The Canadian Journal of Statistics
Vol. 37, No. 2, 2009, Pages 266–281
La revue canadienne de statistique

Robust transformation mixed-effects models for longitudinal continuous proportional data

Peng ZHANG[1]*, Zhenguo QIU[2], Yuejiao FU[3] and Peter X.-K. SONG[4]

[1]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*
[2]*Division of Cancer Epidemiology, Prevention and Screening, Alberta Health Services, Edmonton, Alberta, Canada T5J 3H1*
[3]*Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada M3J 1P3*
[4]*Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109-2029, USA*

Key words and phrases: Logit-normal distribution; logit-*t* distribution; outliers; robustness.

MSC 2000: Primary 62J05; secondary 62F35.

Abstract: The authors propose a robust transformation linear mixed-effects model for longitudinal continuous proportional data when some of the subjects exhibit outlying trajectories over time. It becomes troublesome when including or excluding such subjects in the data analysis results in different statistical conclusions. To robustify the longitudinal analysis using the mixed-effects model, they utilize the multivariate *t* distribution for random effects or/and error terms. Estimation and inference in the proposed model are established and illustrated by a real data example from an ophthalmology study. Simulation studies show a substantial robustness gain by the proposed model in comparison to the mixed-effects model based on Aitchison's logit-normal approach. As a result, the data analysis benefits from the robustness of making consistent conclusions in the presence of influential outliers. *The Canadian Journal of Statistics* 37: 266–281; 2009 © 2009 Statistical Society of Canada

Résumé: Les auteurs proposent un modèle linéaire à effets mixtes robuste pour les études longitudinales des données de proportions continues. Ce modèle s'applique lorsque des sujets ont des trajectoires aberrantes. Le choix d'inclure ou non de tels sujets est difficile à faire surtout lorsque l'analyse des données résulte à des conclusions statistiques différentes. Afin de rendre robuste les analyses longitudinales dans les modèles à effets mixtes, ils suggèrent d'utiliser une distribution T multidimensionnelle pour les effets aléatoires ou les termes d'erreurs. L'estimation et l'inférence à l'aide du modèle suggéré sont faites et illustrées à l'aide de vraies données provenant d'une étude ophtalmologique. Des études de simulation montrent un gain de robustesse substantiel du modèle suggéré par rapport au modèle à effets mixte basé sur l'approche logit-normale d'Aitchison. En conséquence, l'analyse des données bénéficie de la robustesse en obtenant des conclusions cohérentes en présence de valeurs aberrantes influentes. *La revue canadienne de statistique* 37: 266–281; 2009 © 2009 Société statistique du Canada

1. INTRODUCTION

In many biomedical studies, the longitudinal continuous measure with a certain constraint is recorded as a primary response variable of interest. Frequently proportions are observed, which are confined to (0, 1). For instance, percent decrease in glomerular filtration rate at different follow-up times from the baseline are recorded and analysed in nephrological studies, in which such a measure of rate characterizes more directly the loss of renal function than original serum

* *Author to whom correspondence may be addressed.*
E-mail: pengz@ualberta.ca

creatinine levels. Oncologists are often more interested in percent change of tumour size than actual size since the baseline tumour size at diagnosis varies from patient to patient. One key feature of this type of measurement is its confinement within a finite interval, often the unit interval. One issue, that is, often ignored in the analysis of such data is the presence of outliers, which appears more subtle than in the analysis of unconstrained data. This is because a small difference between two proportions, especially when both are close to the interval limit 0 or 1, can potentially correspond to a very large discrepancy in the real line scale. Therefore, outliers in such data are likely hidden as human eyes fail to detect them easily in plots of the data. Misidentifying influential outliers can incur misleading conclusions.

In this article we utilize the eye surgery data of Meyers et al. (1992) to illustrate such subtlety and the merit of robust analysis. The data were collected from a prospective ophthalmology study concerning the use of intraocular gas (C_3F_8) to provide internal tamponade of retinal breaks in the eye in complex retinal surgeries. The outcome variable was the percent of gas left in the eye. The gas, with three different concentration levels, 15%, 20%, and 25%, was injected into the eye before surgery for 31 patients. They were then followed three to eight (average of 5) times over a 3-month period, and the volume of gas in the eye at the follow-up times was recorded as a percentage of the initial gas volume. The primary interest was to investigate whether concentration levels of the gas injected in patients' eyes affect the decay rate of the gas. Figure 1a shows all gas volume decay trajectories across three gas concentration levels. None of the patients' trajectories looks suspicious, except the very bottom one (Patient #15) in the 25% level group, which appears isolated from the others in the same group. The logit-transformed gas volumes are shown in Figure 1b, where the measurements are now in scales from $-\infty$ to ∞, and the solid lines represent the LOESS smoothing curves, respectively. Figure 1b clearly indicates that the separation of patient #15 disappears. However, it is interesting to note that in the 15% concentration group, patient #16's trajectory appears to deviate from the others in the same group. This outlying patient is confirmed by Qiu, Song & Tan (2008) using a diagnostic plot developed in Jensen & Birch (2009) and Williams et al. (2006).

To build a linear mixed model for the logit-transformed proportion, we examine Figure 1b and conclude the following: (i) The LOESS curves appear virtually linear, suggesting that the time covariate should enter into the linear model in a logarithmic scale; (ii) the intercepts of the three concentration levels differ substantially; and (iii) the slopes of the LOESS curves in the 25% and 20% levels may be lower than that of the 15% level. To be specific, let y_{ij} be the proportional observation for the ith patient at the jth follow-up time. Conditional on the random intercept b_{0i}, the logit-transformed variables $\text{logit}(y_{ij})$ are independent and normally distributed with the mean μ_{ij} and the variance σ^2. The conditional mean, μ_{ij}, is modelled by both main and interaction effects as follows:

$$\begin{aligned}\mu_{ij} = {} & \beta_0 + b_{0i} + \beta_1 \log(t_{ij}) + \beta_2 \text{GAS20}_i + \beta_3 \text{GAS25}_i + \beta_4 \text{GAS20}_i \times \log(t_{ij}) \\ & + \beta_5 \text{GAS25}_i \times \log(t_{ij}), \end{aligned} \tag{1}$$

where the random intercepts $b_{0i} \sim N(0, \tau^2)$, and GAS20 and GAS25 are two dummy variables for gas concentration levels 20% and 25%, and t_{ij} is the time in days from gas injection. Here level 15% is regarded as the reference, and the random intercept b_0 describes the patient-specific proportion of the remaining gas volume 1 day after the gas injection. Such cross-patient heterogeneity of the initial gas percent is evident, as pointed out above.

Using SAS PROC MIXED, we fit the data with model (1), and found that both GAS20 and GAS25 covariates are not statistically significant (see the first row in Table 5), with P-values 0.338 and 0.132, respectively. When we excluded patient #16 from the analysis, using exactly the same model (1) and the same software, we found that covariate GAS25 became statistically

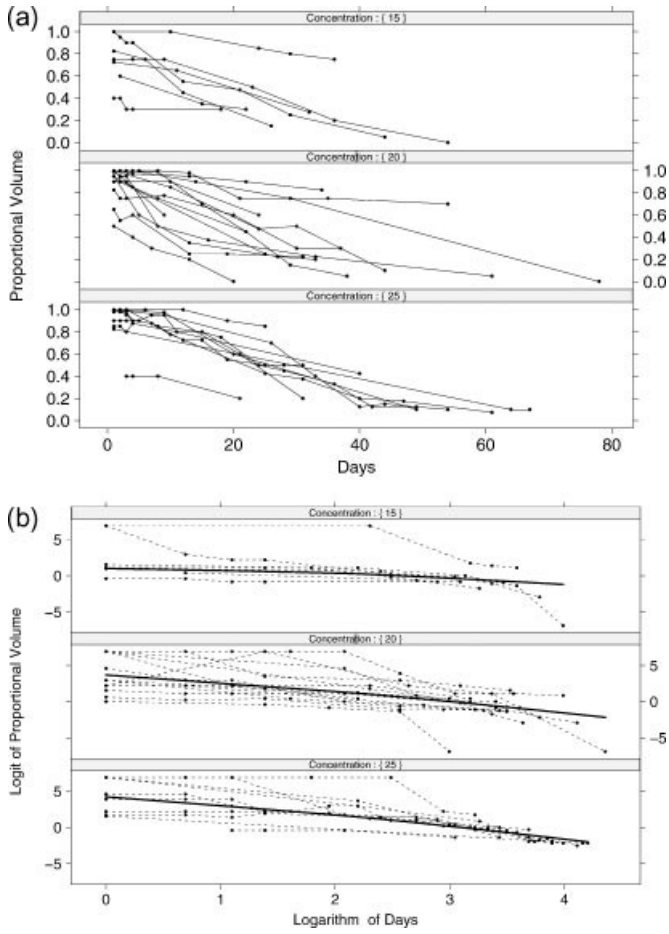

FIGURE 1: (a) Longitudinal plots of observed proportions of the remaining gas volumes in the surgical eyes across three gas concentration levels. (b) Longitudinal plots of logit-transformed proportions of gas volumes with LOESS smoothing curves across three gas concentration levels.

significant (see the second row in Table 5), with P-value 0.018. This difference in the two analyses is not surprising because the normal distributions assumed in model (1) are generally incapable of accommodating outlying data or unduly large (or small) measurements. Practically, whether or not to conclude the significance of covariate GAS25 is troublesome, which will affect making related clinical decisions.

To overcome the impasse, in this article we propose a solution, in pursuit of a robust statistical inference based on the utility of the t-distribution instead of the normal distributions, which results in a logit-t model. In the literature, Lange, Little & Taylor (1989), Pinheiro, Liu & Wu (2001), and Song, Zhang & Qu (2007) have observed the robustness gain from the utility of t-distributions in regression analysis. Due to the fact that the t-distribution is heavy-tailed, especially when the degrees of freedom are low, it can naturally accommodate extreme observations. The method proposed in Song, Zhang & Qu (2007) considers no constraints on the response variable and thus may be regarded as a special case of our development in this article with the identity transformation. Moreover, the logit-t model also includes Aitchison's logit-normal model (Aitchison, 1986) as a special case when the t-distributions have large degrees of freedom.

With regard to the specification of a robust mixed-effects model, the t-distribution can be used to relax the normal distribution assumption either on the error terms or on the random effects. According to Pinheiro, Liu & Wu (2001), outliers can arise from either the source of population heterogeneity (i.e., random effects) or the source of random errors, respectively referred to as b-outliers or e-outliers. Thus, there are four types of models, depending on which sources outliers arise; they are, normal–normal (NN), normal–t (NT), t–normal (TN), and t–t (TT) transformation mixed-effects models. For example, the NT model refers to a transformation mixed-effects model with normally distributed random effects and t-distributed error terms in response to the situation where outliers arise only from the error terms. Through comparisons of these four models, not only can we identify influential outliers but we can also determine the source from which identified outliers arise.

In previous analyses of longitudinal proportional data, Song & Tan (2000) and Song, Qiu & Tan (2004) proposed marginal models based on the GEE method, and Qiu, Song & Tan (2008) proposed a simplex generalized linear mixed model (GLMM) for such data. All these analyses assumed the simplex distribution (Jørgensen, 1997) for the proportional response, with the logit link function on the mean in the setting of generalized linear models. In contrast, in this article we model the median of the logit-transformed proportion using the t-distribution. The resulting model is not in the paradigm of the generalized linear models because the t-distribution does not belong to the class of exponential dispersion models (Jørgensen, 1997). Moreover, the interpretations are not the same regarding the effects of the covariates as those of simplex GLMM given in Qiu, Song & Tan (2008).

The rest of the article is organized as follows. Section 2 extends Aitchison's logit-normal model to the logit-t model. Section 3 presents a likelihood-based inference in the proposed model. Section 4 is devoted to simulation studies to assess the performance of the proposed model and inference, especially in the aspect of robustness against outliers. Section 5 provides a detailed analysis of the eye surgery data, and some concluding remarks are given in Section 6. The Appendix includes related technical details regarding the derivatives of the log-likelihood function.

2. ROBUST LINEAR MIXED-EFFECTS MODELS

To specify the robust LMM for longitudinal proportional data, let us first define the logit-t (LT) distribution. The density of a q-dimensional t distribution, $Mt_q(d, \Omega)$, with d degrees of freedom and a positive definite scale matrix Ω, is given by

$$p(\mathbf{x}|\boldsymbol{\tau}) = \frac{\Gamma((d+q)/2)}{(\pi d)^{q/2}\Gamma(d/2)}|\Omega(\boldsymbol{\tau})|^{-1/2}\left\{1+\frac{\mathbf{x}'\Omega(\boldsymbol{\tau})^{-1}\mathbf{x}}{d}\right\}^{-(q+d)/2}, \quad \mathbf{x}\in\mathcal{R}^q, \tag{2}$$

where $\boldsymbol{\tau}$ denotes a set of distinct parameters in the scale matrix Ω. It is known that the t-distribution given in (2) has mean $\mathbf{0}$ and variance matrix $D(\boldsymbol{\tau}) = (d)/(d-2)\Omega(\boldsymbol{\tau})$, $d > 2$; see Fang, Kotz & Ng (1990) and Kotz & Nadarajah (2004).

For convenience, we rewrite the above density (2) in terms of the variance matrix $D(\boldsymbol{\tau})$ as follows:

$$p(\mathbf{x}|\boldsymbol{\tau}) = \frac{\Gamma((d+q)/2)}{\{\pi(d-2)\}^{q/2}\Gamma(d/2)}|D(\boldsymbol{\tau})|^{-1/2}\left\{1+\frac{\mathbf{x}'D(\boldsymbol{\tau})^{-1}\mathbf{x}}{d-2}\right\}^{-(q+d)/2}, \quad \mathbf{x}\in\mathcal{R}^q.$$

Consequently, the one-dimensional t-density is

$$p(x|\tau^2) = \frac{\Gamma\{(d+1)/2\}}{\sqrt{\pi(d-2)}\tau\Gamma(d/2)} \left\{1 + \frac{x^2}{(d-2)\tau^2}\right\}^{-(d+1)/2}, \quad x \in \mathcal{R},$$

where τ^2 is the variance of the t-distribution. When the degrees of freedom d tend to ∞, the t-distribution converges to the normal distribution with the same mean and variance. Therefore, a t-distribution with a large d can be regarded as an approximation to the normal distribution. The random vector $\mathbf{x}$ is said to follow a q-dimensional t distribution with mean $\boldsymbol{\mu}$, variance matrix $D(\boldsymbol{\tau})$ and degrees of freedom d, if $\mathbf{x} - \boldsymbol{\mu} \sim Mt_q(d, D(\boldsymbol{\tau}))$, denoted by $\mathbf{x} \sim Mt_q(d, \boldsymbol{\mu}, D(\boldsymbol{\tau}))$. Note that the t-distribution is symmetric around the mean $\boldsymbol{\mu}$, so the $\boldsymbol{\mu}$ is also the median. If a componentwise transformation $\mathbf{x} = \log(\mathbf{y}/(1-\mathbf{y})) \sim Mt_q(d, \boldsymbol{\mu}, D(\boldsymbol{\tau}))$ is made, then $\mathbf{y}$ is said to follow the logit-t distribution $LT_q(d, \boldsymbol{\mu}, D(\boldsymbol{\tau}))$, denoted by $\mathbf{y} \sim LT_q(d, \boldsymbol{\mu}, D(\boldsymbol{\tau}))$. Since the logit transformation is monotonic, medians of $\mathbf{x}$ and $\mathbf{y}$ are invariant, but their means are not. This fact gives rises to convenience on the choice of median as the primary population attribute for interpreting the proposed robust LMM throughout this article.

It follows from the logit transformation that the density function of the distribution LT(d, $\mu = 0$, τ^2) is given by

$$p(y|\tau^2) = \frac{\Gamma\{(d+1)/2\}}{\sqrt{\pi(d-2)}\tau\Gamma(d/2)\{y(1-y)\}} \left\{1 + \frac{\log\left(\frac{y}{1-y}\right)^2}{(d-2)\tau^2}\right\}^{-(1+d)/2}, \tag{3}$$

where for the one-dimensional case, the subscript q is omitted in the notation for convenience. Figure 2 displays one-dimensional density functions (in solid lines) of the LT distribution, LT(d, $\mu = 0$, τ^2), at different degrees of freedom $d = 3, 6, 9$ and different values of the dispersion $\tau^2 = 0.1, 1, 3$, together with the logit-normal (LN) densities (in dashed lines) with the same means and variances.

From Figure 2, it is easy to see that (i) with a small variance, the LT distribution has heavier tails than the LN distribution, and the LT distribution approaches the LN distribution when the degrees of freedom increase; and (ii) when the variance is large, the LN density exhibits two modes as opposed to one mode in the density of the LT distribution. The bimodality implies that it is more likely to observe proportions close to the boundary values 0 and 1 than to observe data around the mean or the median (here the mean and the median are the same because of the symmetry) according to the LN distribution. This creates some difficulty in the interpretation of the data generation mechanism, as well as in the maximum likelihood estimation due to the potential problem of multiple roots. Unfortunately, our example of the eye surgery data is found to be in such a situation (see Section 5 for details). Therefore, it is appealing to develop a new model that cannot only handle outlying longitudinal data points but also overcome any potential bimodality. It is worth pointing out that the bimodality is not a fixed feature for the LN distribution and it may disappear if μ is not zero and the variance τ^2 is small.

A robust LMM is specified as follows. Let y_{ij}, $j = 1, \ldots, n_i$ be the jth repeated measurement of subject i. Conditional on random effects $\mathbf{b}_i$, y_{ij} are independently distributed according to LT(d, μ_{ij}, σ^2), where the q-dimensional random effects $\mathbf{b}_i, i = 1, \ldots, n$, are independent and each $\mathbf{b}_i \sim Mt_q(\tilde{d}, 0, D(\boldsymbol{\tau}))$, with variance matrix $D(\boldsymbol{\tau})$. The conditional median μ_{ij} takes the form

$$\mu_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i, \tag{4}$$

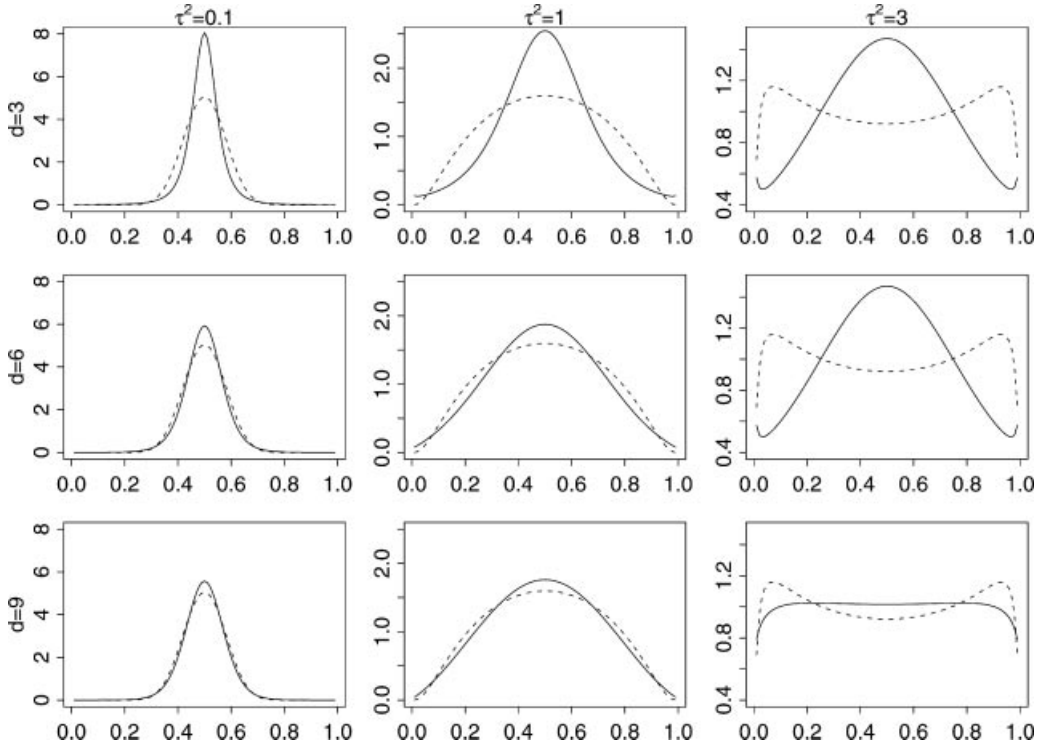

FIGURE 2: Density functions of the logit-t (LT) distribution (solid lines) and the logit-normal (LN) distribution (dashed lines) with dispersion $\tau^2 = (0.1, 1, 3)$ from left to right and degrees of freedom of LT distributions $d = (3, 6, 9)$ from top to bottom.

where $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ are, respectively, the matrices of covariates corresponding to the fixed effects and random effects for the ith subject at the jth time.

Consequently, the normal–normal (NN) LMM corresponds to $d = \tilde{d} = \infty$, the normal–t (NT) LMM, $d < \infty$ and $\tilde{d} = \infty$, the t–normal (TN) LMM, $d = \infty$ and $\tilde{d} < \infty$, and the t–t (TT) LMM, $d = \tilde{d} < \infty$. Note that equal degrees of freedom are required in the TT LMM for the sake of identifiability pertinent to the estimation of the degrees of freedom (Song, Zhang & Qu, 2007). These four models form a primary platform for us to contrast influences of outliers in estimation and inference, as well as to determine from which sources outliers arise.

3. ESTIMATION AND INFERENCE

We develop the maximum likelihood estimation and inference procedures in the proposed models. Suppose that the random effects $\mathbf{b}_i$ follow a q-dimensional distribution with density $p(\mathbf{b}_i|\boldsymbol{\tau}, d)$ and that conditional on $\mathbf{b}_i$ the response vector $\mathbf{y}_i$ follows an n_i-dimensional distribution with density $p(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\eta})$, where $\boldsymbol{\eta} = (\boldsymbol{\beta}', \sigma^2, d)'$. Then, the likelihood function is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int_{\mathcal{R}^q} p(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\eta}) p(\mathbf{b}_i|\boldsymbol{\tau}, d) d\mathbf{b}_i, \tag{5}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\tau}', \sigma^2, d)'$ denotes the vector of all the model parameters to be estimated. The involved integral has a closed form for both NN LMM and TT LMM, but not for NT LMM or TN

LMM. In this article we invoke the Gauss–Hermite quadrature method to numerically evaluate an integral that has no analytic form.

For the TT LMM, it is easy to show that the likelihood for subject i is

$$L_i(\boldsymbol{\theta}) = \frac{\{y_{i1}(1-y_{i1})\cdots y_{in_i}(1-y_{in_i})\}^{-1}\Gamma\left(\frac{d+n_i}{2}\right)}{\{\pi(d-2)\}^{n_i/2}\Gamma(\frac{d}{2})|V_i|^{1/2}}\left\{1+\frac{(\mathbf{w}_i - X_i\boldsymbol{\beta})'V_i^{-1}(\mathbf{w}_i - X_i\boldsymbol{\beta})}{d-2}\right\}^{-(d+n_i)/2}, \tag{6}$$

where $\mathbf{w}_i = (\text{logit}(y_{i1}), \ldots, \text{logit}(y_{in_i}))'$ and $V_i = Z_i D(\boldsymbol{\tau}) Z_i' + \sigma^2 I_{n_i}$. It follows immediately that the log-likelihood of $\boldsymbol{\theta}$ for the ith subject is given, subject to a constant, by

$$\ell_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|V_i| + \log\left\{\frac{\Gamma\left(\frac{d+n_i}{2}\right)}{\Gamma\left(\frac{d}{2}\right)(d-2)^{n_i/2}}\right\} - \frac{d+n_i}{2}\log\left\{1+\frac{(\mathbf{w}_i - X_i\boldsymbol{\beta})'V_i^{-1}(\mathbf{w}_i - X_i\boldsymbol{\beta})}{d-2}\right\}. \tag{7}$$

For TN LMM and NT LMM, we list those formulas in the Appendix.

The maximum likelihood estimation of $\boldsymbol{\theta}$ can be obtained under a standard route by maximizing the log-likelihood function $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. This may be done by solving the score equation $\dot{\ell}(\boldsymbol{\theta}) = 0$, where $\dot{\ell}(\boldsymbol{\theta})$ is the first order derivative of the log-likelihood $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. We use a Gauss–Newton type algorithm (Ruppert, 2005) to search for the roots of the score equation, in which the second order derivatives of the log-likelihood are not needed in iterations. The key feature of this algorithm is to take step-halving that guarantees a steady increase in the likelihood at each iteration. Precisely, the $(k+1)^{\text{th}}$ iteration proceeds as

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k + \delta B^{-1}(\boldsymbol{\theta}^k)\dot{\ell}(\boldsymbol{\theta}^k),$$

where $B(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\dot{\ell}_i(\boldsymbol{\theta})\{\dot{\ell}_i(\boldsymbol{\theta})\}'$ is an approximate Fisher information, and δ is the step-halving term that usually starts at 1 and halves each time until $\ell(\boldsymbol{\theta}^{k+1}) > \ell(\boldsymbol{\theta}^k)$ holds in each iteration. The asymptotic covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$ is estimated by the inverse of $B(\hat{\boldsymbol{\theta}})$ evaluated at the output from the last iteration.

The Appendix includes all the detail concerning the first order derivatives, $\dot{\ell}(\boldsymbol{\theta})$, for the three models, the TT LMM, the NT LMM, and the TN LMM. To deal with constraints on some parameters, such as $\sigma^2 > 0$, $\tau^2 > 0$, and $d > 2$, in the estimation we simply take the log transformation on each of them.

4. SIMULATION

We conducted a simulation study to compare the proposed LT model with the LN model with respect to robustness against both b-outliers and e-outliers. We simulated longitudinal proportional data with b-outliers or e-outliers according to the following two steps. First, we chose a model similar to that used in the eye surgery data analysis in Section 5. That is,

$$w_{ij} = \beta_0 + \beta_1\log(t_j) + \beta_2 x_i + b_{0i} + \varepsilon_{ij}, \quad j = 1, \ldots, 5, \tag{8}$$

where the group covariate x_i was set as 0 or 1, with each subgroup containing 25 subjects, and the time covariate $t_j = 1, 2, \ldots, 5$. The true values of the regression coefficients were set as $(\beta_0, \beta_1, \beta_2) = (2.0, -0.8, 1.0)$, the random effects b_{0i} were *i.i.d.* $N(0, \tau^2)$ with the true $\tau^2 = 1$,

and errors ε_{ij} were *i.i.d.* $N(0, \sigma^2)$ where the true $\sigma^2 = 1$. The proportional observations were produced by $y_{ij} = \frac{\exp(w_{ij})}{1+\exp(w_{ij})} i = 1, \ldots, 25, j = 1, \ldots, 5$.

Second, we generated a data set of $n = 50$ subjects according to the chosen NN LMM (8). This dataset contained no outliers, and the NN LMM should fit the simulated data well.

Third, we created a contaminated dataset with either *b*- or *e*-outliers corresponding to each uncontaminated dataset generated in the second step above. To introduce *b*-outliers into a dataset, we randomly selected 2%, 6%, or 10% subjects as outlying patients whose random effects b_{0i} were then replaced by values generated from a different normal distribution with 100 times larger variance, that is, $N(0, 10^2)$.

In parallel, to bring *e*-outliers into a dataset, we randomly selected 2%, 6%, and 10% subjects and then randomly selected a time point for each of these selected subjects, at which the original value of the error was replaced by a new one generated from $N(0, 5^2)$.

We considered two models in the comparison. One was the NN LMM, the model generating the uncontaminated data. Certainly, any outliers will challenge the performance of this model. The other was the TT LMM, which represents a class of models in which the degrees of freedom determine tail properties of the distribution and hence the flexibility of the model responds to outliers. Therefore the TT LMM is expected to accommodate both *b*- and *e*-outliers and to result in desirable robustness. We ran 200 rounds of simulation to draw conclusions.

To assess to what extent an outlier affects the performance of the proposed model, we utilized two criteria related to the performances on estimation of regression coefficients and estimation of standard errors, respectively. The first one was the standardized relative change (SRC) in the estimation for each regression coefficient (Song, 2007, page 113), defined by

$$\mathrm{SRC}(\beta_j) = \frac{\left|\hat{\beta}_{j,o}^{\mathrm{NN}} - \hat{\beta}_{j,n}^{\mathrm{NN}}\right|}{s.e.(\hat{\beta}_{j,n}^{\mathrm{NN}})} \Bigg/ \frac{\left|\hat{\beta}_{j,o}^{TT} - \hat{\beta}_{j,n}^{\mathrm{TT}}\right|}{s.e.(\hat{\beta}_{j,n}^{\mathrm{TT}})}, \quad j = 0, 1, 2.$$

The second one was the ratio of mean square error changes (RMSEC) between the contaminated and uncontaminated datasets, both of which were fit by the NN LMM and the TT LMM, respectively. That is,

$$\mathrm{RMSEC}(\beta_j) = \sqrt{\frac{\left|\mathrm{MSE}(\beta_{j,o}^{\mathrm{NN}}) - \mathrm{MSE}(\beta_{j,n}^{\mathrm{NN}})\right|}{\left|\mathrm{MSE}(\beta_{j,o}^{\mathrm{TT}}) - \mathrm{MSE}(\beta_{j,n}^{\mathrm{TT}})\right|}}, \quad j = 0, 1, 2,$$

where $\hat{\beta}_{j,o}$ and $\hat{\beta}_{j,n}$ are the MLEs of the fixed effect estimates based on the two datasets with and without outliers, respectively.

Obviously, either SRC > 1 or RMSEC > 1 implies that the TT LMM is more robust than the NN LMM in the estimation of regression coefficients or in the estimation of standard errors.

Table 1 reports a summary of the model performances in light of these two criteria, where we used median instead of mean to evaluate the average performance, simply to avoid some extreme values of the two criteria. Table 1 clearly indicates that (i) the MLE in the TT LMM over-performed that in the NN LMM, as all the values of the two criteria are above 1. In this simple simulation setting, the TT LMM may be up to five times more robust than the NN LMM. (ii) In our simulation setting, the *b*-outliers seem more troublesome to the MLE in the NN LMM than the *e*-outliers, because the values of the two criteria are generally higher for the cases related to the *b*-outliers than those related to the *e*-outliers. In fact, our findings here are in agreement with those from many other studies (e.g., Lange, Little & Taylor, 1989; Pinheiro, Liu & Wu, 2001; Song, Zhang & Qu, 2007); that is, the *t* distribution is helpful in achieving robust estimation

TABLE 1: Median standardized relative change (SRC) and median ratio of MSE changes (RMSEC) for the estimation of each fixed effect between contaminated and uncontaminated datasets fitted by the NN LMM and the TT LMM, based on 200 replications.

Contamination		β_0		β_1		β_2	
Percent	Source	SRC	RMSEC	SRC	RMSEC	SRC	RMSEC
2%	*b*-outlier	4.59	4.17	1.91	5.26	4.83	3.18
6%	*b*-outlier	1.92	1.89	2.12	1.59	2.42	1.64
10%	*b*-outlier	1.63	1.45	1.98	1.47	1.78	1.58
2%	*e*-outlier	3.06	3.08	2.46	3.69	2.64	2.07
6%	*e*-outlier	1.35	1.38	1.07	1.46	2.01	1.30
10%	*e*-outlier	1.28	1.44	1.03	1.59	1.88	1.27

and inference. (iii) It is interesting to note that the values of the criteria decrease as the percent of *b*-outliers increases. This implies that the capability of the TT LMM to handle the outliers was not constant, and, as a matter of fact, it deteriorates as the number of the outliers increases. This is not surprising. Like every other robust model the TT LMM has a breakdown point to its robustness.

To see more details as to how well the MLE performed in the two models, in Table 2 we listed a summary of average estimates of the fixed effects and their 95% quantile intervals. This table shows again that the MLE was affected more by the *b*-outliers than by the *e*-outliers, since wider quantile intervals and bigger bias are obtained in the *b*-outliers cases than in the *e*-outlier cases.

TABLE 2: Average point estimates of the fixed effects with 95% quantile intervals obtained from the NN LMM and the TT LMM, respectively, based on 200 replications.

Model	Outliers	β_0	β_1	β_2
NN	2%*b*	2.40 (1.96, 2.83)	−0.80 (−0.91, −0.71)	0.63 (0.02, 1.28)
TT	2%*b*	2.09 (1.62, 2.53)	−0.80 (−0.90, −0.72)	0.94 (0.36, 1.60)
NN	6%*b*	1.95 (1.32, 2.70)	−0.78 (−0.86, −0.70)	0.97 (−0.01,1.86)
TT	6%*b*	1.98 (1.43, 2.58)	−0.79 (−0.88, −0.71)	1.01 (0.15, 1.76)
NN	10%*b*	1.86 (0.96, 2.69)	−0.76 (−0.85, −0.66)	1.00 (−0.09, 2.09)
TT	10%*b*	1.91 (1.27, 2.62)	−0.78 (−0.87, −0.68)	1.03 (0.27, 1.78)
NN	2%*e*	1.96 (1.28, 2.57)	−0.79 (−0.91, −0.70)	1.03 (0.39, 1.81)
TT	2%*e*	1.98 (1.43, 2.53)	−0.80 (−0.89, −0.73)	1.03 (0.44, 1.78)
NN	2%*e*	2.01 (1.48, 2.57)	−0.80 (−0.93, −0.67)	0.99 (0.33, 1.62)
TT	6%*e*	2.02 (1.48, 2.54)	−0.80 (−0.91, −0.69)	1.01 (0.40, 1.55)
NN	10%*e*	2.00 (1.31, 2.63)	−0.79 (−0.93, −0.65)	0.97 (0.29, 1.77)
TT	10%*e*	2.01 (1.41, 2.59)	−0.80 (−0.92, −0.67)	1.00 (0.25, 1.72)

5. DATA ANALYSIS

Now we turn to the analysis of the eye surgery data introduced in Section 1. We first identified potential outliers and then used the proposed robust TT LMM to carry out estimation and inference. The data were previously analysed by Song & Tan (2000) and Song, Qiu & Tan (2004) using the GEE type marginal model, and by Qiu, Song & Tan (2008) using the generalized linear mixed-effects model (GLMM) with the utility of the simplex distribution. In the simplex GLMM, high-order Laplace approximations were invoked to evaluate integrals, resulting in an approximate inference in the forms of penalized quasi-likelihood (PQL) and restricted maximum likelihood (REML).

5.1. Outliers

To identify *b*-outliers, we began with individual logit-transformed linear regression models,

$$\text{logit}(y_{ij}) = \beta_0 + \beta_1 \log(t_{ij}) + v_i + \varepsilon_{ij}, \quad \text{for } i = 1, \ldots, n,\ j = 1, \ldots, n_i,$$

where $\beta_0 + v_i, i = 1, \ldots, n$ are individual intercepts. We applied the least squares method to estimate the regression coefficients; that is,

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{v}_i, i = 1, \ldots, n) = \arg \min_{\beta_0, \beta_1, v_i, i=1,\ldots,n} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \{\text{logit}(y_{ij}) - \beta_0 - \beta_1 \log(t_{ij}) - v_i\}^2.$$

The distribution of the resulting estimates $\hat{v}_i, i = 1, \ldots, n$ is helpful to identify outlying random effects or *b*-outliers.

Figure 3a displays a side-by-side boxplot of the standardized individual intercepts over three gas concentration levels. The outstanding point in the boxplot of gas level 15% corresponds to patient #16, who has been identified previously as an outlying patient. Two outliers appear in the boxplot of the gas level 20%, which were not identified previously, corresponding to patients #7 and #25. No outliers are found in the boxplot of the gas level 25%. In addition, the size of the box varies among the three gas groups, because the number of subjects in each group is different. In fact, this variation in the random effects has been investigated and found to be unrelated to the gas group level in Qiu (2001).

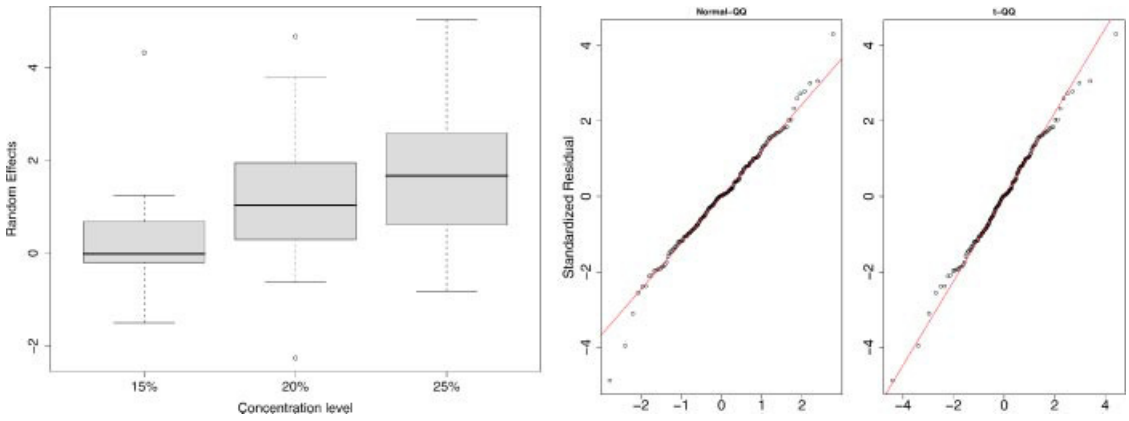

FIGURE 3: (a) Boxplots of random effects at three gas concentration levels. (b) Normal-quantile and *t*-quantile plots of the standardized residuals obtained from individual regression models.

To examine *e*-outliers, Figure 3b displays the normal-quantile plot (the left panel) and the *t*-quantile plot (the right panel) for the standardized residuals obtained from the above individual linear regression models. The quantiles in the right plot were calculated using the *t*-distribution with 5 degrees of freedom. This *t*-quantile plot indicates a clear improvement on the lower left tail to accommodate outliers. This suggests that using a *t* distribution for the error distribution can overcome such moderate violation of normality assumption. Here we did not identify any specific *e*-outliers, but just noted that the conclusion from the preliminary analysis has been in favour of a *t*-distribution for the error terms. We focused on the *b*-outliers in the data analysis.

5.2. Robust Analysis

We considered four LMM in the data analysis in order to assess the influences of the identified *b*-outliers on estimation and inference. Table 3 lists the model components needed to specify these LMMs, where the conditional median μ_{ij} takes the form of Equation (1).

We prepared three datasets in this analysis, including (i) the full dataset with all *b*-outliers, (ii) the sub-dataset with only patient #16 deleted, and (iii) the sub-dataset with all three *b*-outliers (#16, #7, and #25) deleted. We used subscripts *f*, *P*1, and *P*2 to denote the respective analyses of these three datasets. Similar to the investigation in the simulation study, in Table 4 we calculated the standardized relative change (RC) in the estimation of differential effects of two gas concentration levels (20% and 25%). The RC quantifies the robustness of an LMM with the utility of *t*-distribution in the random effects, in the error terms, or in both, versus the naive NN LMM. Clearly, the use of the *t*-distribution is beneficial to acquire consistent results and robust performances for the mixed-effects model in the data analysis.

TABLE 3: Model components in the specification of an LMM for the eye surgery data analysis.

	Distribution		
Model	Conditional $(y_{ij}\|b_{0i})$	Random Intercept (b_{0i})	Parameters $(\boldsymbol{\theta})$
NN LMM	$LN(\mu_{ij}, \sigma^2)$	$N(0, \tau^2)$	$\beta_0, \beta_1, \beta_2, \sigma^2, \tau^2$
TT LMM	$LT(d, \mu_{ij}, \sigma^2)$	$T(d, 0, \tau^2)$	$\beta_0, \beta_1, \beta_2, \sigma^2, \tau^2, d$
TN LMM	$LN(\mu_{ij}, \sigma^2)$	$T(d, 0, \tau^2)$	$\beta_0, \beta_1, \beta_2, \sigma^2, \tau^2, d$
NT LMM	$LT(d, \mu_{ij}, \sigma^2)$	$N(0, \tau^2)$	$\beta_0, \beta_1, \beta_2, \sigma^2, \tau^2, d$

TABLE 4: Standardized relative change (SRC) on the estimation of differential effects of the two higher gas concentration levels in the NT LMM, the TN LMM, and the TT LMM. The estimation from the NN LMM is the reference.

	Without Patient #16			Without Patients #16, #7, and #25		
Covariate	NT LMM	TN LMM	TT LMM	NT LMM	TN LMM	TT LMM
GAS20 (β_2)	1.22	2.49	5.07	1.50	1.78	9.69
GAS25 (β_3)	1.18	2.38	3.57	1.19	1.84	3.80

TABLE 5: The maximum likelihood estimates and standard errors of parameters in the linear mixed-effects models for the eye surgery data.

Model	β_0	β_1	β_2	β_3	β_4	β_5	σ^2	τ^2	d
NN_f	3.07	−1.27	0.68	1.22	−0.27	−0.25	2.01	2.30	—
	(0.63)	(0.19)	(0.77)	(0.81)	(0.23)	(0.23)	(0.23)	(0.68)	—
NN_{p1}	2.34	−1.22	1.31	1.85	−0.31	−0.30	1.98	1.87	—
	(0.62)	(0.21)	(0.74)	(0.78)	(0.24)	(0.25)	(0.23)	(0.58)	—
NN_{p2}	2.32	−1.21	1.29	1.85	−0.28	−0.31	1.91	1.13	—
	(0.51)	(0.20)	(0.62)	(0.64)	(0.25)	(0.24)	(0.23)	(0.40)	—
TN_f	2.70	−1.28	0.95	1.48	−0.26	−0.24	2.00	4.28	2.54
	(0.58)	(0.19)	(0.68)	(0.70)	(0.23)	(0.23)	(0.23)	(1.86)	(2.49)
TN_{p1}	2.30	−0.96	0.82	1.75	−0.68	−0.90	2.58	1.81	2.38
	(0.46)	(0.24)	(0.56)	(0.57)	(0.28)	(0.30)	(0.41)	(0.30)	(0.21)
TN_{p2}	2.47	−1.25	1.26	1.78	−0.37	−0.36	2.03	9.94	2.08
	(0.46)	(0.21)	(0.58)	(0.57)	(0.25)	(0.25)	(0.24)	(0.46)	(0.05)
NT_f	2.41	−0.97	0.76	1.20	−0.37	−0.59	2.86	1.78	2.87
	(0.54)	(0.22)	(0.67)	(0.70)	(0.25)	(0.26)	(1.22)	(0.55)	(0.93)
NT_{p1}	1.70	−0.85	1.21	1.65	−0.48	−0.71	2.79	1.52	2.86
	(0.54)	(0.19)	(0.66)	(0.69)	(0.23)	(0.23)	(1.13)	(0.47)	(0.88)
NT_{p2}	1.64	−0.82	1.11	1.65	−0.40	−0.76	3.06	0.98	2.62
	(0.45)	(0.18)	(0.56)	(0.58)	(0.23)	(0.22)	(1.44)	(0.34)	(0.76)
TT_f	3.44	−0.86	0.91	1.61	−0.47	−0.66	4.30	4.12	2.61
	(0.65)	(0.36)	(0.81)	(0.81)	(0.39)	(0.39)	(2.02)	(1.91)	(1.13)
TT_{p1}	2.87	−0.76	1.04	1.79	−0.54	−0.75	4.91	4.13	2.46
	(0.58)	(0.33)	(0.75)	(0.73)	(0.37)	(0.36)	(1.73)	(1.44)	(1.03)
TT_{p2}	2.98	−0.78	0.98	1.78	−0.49	−0.74	4.52	2.63	2.50
	(0.52)	(0.30)	(0.66)	(0.65)	(0.37)	(0.33)	(0.47)	(0.23)	(0.58)

Subscripts f, $P1$, $P2$ correspond to the full dataset, the sub-dataset without patient #16, and sub-dataset without patients #16, #7, and #25, respectively.

Table 5 lists the estimates and standard errors, where 50 quadrature points were used in Gauss–Hermite quadrature method to evaluate the integrals in the NT LMM and the TN LMM. All the estimated degrees of freedom are between 2 and 3, suggesting that both distributions for the random effects and the error terms are heavy-tailed. This confirmed the findings in the preliminary analysis above. Also, the goodness-of-fit of these four models can be simply compared by their values of the log-likelihood although they are not nested. For the case of the full data analysis, the TT LMM obtained the largest log-likelihood (−339.50), followed by the NT LMM (−343.18), then the TN LMM (−344.05) and the NN LMM (−350.70). The improvement of the TT LMM over the NN LMM is noticeable.

Time is a predominant covariate for the gas decay, and its significance had been little affected by the outliers. However, the significance of gas levels has been affected by the outliers, especially

for β_3, the differential effect between the two levels of 15% and 25% gas concentrations. The two models with the utility of t-distribution for the random effects b_{0i}, namely the TN LMM and the TT LMM, all indicated the significant differential effect in either the full data and partial data analyses. This conclusion agrees with the conclusion made from the NN LMM with the outliers deleted. In addition, by comparison the estimates of β_3 and the standard errors between the first sub-dataset ($P1$) and the second sub-dataset ($P2$), it seems that two outlying patients (#7 and #25) did not affect the estimates much but did affect the standard errors substantially, judged by the fact that a smaller dataset (i.e., $P2$-dataset) gave smaller standard errors. However, β_2 became significant in the NN, NT, and TN models after deleting #7 and #25, but it remained insignificant in the TT model. It further showed that the TT model is more robust than the other three models with respect to outliers.

The estimated marginal variances of the response are in the NN LMM $\hat{\sigma}^2 + \hat{\tau}^2 = 4.31$, and in the TT LMM $\hat{\sigma}^2 + \hat{\tau}^2 = 8.42$, respectively. As suggested in Figure 2, some caution may be needed for the use of the LN distribution to model the proportional data when the corresponding LT distribution has a large variance (8.42) and small degrees of freedom (2.5). Summarizing all the analyses above, we prefer the TT LMM as our final model to draw conclusions: time and the differential effect of gas concentration level 25% versus level 15% are statistically significant. This means patients who were assigned to the gas level 25% took longer time to absorb the gas to a given level than the patients in gas level 15% group. This finding is in agreement with that given in Qiu, Song & Tan (2008).

6. CONCLUDING REMARKS

We have proposed a robust transformation mixed-effects models for the analysis of longitudinal proportional data, based on a new logit-t distribution. It includes Aitchison's logit-normal as a special case. Robustness is a desirable property in both modelling and statistical inference when some outlying observations are present. Due to the hierarchical structure of a mixed-effect model, the t-distribution can be assumed for random effects, error terms or both to accommodate different types of outliers. Outliers in the mixed-effect model analysis of longitudinal data can be classified as b-outliers or e-outliers, according to if they come from random effects level or the conditional error terms. A preliminary analysis is an important step to identify potential outliers. This article has suggested four models, namely NN LMM, NT LMM, TN LMM and TT LMM, to form a primary platform to assess the influence of potential outliers and to determine from which source influential outliers arise.

The proposed robust model requires some analytical and numerical effort in the implementation of the MLE, compared to a straightforward application of the normal linear mixed model to fit logit-transformed data. However, the robust gain in the proposed model is substantial, where no extra model assumptions are required additional to those similarly given in the normal linear mixed model. As far as robustness is concerned, the proposed model is conceptually simpler and numerically easier than the simplex mixed-effects model introduced in Qiu, Song & Tan (2008). In particular, the TT LMM has a closed form in its likelihood function, and hence the related estimation and inference are as simple as the normal linear mixed model.

Since the numerical evaluation of integrals in the likelihood is required in the NT LMM and the TN LMM, the dimension of the random effects is restricted to be low, say 3 or less. This limitation of the quadrature numerical method can be alleviated by the application of Monte Carlo simulation method (e.g., McCulloch, 1997). If one is only interested in a robust analysis, then the TT LMM is recommended because the likelihood in the TT LMM has a closed analytic form and hence related computation is fast no matter what the dimension of the random effects. Of course, with no help from the other NT LMM and TN LMM, one may not be able to assess the influence of an outlier on goodness-of-fit.

APPENDIX: LIKELIHOOD FUNCTION AND FIRST ORDER DERIVATIVES

To ensure the positivity of parameters σ^2, τ^2 and $d-2$, let $\log(\sigma^2)=\xi$, $\log(\tau^2)=\zeta$, $\log(d-2)=\omega$. Then $\sigma^2=e^{\xi}$, $\tau^2=e^{\zeta}$, $d=e^{\omega}+2$.

For the NT LMM, the likelihood is evaluated under the following given distributions:

$$p(\mathbf{y}_i|\mathbf{b}_i)=\prod_{j=1}^{n_i}\frac{\Gamma\{(d+1)/2\}}{\sqrt{\pi(d-2)}\sigma\Gamma(d/2)y_{ij}(1-y_{ij})}\left\{1+\frac{(w_{ij}-X_{ij}'\boldsymbol{\beta}-Z_{ij}'\mathbf{b}_i)^2}{(d-2)\sigma^2}\right\}^{d+1/2}, \tag{9}$$

and

$$p(\mathbf{b}_i)=\frac{1}{(2\pi)^{q/2}|D(\boldsymbol{\tau})|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{b}_i'D(\boldsymbol{\tau})^{-1}\mathbf{b}_i\right\}. \tag{10}$$

It follows from (5) that $\ell_i(\boldsymbol{\theta})=\log\int p(\mathbf{y}_i|b_{0i})p(b_{0i})\mathrm{d}b_{0i}=\int p(\mathbf{y}_i,b_{0i})\mathrm{d}b_{0i}$. Thus,

$$\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\left\{\int\frac{\partial\log p(\mathbf{y}_i,b_{0i})}{\partial\boldsymbol{\theta}}p(\mathbf{y}_i|b_{0i})p(b_{0i})\mathrm{d}b_{0i}\right\}\left\{\int p(\mathbf{y}_i|b_{0i})p(b_{0i})\mathrm{d}b_{0i}\right\}^{-1},$$

where the integrals are evaluated by Gauss–Hermite quadrature method with 50 quadrature points. Let $w_{ij}=\mathrm{logit}(y_{ij})$, $j=1,\ldots,n_i;i=1,\ldots,n$. Then, $\mathbf{w}_i=(w_{i1},\ldots,w_{in_i})'$, $i=1,\ldots,n$. From (9) and (10), we obtain the log-likelihood function and the scores as follows:

$$\begin{aligned}\log p(\mathbf{y}_i,b_{0i})\propto{}& n_i\left\{\log\Gamma\left(\frac{d+1}{2}\right)-\frac{1}{2}\log(d-2)-\frac{1}{2}\log\sigma^2-\log\Gamma\left(\frac{d}{2}\right)\right\}\\&-\frac{d+1}{2}\sum_{j=1}^{n_i}\log\left\{1+\frac{(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})^2}{(d-2)\sigma^2}\right\}-\frac{1}{2}\log\tau^2-\frac{b_{0i}^2}{2\tau^2},\end{aligned}$$

$$\frac{\partial\log p(\mathbf{y}_i,b_{0i})}{\partial\boldsymbol{\beta}}=(d+1)\sum_{j=1}^{n_i}\frac{\mathbf{X}_{ij}(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})}{(d-2)\sigma^2+(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})^2},$$

$$\frac{\partial\log p(\mathbf{y}_i,b_{0i})}{\partial\xi}=-\frac{n_i}{2}+\frac{d+1}{2}\sum_{j=1}^{n_i}\frac{(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})^2}{(d-2)\sigma^2+(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})^2},$$

$$\frac{\partial\log p(\mathbf{y}_i,b_{0i})}{\partial\zeta}=\frac{1}{2\tau^2}(-\tau^2+b_{0i}^2),$$

$$\begin{aligned}\frac{\partial\log p(\mathbf{y}_i,b_{0i})}{\partial\omega}=\frac{d-2}{2}\Bigg[{}& n_i\left\{\psi_0\left(\frac{d+1}{2}\right)-\frac{1}{d-2}-\psi_0\left(\frac{d}{2}\right)\right\}-\sum_{j=1}^{n_i}\log\left\{1+\frac{(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})^2}{(d-2)\sigma^2}\right\}\\&+\frac{d+1}{d-2}\sum_{j=1}^{n_i}\frac{(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})^2}{(d-2)\sigma^2+(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})^2}\Bigg],\end{aligned}$$

where $\psi_0(\cdot)$ is the logarithmic derivative of the gamma function.

For the TN LMM, the distributions involved in the likelihood function are

$$p(\mathbf{y}_i|\mathbf{b}_i)=\prod_{j=1}^{n_i}\frac{1}{\sqrt{2\pi}\sigma y_{ij}(1-y_{ij})}\exp\left\{-\frac{(w_{ij}-X_{ij}'\boldsymbol{\beta}-Z_{ij}'\mathbf{b}_i)^2}{2\sigma^2}\right\}, \tag{11}$$

and

$$p(\mathbf{b}_i) = \frac{\Gamma((d+q)/2)}{\{\pi(d-2)\}^{q/2}\Gamma(d/2)}|D(\boldsymbol{\tau})|^{-1/2}\left\{1+\frac{\mathbf{b}_i' D(\boldsymbol{\tau})^{-1}\mathbf{b}_i}{d-2}\right\}^{-(d+q)/2}. \tag{12}$$

From (11) and (12), the log-likelihood and scores are given by

$$\begin{aligned}
\log p(\mathbf{Y}_i, b_i) \propto & -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{n_i}(y_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_i)^2 - \frac{d+1}{2}\log\left\{1+\frac{b_i^2}{(d-2)\tau^2}\right\} \\
& + \log\Gamma\left(\frac{d+1}{2}\right) - \frac{1}{2}\log(d-2) - \log\Gamma\left(\frac{d}{2}\right) \\
\frac{\partial \log p(\mathbf{y}_i, b_{0i})}{\partial \boldsymbol{\beta}} = & \frac{1}{\sigma^2}\sum_{j=1}^{n_i}\mathbf{X}_{ij}(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i}), \\
\frac{\partial \log p(\mathbf{y}_i, b_{0i})}{\partial \xi} = & \frac{1}{2\sigma^2}\left\{\sum_{j=1}^{n_i}(w_{ij}-\mathbf{X}_{ij}'\boldsymbol{\beta}-b_{0i})^2 - n_i\sigma^2\right\}, \\
\frac{\partial \log p(\mathbf{y}_i, b_{0i})}{\partial \zeta} = & \frac{1}{2}\frac{db_{0i}^2-(d-2)\tau^2}{(d-2)\tau^2+b_{0i}^2}, \\
\frac{\partial \log p(\mathbf{y}_i, b_{0i})}{\partial \omega} = & \frac{d-2}{2}\left[\left\{\psi_0\left(\frac{d+1}{2}\right)-\frac{1}{d-2}-\psi_0\left(\frac{d}{2}\right)\right\}\right. \\
& \left. - \log\left\{1+\frac{b_{0i}^2}{(d-2)\tau^2}+\frac{(d+1)b_{0i}^2}{(d-2)^2\tau^2+(d-2)b_{0i}^2}\right\}\right].
\end{aligned}$$

For the TT LMM, from (7) we obtain

$$\begin{aligned}
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = & \frac{d+n_i}{2}\frac{\mathbf{X}_i\mathbf{V}_i^{-1}(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})}{d-2+(\mathbf{y}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})} \\
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \xi} = & -\frac{\sigma^2}{2}\left\{\mathrm{tr}(\mathbf{V}_i^{-1}I_{n_i}) - \frac{(d+n_i)(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}\mathbf{V}_i^{-1}(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})}{d-2+(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i'\boldsymbol{\beta})}\right\}, \\
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \zeta} = & -\frac{\tau^2}{2}\left\{\mathrm{tr}(\mathbf{V}_i^{-1}J_{n_i}) - \frac{(d+n_i)(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}J_{n_i}\mathbf{V}_i^{-1}(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})}{d-2+(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})}\right\}, \\
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \omega} = & \frac{d-2}{2}\left[\psi_0\left(\frac{d+n_i}{2}\right) - \frac{n_i}{d-2} - \psi_0\left(\frac{d}{2}\right) - \log\left\{1+\frac{(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})}{d-2}\right\}\right. \\
& \left. + \frac{(d+n_i)(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})}{2\{(d-2)^2+(d-2)(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{w}_i-\mathbf{X}_i'\boldsymbol{\beta})\}}\right].
\end{aligned}$$

ACKNOWLEDGMENTS

This research is supported by the NSERC individual discovery grants of the first and third authors. The authors are grateful to the AE and two anonymous referees for their constructive comments. The authors thank Professor D. Wiens for his helpful suggestions; they led to an improved version of this manuscript.

BIBLIOGRAPHY

J. Aitchison, "*The Statistical Analysis of Compositional Data*," Chapman and Hall, London (1986).

K.-T. Fang, S. Kotz & K. W. Ng, "*Symmetric Multivariate and Related Distributions*," Chapman and Hall, London (1990).

W. A. Jensen & J. B. Birch (2009). Profile monitoring via nonlinear mixed models. *Journal of Quality Technology*, 41, 18–34.

B. Jørgensen, "*The Theory of Dispersion Models*," Chapman and Hall, London (1997).

S. Kotz & S. Nadarajah, "*Multivariate T Distributions and Their Applications*," Cambridge University Press, Cambridge (2004).

K. L. Lange, R. J. A. Little & J. M. G. Taylor (1989). Robust statistical modelling using *t*-distribution. *Journal of the American Statistical Association*, 84, 881–896.

C. E. McCulloch (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92, 162–170.

S. M. Meyers, J. S. Ambler, M. Tan, J. C. Werner & S. S. Huang (1992). Variation of perfluorpropane disappearance after vitrectomy. *Retina*, 12, 359–363.

J. C. Pinheiro, C. Liu & Y. N. Wu (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate *t* distribution. *Journal of Computational and Graphical Statistics*, 10, 249–276.

Z. Qiu (2001). *Simplex models for longitudinal proportional data*. Ph.D. Dissertation, Department of Mathematics and Statistics, York University, Toronto, Canada.

Z. Qiu, P. X.-K. Song & M. Tan (2008). Simplex mixed-effects models for longitudinal proportional data. *Scandinavian Journal of Statistics*, 35, 577–596.

D. Ruppert (2005). Discussion of maximization by parts in likelihood inference by Song, Fan, and Kalbfleish. *Journal of the American Statistical Association*, 100, 1161–1163.

P. X.-K. Song. "*Correlated Data Analysis: Modeling, Analytics, and Applications*," Springer, New York (2007).

P. X.-K. Song, Z. Qiu & M. Tan (2004). Modelling heterogeneous dispersion in marginal models for longitudinal proportional data. *Biometrical Journal*, 46, 540–553.

P. X.-K. Song & M. Tan (2000). Marginal models for longitudinal continuous proportional data. *Biometrics*, 56, 496–502.

P. X.-K Song, P. Zhang & A. Qu (2007). Maximum likelihood inference in robust linear mixed-effects models using multivariate *t* distributions. *Statistica Sinica*, 17, 929–943.

J. D. Williams, W. H. Woodall, J. B. Birch & J. H. Sullivan (2006). Distribution of Hotelling's T^2 statistic based on the successive differences estimator. *Journal of Quality Technology*, 38, 217–229.

Received 22 April 2008
Accepted 3 March 2009