

# Linear mixed model with Laplace distribution (LLMM)

Fulya Gokalp Yavuz<sup>1</sup> · Olcay Arslan<sup>2</sup>

Received: 8 March 2015 / Revised: 3 March 2016 / Published online: 29 March 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Linear mixed modeling (LMM) is a comprehensive technique used for clustered, panel and longitudinal data. The main assumption of classical LMM is having normally distributed random effects and error terms. However, there are several situations for that we need to use heavier tails distributions than the (multivariate) normal to handle outliers and/or heavy tailness in data. In this study, we focus on LMM using the multivariate Laplace distribution which is known as the heavy tailed alternative to the normal distribution. The parameter estimators of interest are generated with EM algorithm for the proposed model. A simulation study is provided to illustrate the performance of the Laplace distribution over the normal distribution for LMM. Also, a real data example is used to explore the behavior of the proposed estimators over the counterparts.

**Keywords** Laplace · Mixed models · Robust distributions · EM · Orthodont

## 1 Introduction

Linear mixed models (LMMs) are regression-type models with an extra random term which represents the individual/subject effects on repeated measures. LMMs include both fixed and random effects; fixed-effect parameters represent the relations of the covariates to the dependent variable for the population; while random-effect parameters represent the clusters or individuals within a population (West et al. 2007), named individual (subject) specific parameters.

---

✉ Fulya Gokalp Yavuz  
fulyagokalp@gmail.com

<sup>1</sup> Department of Statistics, Yildiz Technical University, Davutpaşa, 34220 Istanbul, Turkey

<sup>2</sup> Department of Statistics, Ankara University, Tandogan, 06100 Ankara, Turkey

Linear mixed models were first used for animal breeding studies by [Eisenhart \(1947\)](#) and by [Henderson \(1950\)](#). As a theoretical point of view, [Harville \(1977\)](#) reviewed the maximum likelihood (ML) approach to variance component estimation related to LMMs. Based on Harville's work (1977), [Laird and Ware \(1982\)](#) introduced a two-staged model for repeated measurements and [Lindstrom and Bates \(1988\)](#) further studied on the same issue includes both Newton-Raphson and EM-algorithms. [Laird and Ware \(1982\)](#) introduced population parameters, individual effects and within-person variation at the first stage, and between-person variation at the second stage in their model. Their work included EM algorithm for ML estimates, following parameter estimations.

LMMs include more parameters, such as random effect coefficients to be estimated than classical regression-based models. Depending on that fact, integrating process includes more computational burden and the parameter estimations includes other unknown-parameters in the final equations. Therefore, some iterative algorithms such as EM-type ([Dempster et al. 1977](#)) are used in LMM to find ML estimators. EM is used for a wide variety of situations such as random effects, mixtures, log linear models, latent class and latent variable situations ([Healy and Westmacott 1956](#); [Kleinbaum 1973](#); [McLachlan and Krishnan 1997](#)) in addition to incomplete data structures.

Classical LMMs include the normality assumption for both random and error terms, but it may be too restrictive especially for repeated measures and clustered data. The existence of missing data or e- and b-outliers makes normality assumption being doubtful for the model. Even normality assumption allows applying the model easily, it is too restrictive for the data with heavy tails and/or skewed. So, it is critical for LMMs to choose a robust distribution for random effects/error terms rather than normal distribution.

There has been wide ranges of studies robustifying linear mixed models ([Lin and Lee 2008](#); [Lin 2008](#); [Pinheiro et al. 2001](#); [Verbeke and Lesaffre 1996](#)) following studies robustifying linear regression models ([Hampel et al. 1986](#); [Huber 1973](#); [Huber 1981](#); [Lange et al. 1989](#); [Zellner 1976](#)). [Verbeke and Lesaffre \(1996\)](#) used finite mixtures of normal distribution for random effects to detect the subgroups in the population. [Pinheiro et al. \(2001\)](#) proposed a linear mixed model with both random effects and residuals which were assumed to be multivariate t-distributed with the same degrees of freedom and they used an EM-type algorithm for the estimation of parameters. [Lin \(2008\)](#) proposed a t-linear mixed effects model with AR(p) dependence structure for repeated measures with thick tails and serial correlations. Likelihood inference is then required degrees-of-freedom parameter which is not required for classical LMM and the proposed model of us. In the study of [Osorio et al. \(2007\)](#), both light and heavy-tailed distributions were covered for elliptical linear models under different perturbation schemes. They conducted the study for the different types of elliptical distributions with error terms, not for the random terms.

In this study, we use multivariate Laplace distribution which is a special case of multivariate exponential power (EP) distribution ([Arslan 2010](#); [Gómez-Sánchez-Manzano et al. 2006](#); [Gómez-Sánchez-Manzano et al. 2008](#)) to handle with the situations being doubtful to assume a normal distribution. It has heavier tails than normal distribution and the number of parameters of Laplace distribution is less than t-distribution that makes the estimation procedures simpler than the estimations based on t-distribution.

Additionally, EM algorithm can be implemented easily to estimate the parameters of interest, since Laplace distribution is written as a scale mixture of normal distribution.

The paper is structured as follows: Sect. 2 describes LMMs briefly. Section 3 includes multivariate EP distribution, the mixture form of it, and also the integration of this distribution into LMMs. Section 4 includes the general scheme of EM algorithm, and EM for Laplace LMM. The application part of the study is composed of a simulation study and a real life data example at the 5th and the 6th sections, respectively. Our conclusions are presented in Sect. 7. The originality of this study comes from using Laplace distribution with LMMs and also describes ECM with this new form.

## 2 Linear mixed models

The LMM has the following form (Laird and Ware 1982)

$$y_i = X_i\beta + Z_iu_i + e_i, \quad i = 1, 2, \dots, n, \tag{1}$$

where,

$$e_i \sim N_{n_i}(0, R_i),$$

$$u_i \sim N_r(0, D),$$

$y_i$  denotes a  $(n_i \times 1)$  vector of continuous responses for the  $i$ -th subject,  $\beta$  denotes a  $(p \times 1)$  vector of unknown population parameters,  $X_i$  is a known  $(n_i \times p)$  design matrix,  $u_i$  denotes a  $(r \times 1)$  vector of unknown individual effects,  $Z_i$  is a known  $(n_i \times r)$  design matrix, and  $e_i$  denotes a  $(n_i \times 1)$  vector of residual errors assumed to be independent of  $u_i$ . The marginal distribution of  $y_i$  can be obtained as normal with mean  $X_i\beta$  and the covariance  $Z_iDZ_i' + R_i$  ( $y_i \sim N(X_i\beta, Z_iDZ_i' + R_i)$ ).

The general form of the LMM is defined as in model (1) for a given subject  $i$ . This model definition is not the only way to define the model, there is an alternative specification of the LMM which is “stacking” the vectors of the model such as  $y_i$  or  $u_i$  for all subjects vertically. This specification is used in SAS, R, SPSS, S documentations; while HLM uses the hierarchical approach of Raudenbush and Bryk (2002) (West et al. 2007).

Comprehensive work on LMMs is found in Searle et al. (1992), Verbeke and Molenberghs (2000), McCulloch and Searle (2001) and Demidenko (2004), among others. Parameter estimations are described in these studies under the assumptions above.

## 3 Linear mixed models with multivariate Laplace distribution

### 3.1 Multivariate exponential power distribution to Laplace distribution and the mixture form

The exponential power family was introduced by Box and Tiao (1973) and it has been found a place in the Bayesian modelling for robustness context. The normal scale

mixture property of EP distributions and its relationship with the class of stable distributions were discussed by West (1987). The multivariate EP distribution is defined as the generalization of multivariate normal distribution with an extra parameter  $K$  which governs the kurtosis, and this parameter indicates the difference between EP distribution and the normal distribution (Gómez-Villegas et al. 2011).

An  $n$ -dimensional random vector  $\mathbf{X} = (X_1, \dots, X_n)'$  is said to have EP distribution ( $EP_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, K)$ ) with parameters,  $\boldsymbol{\mu} \in \mathbb{R}$ ,  $\boldsymbol{\Sigma}$  is a positive definite symmetric matrix and  $K \in (0, \infty)$ , if its density function is as follow

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, K) = d |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}((\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}))^K\right), \tag{2}$$

$$d = \frac{n\Gamma(\frac{n}{2})}{\pi^{\frac{n}{2}} \Gamma(1 + \frac{n}{2K}) 2^{1+\frac{n}{2K}}},$$

$K$  is the kurtosis parameter [i.e., see (Arslan 2010; Fang et al. 1990; Gómez-Sánchez-Manzano et al. 2008; Gómez et al. 1998)]. It is obvious that if  $K=1$ , it turns into the normal distribution.

EP distribution is written as a scale mixture of normal distributions if and only if  $K \in (0, 1]$ , and it is denoted as  $\mathbf{X} \sim \text{SMN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_K)$  where the mixing distribution function  $h_K$  is the absolutely continuous distribution function with (Gómez-Sánchez-Manzano et al. 2008)

$$h_K(v) = \frac{2^{1+\frac{n}{2}-\frac{n}{2K}} \Gamma(1 + \frac{n}{2})}{\Gamma(1 + \frac{n}{2K})} v^{n-3} S_K\left(v^{-2}; 2^{1-\frac{1}{K}}\right), \quad v > 0.$$

$S_K$  is referred to stable distribution which has the characteristic function for  $K \in (0, 1)$  as (Samorodnitsky and Taqqu 1994)

$$\varphi(t) = \exp\left\{-\sigma^K |t|^K e^{-i\frac{\pi}{2}K \text{sign}(t)}\right\}.$$

Gómez-Sánchez-Manzano et al. (2008) use  $\sigma = 2^{-\frac{1}{K}}$  and the Laplace transform of its distribution function becomes  $\mathcal{L}_F(t) = \exp\left\{-\frac{1}{2}t^K\right\} = g_K(t)$ .

As a special case of multivariate EP distribution; if  $K = \frac{1}{2}$ , it is considered as the multivariate Laplace distribution (double exponential distribution) and the mixture form is denoted as  $\text{SMN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h_{1/2})$  with the following density of mixing function

$$h_{1/2}(v) = \frac{\Gamma(\frac{n}{2})}{2^{\frac{n}{2}} \Gamma(n)} v^{n-3} S_{\frac{1}{2}}\left(v^{-2}; \frac{1}{2}\right),$$

for  $y > 0$ ,  $S_{\frac{1}{2}}(y; \sigma) = \frac{1}{2} \pi^{-\frac{1}{2}} \sigma^{\frac{1}{2}} y^{-\frac{3}{2}} \exp(-\frac{1}{4} \sigma y^{-1})$  is an inverted-gamma distribution  $IG(\frac{1}{2}, \frac{\sigma}{4})$ ; so the mixing distribution becomes as (Gómez-Sánchez-Manzano et al. 2008)

$$h_{1/2}(v) = \frac{1}{2^{\frac{3n+1}{2}} \Gamma(\frac{n+1}{2})} v^n \exp\left(-\frac{1}{8}v^2\right) I_{(0,\infty)}(v). \tag{3}$$

Further, if  $X$  is a random vector with  $Laplace_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , the expectation and the variance of  $X$  are  $E(X) = \boldsymbol{\mu}$  and  $Var(X) = \frac{4\Gamma(n+2)}{n\Gamma(n)}\boldsymbol{\Sigma}$ . Scale mixture of normal distribution property of Laplace distribution provides more modeling alternatives, since it creates longer tails than normal distribution (Arslan 2010).

### 3.2 Linear mixed models with multivariate Laplace distribution

The joint distribution of  $y_i$  and  $u_i$  in classical LMM is defined as follow

$$\begin{bmatrix} y_i \\ u_i \end{bmatrix} \sim N_{n_i+r} \left( \begin{bmatrix} X_i\boldsymbol{\beta} \\ 0 \end{bmatrix}, \begin{bmatrix} Z_i\mathbf{D}Z_i' + R_i & Z_i\mathbf{D} \\ \mathbf{D}Z_i' & \mathbf{D} \end{bmatrix} \right), \quad i = 1, \dots, n. \tag{4}$$

We replace multivariate normal distribution in model (4) with multivariate Laplace distribution for both random effects and subject-specific errors (LLMM or l-LMM) by proceeding as in Lange et al. (1989) and Pinheiro et al. (2001) as follow

$$\begin{bmatrix} y_i \\ u_i \end{bmatrix} \sim Laplace_{n_i+r} \left( \begin{bmatrix} X_i\boldsymbol{\beta} \\ 0 \end{bmatrix}, \begin{bmatrix} Z_i\mathbf{D}Z_i' + R_i & Z_i\mathbf{D} \\ \mathbf{D}Z_i' & \mathbf{D} \end{bmatrix} \right), \quad i = 1, \dots, n. \tag{5}$$

From this distribution, the marginal distribution of  $y_i$  is obtained as

$$y_i \sim Laplace_{n_i}(X_i\boldsymbol{\beta}, Z_i\mathbf{D}Z_i' + R_i),$$

(Arslan 2010; Gómez et al. 1998). According to the scale mixture property of Laplace distribution and Theorem 3.1 in Gómez-Sánchez-Manzano et al. (2008), the corresponding conditional distributions of  $[y_i', u_i']'$  are as follow ( $V_i = Z_i\mathbf{D}Z_i' + R_i$ )

$$\begin{aligned} \begin{bmatrix} y_i \\ u_i \end{bmatrix} | \tau_i &\sim N_{n_i+r} \left( \begin{bmatrix} X_i\boldsymbol{\beta} \\ 0 \end{bmatrix}, \tau_i^2 \begin{bmatrix} V_i & Z_i\mathbf{D} \\ \mathbf{D}Z_i' & \mathbf{D} \end{bmatrix} \right), \\ \tau_i &\sim h_{1/2}(\tau_i). \end{aligned} \tag{6}$$

It follows from (6) that the hierarchical representations of the conditional distributions of  $y_i$  and  $u_i$  vectors are as follow

$$\begin{aligned} y_i | u_i, \tau_i &\sim N(X_i\boldsymbol{\beta} + Z_i u_i, \tau_i^2 R_i), \quad i = 1, \dots, n, \\ u_i | \tau_i &\sim N(0, \tau_i^2 \mathbf{D}), \\ \tau_i &\sim h_{1/2}(\tau_i). \end{aligned} \tag{7}$$

The maximum likelihood estimation of unknown parameters in this hierarchical representation of Laplace LMM leads to EM implementations. Different than classical LMM, there is an extra unknown term,  $\tau_i$  needs to be estimated from the data. The next section includes both parameter estimations and EM algorithm of (7).

### 4 EM algorithm

EM algorithm takes its name from the study of [Dempster et al. \(1977\)](#) and it is composed of two steps; the expectation step (E-step) and the maximization step (M-step). The conditional expectation of the complete data log likelihood is calculated for given parameter estimates and observed data is computed in E-step. M-step updates the parameter estimates by maximizing conditional expectation in the E-step. These two steps are iterated until the convergence and the final parameter is taken as MLE or a local maximizer. EM is used for a wide variety of situations such as incomplete data, random effects, mixtures, log linear models, latent class and latent variable situations ([McLachlan and Krishnan 1997](#)).

We illustrate a general scheme for EM for classical/Gaussian LMM and it has similar approach for Laplace LMM with extra CM-steps explained in the next section.

Consider model (7) and take  $u_i$  as “missing data” and the complete data is  $\{(y_i, u_i), i = 1, \dots, n\}$ , where  $\{y_i, i = 1, \dots, n\}$  is observed data. Let  $\theta = (\beta, R, D)$  denote all parameters. The complete data likelihood for subject  $i$  is given by

$$l_i^{(c)}(\theta|y_i, u_i) = \ln f(y_i|u_i, \beta, R) + \ln f(u_i|D).$$

Begin with a starting value  $\theta^0$ . At the k-th iteration, the conditional expectation is computed as following:

$$\begin{aligned} \text{E-Step : } Q(\theta|\theta^k) &= E \left[ \sum_{i=1}^n l_i^{(c)}(\theta|y_i, u_i)|y_i, \theta^{(k)} \right] \\ &= E \left[ \sum_{i=1}^n \ln f(y_i|u_i, \beta, R) + \ln f(u_i|D)|y_i, \theta^{(k)} \right]. \end{aligned}$$

M-Step requires to maximize  $Q(\theta|\theta^k)$  with respect to  $\theta$  to acquire an updated estimate  $\theta^{(k+1)}$ . At the k-th iteration of the EM algorithm, since  $\theta^{(k+1)}$  maximizes  $Q(\theta|\theta^k)$ :

$$Q(\theta^{(k+1)}|\theta^k) \geq Q(\theta|\theta^k), \quad k = 0, 1, 2, \dots$$

This property shows that EM is guaranteed to converge to a maximum ([Wu 1983](#)).

#### 4.1 EM for Laplace LMM

This section includes the maximum likelihood estimation of unknown parameters of Laplace linear mixed model (5) with EM-type algorithm. The maximization process for log-likelihood of LMM leads to some non-linear equations; consequently, solutions could be too complex to be solved. Solving these equations are not solely enough for implementations; since we need to count for acquiring our parameters to be in parameters space ([Searle et al. 1992](#)). For the maximum likelihood estimations (MLE) of the parameters, we utilize the EM algorithm by using hierarchical model with both  $\tau_i$

and  $u_i$  are treated as missing. Since maximization (M) step of EM is computationally unattractive for LMM, expectation-conditional maximization (ECM) (Meng and Rubin 1993) is used in this study. The ECM algorithm, which is the extension of EM algorithm, implemented by conditioning on some function of the parameters under estimation and replaced M-steps with computationally easier CM-step. ECM has the same stable monotone convergence property of EM (McLachlan and Krishnan 1997).

Let  $\theta = [\beta', D', R']'$  denotes all parameters,  $y = [y'_1, \dots, y'_n]'$  is observed data  $u = [u'_1, \dots, u'_n]'$  and  $\tau = [\tau_1, \dots, \tau_n]$  are treated as “missing data”, so the “complete data” are  $\{y, u, \tau\}$ . Because of the conditional structure of model (7), the joint density of complete data can be factored into the product of the conditional densities, so the complete data likelihood based on model (7) is given by

$$\ln L(\beta, D, R|y, u, \tau) = \ln L_1(\beta, R|y, u, \tau) + \ln L_2(D|u, \tau) + constant.$$

‘Constant’ represents the terms without any parameters to be estimated. E-step of EM algorithm requires conditional expectations of log-likelihood function acquired as below

$$\begin{aligned} E[\ln L(\beta, D, R|y, u, \tau)|y, \hat{\theta}] &= \\ E[\ln L_1(\beta, R|y, u, \tau)|y, \hat{\theta}] &+ E[\ln L_2(D|u, \tau)|y, \hat{\theta}] + constant \\ &= -\frac{1}{2} \sum_{i=1}^n \{(y_i - X_i\beta - Z_i\hat{u}_i)'(\hat{\tau}_i^2 R_i)^{-1}(y_i - X_i\beta - Z_i\hat{u}_i) + \ln |R_i|\} \\ &\quad -\frac{1}{2} \left\{ n \ln |D| + \sum_{i=1}^n \tau_i^{-2} u'_i D^{-1} u_i \right\} + constant. \end{aligned}$$

Letting

$$\begin{aligned} \hat{u}_i &= E(u_i | \theta = \hat{\theta}, y), \\ \hat{\Omega}_i &= cov(u_i | \theta = \hat{\theta}, y), \\ \hat{\tau}_i^2 &= E(\tau_i^2 | \theta = \hat{\theta}, y), \\ \hat{\tau}_i^{-2} &= E(\tau_i^{-2} | \theta = \hat{\theta}, y), \end{aligned}$$

$$E(u_i u'_i) = cov(u_i u'_i) + E(u_i)E(u'_i) = \hat{\Omega}_i + \hat{u}_i \hat{u}'_i,$$

we find following estimations

$$\begin{aligned} \hat{u}_i &= \hat{D} Z'_i \hat{V}^{-1} (y_i - X_i \hat{\beta}) = \hat{D} Z'_i (Z_i \hat{D} Z'_i + \hat{R}_i)^{-1} (y_i - X_i \hat{\beta}) \\ &= (\hat{D}^{-1} + Z'_i \hat{R}_i^{-1} Z_i)^{-1} Z'_i \hat{R}_i^{-1} (y_i - X_i \hat{\beta}), \\ \hat{\Omega}_i &= \hat{\tau}_i^2 \hat{D} - \hat{\tau}_i^2 \hat{D} Z'_i (\hat{\tau}_i^2 \hat{V}_i)^{-1} Z_i \hat{D} \hat{\tau}_i^2 = \hat{\tau}_i^2 (\hat{D} - \hat{D} Z'_i \hat{V}_i^{-1} Z_i \hat{D}) \\ &= \hat{\tau}_i^2 (\hat{D} - \hat{D} Z'_i (Z_i \hat{D} Z'_i + \hat{R}_i)^{-1} Z_i \hat{D}) = \hat{\tau}_i^2 (\hat{D}^{-1} + Z'_i \hat{R}_i^{-1} Z_i)^{-1}. \end{aligned}$$

$\hat{\beta}$ ,  $\hat{D}$  and  $\hat{R}_i$  are found by differentiating expected values of conditional log-likelihood functions with respect to parameters. Different from the Gaussian LMM

(1), the Laplace LMM (7) requires  $\tau_i$  that is unobserved but imputed from the data for each subject. To find  $\hat{\tau}_i^2$  and  $\hat{\tau}_i^{-2}$ , the conditional distribution of  $\tau_i$  is required. We find it with the Bayesian approach,  $f(\tau_i|y_i) = \frac{f(y_i|\tau_i)f(\tau_i)}{f(y_i)}$ . After some algebra, the final equation is found as below

$$f(\tau_i|y_i) = \frac{\Gamma(n+1)2^{\frac{n}{2}+1}}{n\Gamma(\frac{n}{2})} \tau_i^{-1} \exp \left\{ -\frac{1}{2} (y_i - X_i\beta)' (\tau_i^2 V_i)^{-1} (y_i - X_i\beta) + \frac{1}{2} ((y_i - X_i\beta)' (V_i)^{-1} (y_i - X_i\beta))^{\frac{1}{2}} \right\} \times h_{1/2}(\tau_i),$$

where  $h_{1/2}(\tau_i)$  is described in (3). Adapting this mixing function into the conditional distribution above leads to

$$f(\tau_i|y_i) = \frac{\Gamma(n)}{\Gamma(\frac{n}{2})\Gamma(\frac{n+1}{2})} 2^{\frac{1}{2}-n} \exp \left( \frac{1}{2} C_i^{\frac{1}{2}} \right) \tau_i^{n-1} \exp \left\{ -\frac{1}{2} (\tau_i^{-2} C_i + \frac{1}{4} \tau_i^2) \right\} I_{(0,\infty)}(v),$$

where,

$$C_i = (y_i - X_i\beta)' (V_i)^{-1} (y_i - X_i\beta),$$

$$f(\tau_i|y_i) \propto \tau_i^{n-1} \exp \left[ -\frac{1}{2} (\tau_i^{-2} C_i + \frac{1}{4} \tau_i^2) \right].$$

Redefining the function by letting  $\xi_i = \tau_i^2$ ,  $\tau_i = \xi_i^{1/2}$ ,  $\frac{\partial \tau_i}{\partial \xi_i} = \frac{1}{2\xi_i^{1/2}}$  gives

$$f(\xi_i|y_i) \propto \xi_i^{\frac{n-1}{2}} \exp \left[ -\frac{1}{2} (\xi_i^{-1} C_i + \frac{1}{4} \xi_i) \right] \frac{1}{2\xi_i^{\frac{1}{2}}} = \frac{1}{2} \xi_i^{\frac{n}{2}-1} \exp \left[ -\frac{1}{2} (\xi_i^{-1} C_i + \frac{1}{4} \xi_i) \right].$$

The final distribution takes the form of generalized inverse Gaussian distribution. The moment of this distribution is as follow (Barndorff-Nielsen 1978; Barndorff-Nielsen and Halgreen 1977)

$$E(\xi_i^\alpha) = \left( \frac{C_i}{1/4} \right)^{\alpha/2} \frac{M_{\frac{n}{2}+\alpha} \left( \sqrt{C_i \frac{1}{4}} \right)}{M_{\frac{n}{2}} \left( \sqrt{C_i \frac{1}{4}} \right)},$$

$M_{(\cdot)}$  is modified Bessel function of the third kind (Jorgensen 1982; Watson 1966). By using this property, the moments are as followings for  $\alpha = 1$  and  $\alpha = -1$ , respectively

$$E(\xi_i) = \left( \frac{C_i}{1/4} \right)^{1/2} \frac{M_{\frac{n}{2}+1} \left( \sqrt{C_i \frac{1}{4}} \right)}{M_{\frac{n}{2}} \left( \sqrt{C_i \frac{1}{4}} \right)}, E(\xi_i^{-1}) = \left( \frac{C_i}{1/4} \right)^{-1/2} \frac{M_{\frac{n}{2}-1} \left( \sqrt{C_i \frac{1}{4}} \right)}{M_{\frac{n}{2}} \left( \sqrt{C_i \frac{1}{4}} \right)}.$$



Having starting values of the parameters  $\theta^0 = (\beta^0, D^0, R^0)$ ; an iterative algorithm in which current values of the unknown parameters are updated in the following steps:

*E-Step* For a given  $\hat{\theta}$ , calculate  $\hat{u}_i, \hat{\tau}_i^{-2}, \hat{\tau}_i^2$ , and  $\hat{\Omega}_i \quad i = 1, \dots, n$ .

$$\begin{aligned} \hat{u}_i^{(m)} &= \left( \hat{D}^{-1(m)} + Z_i' \hat{R}_i^{-1(m)} Z_i \right)^{-1} Z_i' \hat{R}_i^{-1(m)} \left( y_i - X_i \hat{\beta}^{(m)} \right), \\ \hat{\Omega}_i^{(m)} &= \hat{\tau}_i^{2(m)} \left( \hat{D}^{-1(m)} + Z_i' \hat{R}_i^{-1(m)} Z_i \right)^{-1}, \\ \hat{\tau}_i^{2(m)} &= \left( \frac{\hat{C}_i}{1/4} \right)^{1/2} \frac{M_{\frac{n}{2}+1} \left( \sqrt{\hat{C}_i \frac{1}{4}} \right)}{M_{\frac{n}{2}} \left( \sqrt{\hat{C}_i \frac{1}{4}} \right)}, \\ \hat{\tau}_i^{-2(m)} &= \left( \frac{\hat{C}_i}{1/4} \right)^{-1/2} \frac{M_{\frac{n}{2}-1} \left( \sqrt{\hat{C}_i \frac{1}{4}} \right)}{M_{\frac{n}{2}} \left( \sqrt{\hat{C}_i \frac{1}{4}} \right)}. \end{aligned}$$

*CM-Step 1* Fix  $R_i = \hat{R}_i$  and update  $\hat{\beta}$  for  $i = 1, \dots, n$  by maximizing  $E \left[ L_1(\beta, \hat{R} | y, u, \tau) | y, \hat{\theta} \right]$  over  $\beta$ .

$$\hat{\beta}^{(m+1)} = \left( \sum_{i=1}^n \hat{\tau}_i^{-2(m)} X_i' \hat{R}_i^{-1(m)} X_i \right)^{-1} \sum_{i=1}^n \hat{\tau}_i^{-2(m)} X_i' \hat{R}_i^{-1(m)} (y_i - Z_i \hat{u}_i^{(m)}).$$

*CM-Step 2* Fix  $\beta = \hat{\beta}$  and update  $\hat{R}_i$  for  $i = 1, \dots, n$  by maximizing  $E \left[ L_1(R, \hat{\beta} | y, u, \tau) | y, \hat{\theta} \right]$  over  $R_i$ .

$$\begin{aligned} \hat{R}_i^{(m+1)} &= \frac{1}{n} \sum_{i=1}^n \hat{\tau}_i^{-2(m)} \left[ (y_i - Z_i \hat{u}_i^{(m)}) (y_i - Z_i \hat{u}_i^{(m)})' + Z_i \hat{\Omega}_i^{(m)} Z_i' \right. \\ &\quad \left. + X_i \hat{\beta}^{(m)} \hat{\beta}^{(m)'} X_i' - X_i \hat{\beta}^{(m)} (y_i - Z_i \hat{u}_i^{(m)})' \right]. \end{aligned}$$

*CM-Step 3* Update  $\hat{D}$  by maximizing  $E \left[ L_2(D | u, \tau) | y, \hat{u} \right]$  over  $D$ .

$$\hat{D}^{(m+1)} = \frac{1}{n} \sum_{i=1}^n \hat{\tau}_i^{-2(m)} (\hat{\Omega}_i^{(m)} + \hat{u}_i^{(m)} \hat{u}_i^{(m)'})$$

Repeating the cycles creates sequences of related parameters.

### 5 Simulation studies

We conducted a simulation study similar to [Pinheiro et al. \(2001\)](#) and generated the data from the mixture of normals. The “rmnorm” function in mvtnorm library in R

is used to generate the data after the model definition. In addition to their study, we added an extra scenario (2nd one) with Laplace distribution (8).

Let us define a model as below:

$$y_i = X(\beta + u_i) + e_{ij}, \quad i = 1, \dots, 27, \quad j = 1, \dots, 4, \quad X = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix}.$$

*First scenario*

$$u_i \sim ind(1 - p_b)\mathcal{N}_2(0, \mathbf{D}) + p_b f \mathcal{N}_2(0, \mathbf{D}),$$

$$e_{ij} \sim ind(1 - p_e)\mathcal{N}_4(0, \mathbf{R}) + p_e f \mathcal{N}_4(0, \mathbf{R}), \quad i = 1, \dots, 27, \quad j = 1, \dots, 4.$$

*Second scenario*

$$u_i \sim ind(1 - p_b)Laplace_2(0, \mathbf{D}) + p_b f Laplace_2(0, \mathbf{D}),$$

$$e_{ij} \sim ind(1 - p_e)Laplace_4(0, \mathbf{R}) + p_e f Laplace_4(0, \mathbf{R}),$$

$$i = 1, \dots, 27, \quad j = 1, \dots, 4. \tag{8}$$

$p_b$  and  $p_e$  denotes the expected percentage of b- and e-outliers.  $f$  denotes contamination factor.  $p_b$  and  $p_e$  are taken 0, 0.05, 0.10 values and  $f$  is taken 0, 1, 2 values for each replicates.  $Var(u_i)$  and  $Var(e_{ij})$  is the natural extension of (8)

$$E(y_i) = X\beta,$$

$$Var(y_i) = XVar(u_i)X' + Var(e_{ij}),$$

$$Var(u_i) = (1 - p_b)\mathbf{D} + p_b f^2 \mathbf{D} = [1 + (f^2 - 1)p_b] \mathbf{D},$$

$$Var(e_{ij}) = (1 - p_e)\mathbf{R} + p_e f^2 \mathbf{R} = [1 + (f^2 - 1)p_e] \mathbf{R}.$$

As a first step, we generated  $y_i$ s depending on the parameters above. Four replicates for 27 subjects with 100 Monte Carlo replicates for each scenario are generated. Then, we obtain parameter estimations of our proposed model (7) and also parameter estimations of classical LMM (1). The proposed model is implemented by generating new functions in R. For classical LMM, the *lmer* function in ‘nlme’ package is used in R. See the target values (starting values) below

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 17 \\ 0.8 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 4 & 0 \\ 0 & 0.0225 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The parameter comparisons are directly implemented for  $\beta_0$  and  $\beta_1$ . On the other hand, we need to do the transformation for  $\mathbf{D}$  comparisons defined as

$$\mathbf{D}_{Laplace} = \frac{2^2\Gamma(4)}{2\Gamma(2)} \Sigma = (12)\mathbf{D}.$$

The graphs below (Table 1) show the distributions of  $|\hat{\theta} - \theta_0|$  for each parameter. Black lines show the differences between our proposed model estimations and starting values, purple dashed lines show the differences between *lmer* function estimations and starting values in all graphs.

Because of the space limitations, we preferred to display two extreme combinations of  $f$ ,  $p_b$  and  $p_e$ . The first one is for  $f = p_b = p_e = 0$  which indicates the non-contamination factor without b- and e-outliers. The second one is  $f = 2$ ,  $p_b = 0.10$ ,  $p_e = 0.10$  located on the second part of the Table 1. We acquired similar results for Laplace LMM and classical LMM approaches, especially for the first scenario, as it is seen on the left side of the Table 1. The differences of parameter estimations from the target values (starting values) are getting higher for the second scenario which are seen on the right side of the Table 1. We observe more gaps in the differences for the second combination with 10% expected b- and e-outliers and also contamination factor, than the first one, especially for variance components of  $D$  and  $R$ . For the data sampled from a distribution shaped more like Laplace distribution, our model outperforms the classical LMM. AIC values comparisons of two models for both scenarios are located at the final section.

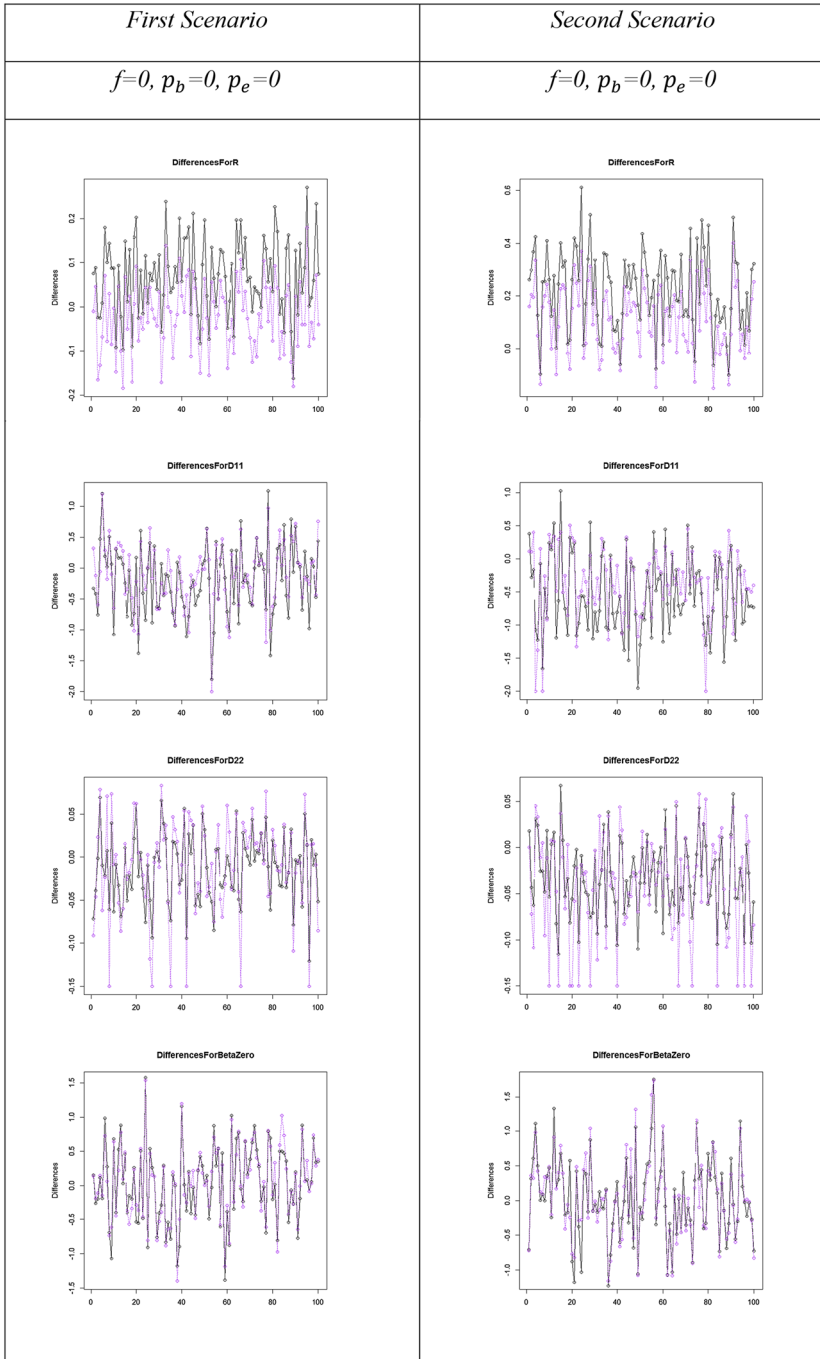
## 6 Orthodont data

Orthodont data is firstly reported by [Potthoff and Roy \(1964\)](#) and named as Orthodont Study/Data found in *nlme* library in R. The data represents the distance (in millimeters) between the pituitary and the pterygomaxillary (we call it ‘distance’ for the rest of the study) measured at 8, 10, 12 and 14 years of each girl and boy. The graph below shows age versus distance for each subject. We mainly observe a linear trend over time in addition to some within and between variations (Fig. 1).

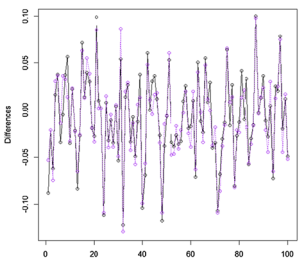
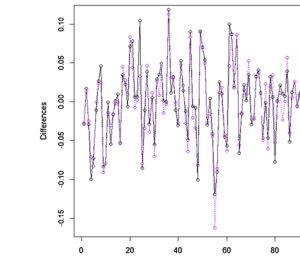
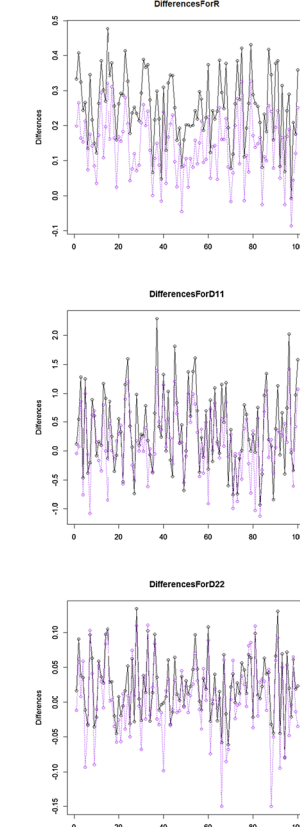
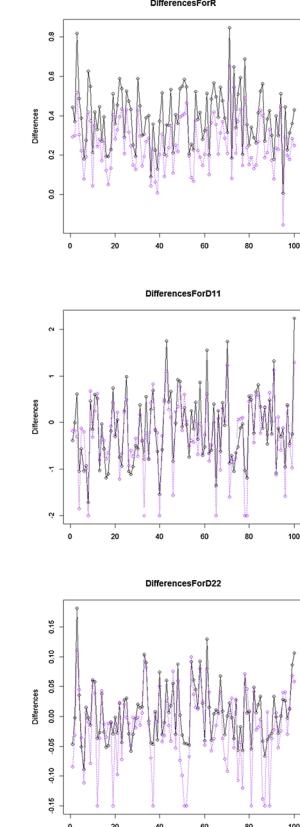
To better understand these variations, we conduct a single linear regression for each subject and acquire intercept and slope with centered age variable versus distance with *lmList* function in R. The following graphs (Fig. 2) are related to these regression models.

The figure on the left depicts the confidence intervals of intercepts and slopes with standardized data obtained with a single regression for each subject (Fig. 2). The regressions lines for each subject with age versus distance values are located on the right figure (Fig. 2). Both graphs give us a clue to fit the model on behalf of the one with random effects. The variations among confidence intervals and also regression lines are observed in both graphs. They also reveal the potential outliers of the data. For example, the slope for M13 is larger than others which may be indicator of being a b-outlier. Two observations of M09 may be an e-outlier because of the more variations around the regression line. We observe the linear trend on the distance with different ratios. It is also obvious that the first measurements for distances are different for each subject. These findings indicate random intercept and potentially random slope for the data. Using random terms in addition to fixed parameters most probably makes the model more consistent and persistent. Detailed investigations and related R codes for Orthodont data can be found in [Pinheiro and Bates \(2000\)](#).

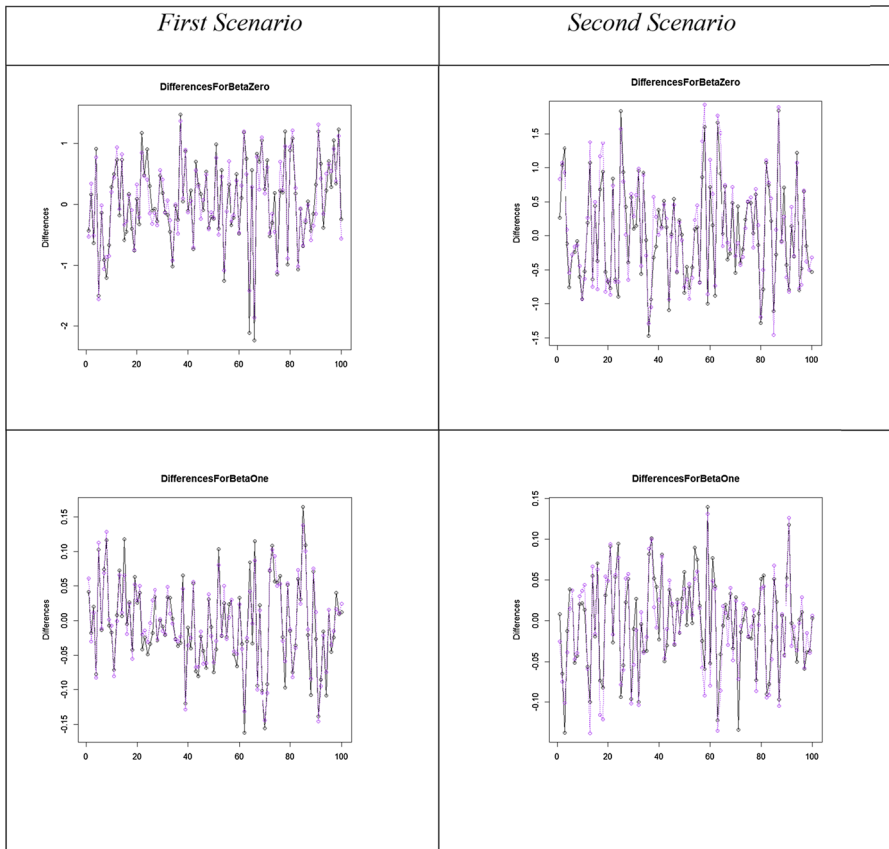
**Table 1** The figures of first scenario (on the left) and second scenario (on the right) simulations



**Table 1** continued

<i>First Scenario</i>	<i>Second Scenario</i>
	
$f=2, p_b=0.10, p_e=0.10$	$f=2, p_b=0.10, p_e=0.10$
	

**Table 1** continued



All graphs shows  $|\hat{\theta} - \theta_0|$  for LMM (with purple-dashed line) and for Laplace LMM (black-straight line)

Both classical LMM and our proposed model with Laplace distributed error and random terms are applied for this data. The related parameter estimations are located at Table 2 (the gender effect is omitted for the ease of the computation):

We observe similar behavior from both models. This is actually expected since the behavior of ‘Orthodont’ data is close to normality. However, we observe the superiority of Laplace LMM over the normal LMM when we have atypical observations in the data in simulation study. Additionally, there is a gap between  $\sigma^2$  obtained from *lmer* and our base-code. It is because of the fact that *lmer* function takes  $\mathbf{R}$  matrix as  $\sigma^2 I_{n_i}$ , as a consequence, the output includes single value. Our base code gives an estimated  $\mathbf{R}$  matrix, which has the same dimensional with target value. We use the mode of the single value decomposition of  $\mathbf{R}$  matrix as  $\sigma^2$ . The figure below shows the fitted values from our model versus classical LMM for each subject (Fig. 3).

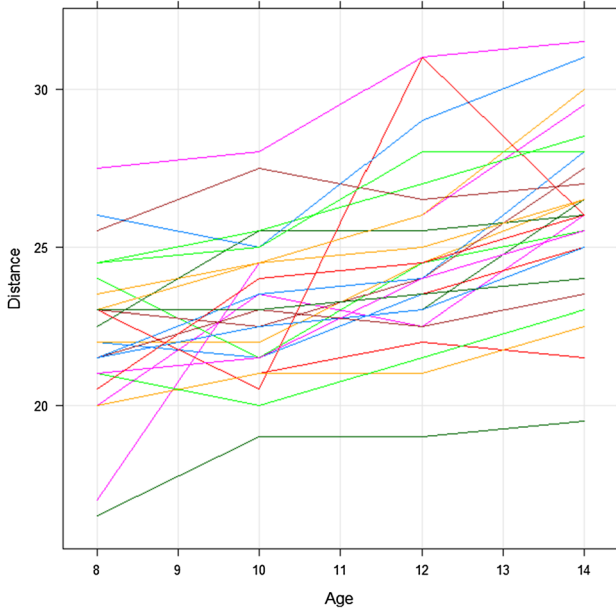


Fig. 1 Age versus distance for each subject

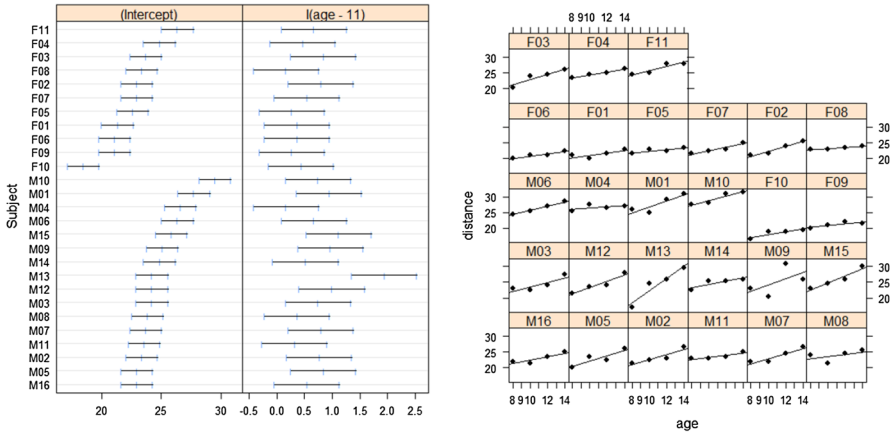


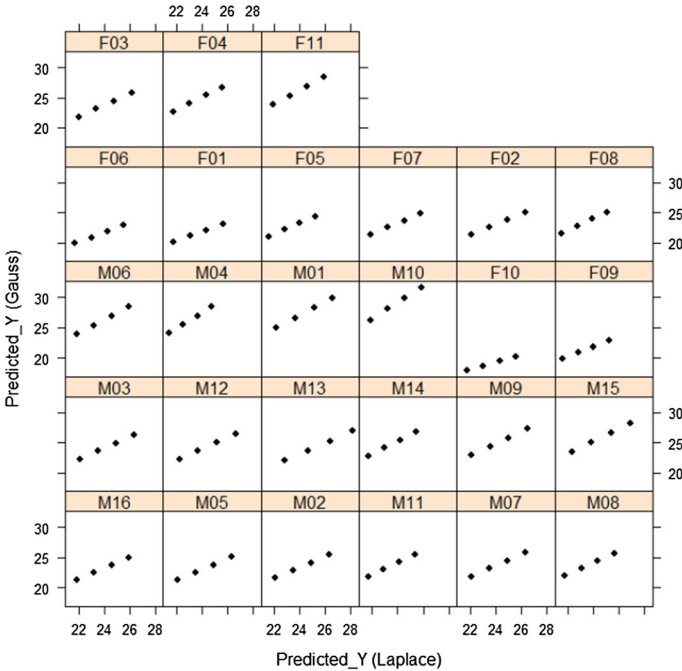
Fig. 2 Fitted parameters with confidence intervals for each subject on the left. Fitted regression lines for each subject with age versus distance on the right

### 7 Conclusion and discussion

In this study, we aimed to enhance LMMs with different distributions than normal distribution. As a first step, we used Laplace-normal hierarchical structure and implemented corresponding mathematical inferences. To the extend of our knowledge, it is the first attempt in literature using Laplace distribution in LMMs with both random effects and error terms are assumed to follow Laplace distributions. It gives us the

**Table 2** Parameter estimation results from LMM versus Laplace LMM

Parameters	Gauss	Laplace
$\beta_0$	16.761	16.529
$\beta_1$	0.660	0.674
$D_{11}$	1.825	2.739
$D_{22}$	0.021	0.030
$\sigma^2$	1.859	2.192



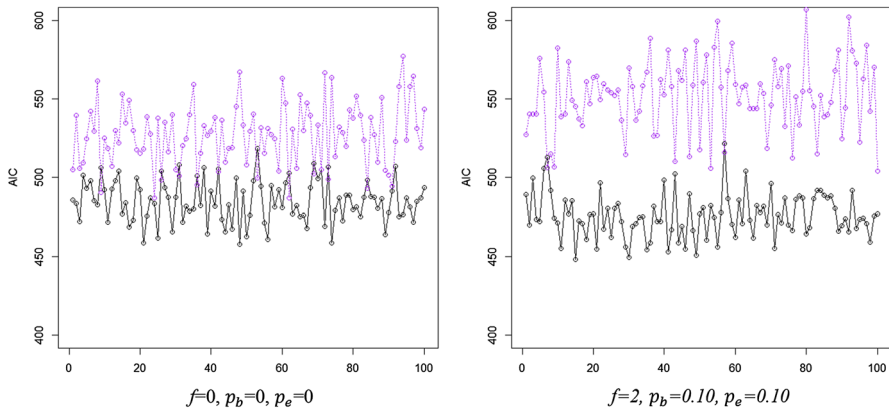
**Fig. 3** Fitted values from proposed method versus classical LMM

opportunity to use a robust distribution with less parameter than t-distribution (e.g., [Pinheiro et al. 2001](#); [Lin 2008](#)) as an alternative to normal distribution.

Laplace LMM allows to use hierarchical structure and to define the conditional distributions of response and error/random terms. After defining these conditional distributions given the scalar with a mixing distribution (7), the conditional distribution of that scalar is found for estimation process of parameters. The originality of this paper also comes from the definition of this conditional distribution of the scalar ( $\tau_i$ ), since we come up with the generalized inverse Gaussian (GIG) distribution after some algebra and defined the conditional distribution of this scalar. Estimation process and EM algorithm is finalized using the properties of GIG distribution.

Simulation studies are implemented with two different scenarios; the first one is generated from mixtures of normal distributions, while the second scenario is gen-





**Fig. 4** AIC values for LMM (purple-dashed line) and for Laplace LMM (black straight line)

erated from mixtures of Laplace distributions with/without contamination factor. We apply these inferences with orthodont data and acquired similar predictions with classical LMM. The parameter estimations from the proposed method are close to classical LMM estimations, since the behavior of this data is very likely normal.

See the AIC values of LMM (purple-dashed line) and Laplace LMM (black straight line) for the second scenario (samples are generated from mixtures of Laplace distribution) at Fig. 4. The one on the left is for  $f = 0$ ,  $p_b = 0$ ,  $p_e = 0$ ; while the one on the right is for  $f = 2$ ,  $p_b = 0.10$ ,  $p_e = 0.10$ . For each model, AICs for Laplace LMM is smaller than AICs for LMM. And also, the gap between AIC values is getting bigger for the case with contamination factor and higher expected values for b- and e-outliers (Fig. 4).

For the future work, we plan to use skew Laplace distribution for LMM as an alternative to skew-normal (Arellano-Valle et al. 2005) and skew-t distributions (Choudhary et al. 2014; Lachos et al. 2009). We also, want to enhance and redefine our base R-code in R-package format and add additional properties such as hypothesis tests for parameters.

**Acknowledgements** The authors thank the Editor and two anonymous referees for valuable suggestions that greatly improved the paper.

## References

- Arellano-Valle RB, Bolfarine H, Lachos VH (2005) Skew-normal linear mixed models. *J Data Sci* 3:415–438
- Arslan O (2010) An alternative multivariate skew Laplace distribution: properties and estimation. *Stat Pap* 51:865–887
- Barndorff-Nielsen O (1978) Hyperbolic distributions and distributions on hyperbolae. *Scand J Stat* 5:151–157
- Barndorff-Nielsen O, Halgreen C (1977) Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z Wahrscheinlichkeitstheorie verw Gebiete* 38:309–311
- Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Wiley, New York

- Choudhary PK, Sengupta D, Cassey P (2014) A general skew-t mixed model that allows different degrees of freedom for random effects and error distributions. *J Stat Plan Inference* 147:235–247
- Demidenko E (2004) *Mixed models: theory and applications*. Wiley, New York
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B (Methodol)* 39:1–38
- Eisenhart C (1947) The assumptions underlying the analysis of variance. *Biometrics* 3:1–21
- Fang KT, Kotz S, Ng FW (1990) *Symmetric multivariate and related distributions*. Chapman and Hall, London
- Gómez E, Gomez-Viilegas MA, Marín JM (1998) A multivariate generalization of the power exponential family of distributions. *Commun Stati Theory Methods* 27:589–600
- Gómez-Sánchez-Manzano E, Gómez-Villegas MA, Marín JM (2006) Sequences of elliptical distributions and mixtures of normal distributions. *J Multivar Anal* 97:295–310
- Gómez-Sánchez-Manzano E, Gómez-Villegas MA, Marín JM (2008) Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications. *Commun Stat Theory Methods* 37:972–985
- Gómez-Villegas M, Gómez-Sánchez-Manzano E, Maín P, Navarro H (2011) The effect of non-normality in the power exponential distributions. Understanding complex systems. In: Pardo L, Balakrishnan N, Gil M (eds) *Modern mathematical tools and techniques in capturing complexity*. Springer, Berlin Heidelberg, pp 119–129
- Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) *Robust statistics: the approach based on influence functions*. Wiley, New York
- Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* 72:320–338
- Healy M, Westmacott M (1956) Missing values in experiments analysed on automatic computers. *J R Stat Soc Ser C (Appl Stat)* 5:203–206
- Henderson CR (1950) Estimation of genetic parameters. *Ann Math Stat* 21:309–310
- Huber PJ (1973) Robust regression: asymptotics, conjectures and Monte Carlo. *Ann Math Stat* 1:799–821
- Huber PJ (1981) *Robust statistics*. Wiley, New York
- Jørgensen B (1982) Statistical properties of the generalized inverse Gaussian distributions, vol 9. *Lecture Notes in statistics*, Springer, New York
- Kleinbaum DG (1973) A generalization of the growth curve model which allows missing data. *J Multivar Anal* 3:117–124
- Lachos VH, Dey DK, Cancho VG (2009) Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective. *J Stat Plan Inference* 139:4098–4110
- Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38:963–974
- Lange KL, Little RJA, Taylor JMG (1989) Robust statistical modeling using the t distribution. *J Am Stat Assoc* 84:881–896
- Lin TI (2008) Longitudinal data analysis using t linear mixed models with autoregressive dependence structures. *J Data Sci* 6:333–355
- Lin T, Lee J (2008) Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Stat Med* 27:1490–1507
- Lindstrom MJ, Bates DM (1988) Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J Am Stat Assoc* 83:1014–1022
- McCulloch CE, Searle SR (2001) *Generalized, linear, and mixed models*. Wiley, New York
- McLachlan GJ, Krishnan T (1997) *The EM algorithm and extensions*. Wiley Series in Probability and Statistics, New York
- Meng X-L, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80:267–278
- Osorio F, Paula GA, Galea M (2007) Assessment of local influence in elliptical linear models with longitudinal structure. *Comput Stat Data Anal* 51:4354–4368
- Pinheiro JC, Bates DM (2000) *Mixed-effects models in S and S-Plus*. Springer-Verlag, New York
- Pinheiro JC, Liu C, Wu YN (2001) Efficient algorithms for Robust estimation in linear mixed-effects models using the multivariate t distribution. *J Comput Graph Stat* 10:249–276
- Potthoff RF, Roy SN (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51:313–326
- Raudenbush SW, Bryk AS (2002) *Hierarchical linear models: applications and data analysis methods*. Sage Publications, Newbury Park

- Samorodnitsky G, Taqqu MS (1994) Stable non-Gaussian random processes: stochastic models with infinite variance. Chapman and Hall, New York
- Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York
- Verbeke G, Lesaffre E (1996) A linear mixed-effects model with heterogeneity in the random-effects population. *J Am Stat Assoc* 91:217–221
- Verbeke G, Molenberghs G (2000) Linear mixed models for longitudinal data. Springer, New York
- Watson GN (1966) A treatise on the theory of Bessel functions. Cambridge University Press, London
- West BT, Welch KB, Galecki AT (2007) Linear mixed models: a practical guide using statistical software. CRC Press LLC, Boca Raton
- West M (1987) On scale mixtures of normal distributions. *Biometrika* 74:646–648
- Wu CFJ (1983) On the convergence properties of the EM algorithm. *Ann Stat* 11:95–103
- Zellner A (1976) Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms. *J Am Stat Assoc* 71:400–405