M_k}\sum_{i=1}^{M_{kj}}x_{i(jk)}{}^2 x_{i(jk)}{}^2},\frac{1}{\frac{1}{\sigma_{\beta_1}^2}+\frac{1}{\sigma_\varepsilon^2}\sum_{j=1}^{M_k}\sum_{i=1}^{M_{kj}}x_{i(jk)}}\right) \quad \text{(A3)}$$

$$(\lambda_u|\beta_0,\beta_1,\sigma_v^2,\sigma_u^2,\sigma_\varepsilon^2,Y,U,W^u,T^{u_1},V)\sim N\left(\frac{\frac{\mu_{\lambda_u}}{\sigma_{\lambda_u}^2}+\frac{1}{\sigma_u^2}\sum_{k=1}^{M}\sum_{j=1}^{M_k}\sqrt{w_{j(k)}^u}|t_{j(k)}^{u_1}|}{\frac{1}{\sigma_{\lambda_u}^2}+\frac{1}{\sigma_u^2}\sum_{k=1}^{M}\sum_{j=1}^{M_k}t_{j(k)}^{u_1\,2}},\frac{1}{\frac{1}{\sigma_{\lambda_u}^2}+\frac{1}{\sigma_u^2}\sum_{j=1}^{M_k}\sum_{i=1}^{M_{kj}}t_{j(k)}^{u_1\,2}}\right) \quad \text{(A4)}$$

$$(\sigma_\varepsilon^2|\beta_0,\beta_1,\lambda_u,\sigma_v^2,\sigma_u^2,Y,U,W^u,T^{u_1},V)\sim \text{IGamma}\left(I(\lambda_\varepsilon+1)\,n_{00}+\alpha_\varepsilon+1, 0.5\sum_{K=1}^{M}\sum_{j=1}^{M_k}\sum_{i=1}^{M_{kj}}\left(y_{i(jk)}-\beta_0-\beta_1 x_{i(jk)}-v_k-u_{j(k)}\right)^2+\gamma_\varepsilon\right) \quad \text{(A5)}$$

$$(\sigma_\varepsilon^2|\beta_0,\beta_1,\lambda_u,\sigma_v^2,\sigma_\varepsilon^2,Y,U,W^u,T^{u_1},V)\sim \text{IG}\left((I(\lambda_u+1)\,n_0+\alpha_u+1, 0.5\times\sum_{K=1}^{M}\sum_{j=1}^{M_k}(t_{j(k)}^{u^1\,2}+w_{j(k)}^u\left(u_{j(k)}-\frac{\lambda_u|t_{j(k)}^{u^1}|}{\sqrt{w_{j(k)}^u}}\right)^2)+\gamma_\varepsilon\right) \quad \text{(A6)}$$

$$(\sigma_v^2|\beta_0,\beta_1,\lambda_u,\sigma_u^2,\sigma_\varepsilon^2,Y,U,W^u,T^{u_1},V)\sim \text{IG}\left((I(\lambda_v+1)\,M+\alpha_v+1, 0.5\times\sum_{k=1}^{M}(w_k^v v_k^2)+\gamma_\varepsilon\right) \quad \text{(A7)}$$

$$v_u \propto \pi_u(v)\times \text{Gamma}\left(\sqrt{\prod_{k=1}^{M}\prod_{j=1}^{M_k}w_{j(k)}^u}, 0.5\sum_{k=1}^{M}\sum_{j=1}^{M_k}w_{j(k)}^u+v^u\right)$$

That $\pi_u(v)=\frac{v^{\frac{n_0 V}{2}}}{(2^{\frac{v}{2}}\Gamma(\frac{v}{2}))^{n_0}}$

Notation N: normal distribution, IGamma: Inverse Gamma Distribution, Gamma: Gamma Distribution