A family of linear mixed-effects models using the generalized Laplace distribution

Marco Geraci[1*] Alessio Farcomeni[2]

[1]Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina
[2]Department of Economics and Finance, University of Rome "Tor Vergata"

Abstract

We propose a new family of linear mixed-effects models based on the generalized Laplace distribution. Special cases include the classical normal mixed-effects model, models with Laplace random effects and errors, and models where Laplace and normal variates interchange their roles as random effects and errors. By using a scale-mixture representation of the generalized Laplace, we develop a maximum likelihood estimation approach based on Gaussian quadrature. For model selection, we propose likelihood ratio testing and we account for the situation in which the null hypothesis is at the boundary of the parameter space. In a simulation study, we investigate the finite sample properties of our proposed estimator and compare its performance to other flexible linear mixed-effects specifications. In two real data examples, we demonstrate the flexibility of our proposed model to solve applied problems commonly encountered in clustered data analysis. The newly proposed methods discussed in this paper are implemented in the `R` package `nlmm`.
Keywords: best linear predictor; chi-bar squared; convolution; heterogeneity of treatment effects; longitudinal data; meta-analysis.

1 Introduction

The normal (or Gaussian) distribution historically has played a prominent role not only as limiting distribution of a number of sample statistics, but also for modeling data obtained in empirical studies. Its probability density is given by

$$f_N(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}, \tag{1.1}$$

for $-\infty < y < \infty$. The Laplace (or double exponential) distribution, like the normal, has a long history in statistics. However, despite being of potentially great value in applied research, it has never received the same attention. Its density is given by

$$f_L(y) = \frac{1}{\sqrt{2}\sigma} \exp\left\{-\sqrt{2}\left|\frac{y-\mu}{\sigma}\right|\right\}. \tag{1.2}$$

Throughout this paper, these distributions will be denoted by $\mathcal{N}(\mu, \sigma)$ and $\mathcal{L}(\mu, \sigma)$, respectively.

*Address for correspondence: Marco Geraci, Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, 915 Greene Street, Columbia, SC 29208, USA. E-mail: geraci@mailbox.sc.edu

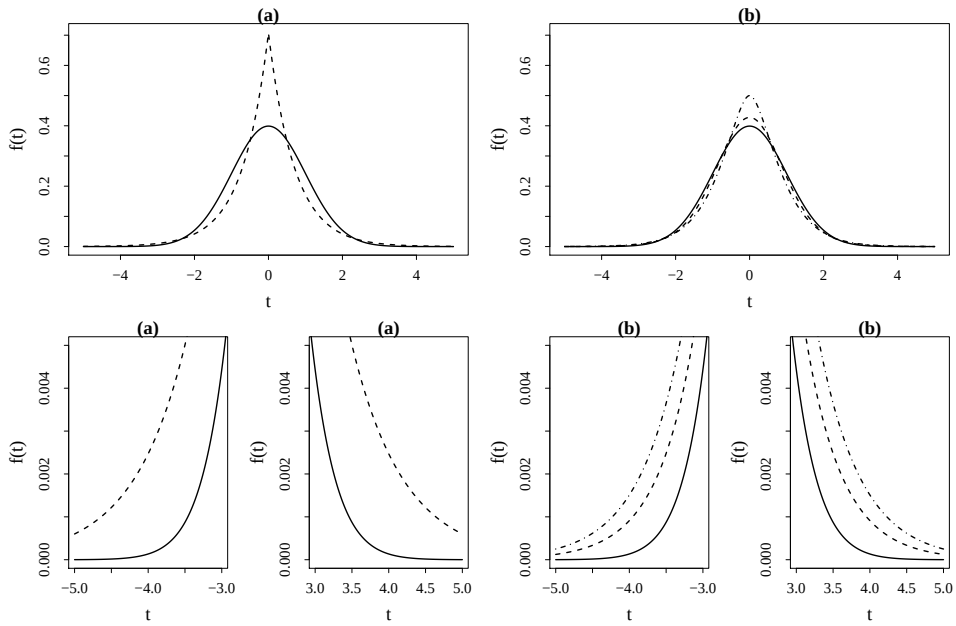

Figure 1: (a) Left: The normal (solid line) and double exponential (dashed line) densities. The location parameter is set to 0 and the variance is set to 1. (b) Right: The Normal-Normal (solid line), Normal-Laplace (dashed line), and Laplace-Laplace (dot-dash line) densities. The location parameter is set to 0 and the variance is set to 1.

In (1.1) and (1.2), μ and σ, where $-\infty < \mu < \infty$ and $\sigma > 0$, represent a location and a scale parameters, respectively. These two densities are shown in the left plot of Figure 1. The normal and Laplace distributions are both symmetric about μ and have variance equal to σ^2. As compared to the normal one, the Laplace density has a more pronounced peak (a characteristic technically defined *leptokurtosis*) and fatter tails. While visually the densities may seem to overlap on the tails, one can appreciate the difference between a normal and a Laplace distribution in terms of the cumulative probability: at $y = 3$, this is approximately 99.9% for a standard normal, but 99.3% for a Laplace with the same median and variance. Therefore the occurrence of an 'extremely' large value in a Laplace population is more than five times as likely as in a normal population.[1]

Interestingly, the Laplace distribution can be represented as a scale mixture of normal distributions. Let $Y \sim \mathcal{L}(\mu, \sigma)$, then[2]

$$Y \stackrel{d}{=} \mu + \sigma\sqrt{W}U, \tag{1.3}$$

where W and U are independent standard exponential and normal variables, respectively. That is, the Laplace distribution emerges from heterogeneous normal sub-populations. Both laws were proposed by Pierre-Simon Laplace: the double exponential in 1774 and the normal in 1778 (for an historical account, see Wilson[3]).

It is well known that, under the Gaussian error law (1.1), the maximum likelihood estimate of μ is the sample mean but, under the double exponential error law (1.2), it is the sample median. The former is the minimizer of the least squares (LS) estimator, while the latter is the minimizer of the least absolute deviations (LAD) estimator. In the past few years, theoretical developments related to least absolute error regression[4,5] have led to a renewed interest in the Laplace distribution and its asymmetric extension as pseudo-likelihood for quantile regression models of which median regression is a special case.[6–10] In parallel, computational advances based on interior point algorithms have

made LAD estimation a serious competitor of LS methods.[11,12] Another reason for the 'comeback' of the double exponential is related to its robustness properties which makes this distribution and distributions alike desirable in many applied research areas.[13]

In this paper we are interested in processes where the source of randomness can be attributed to more than one 'error' (a hierarchy of errors is also established). For instance, this is the case of longitudinal studies where part of the variation is attributed to an individual source of heterogeneity (called 'random effect'), say ε_1, independently from the noise, ε_2, where the distributions of ε_1 and ε_2 are often assumed to be symmetric about zero. The resulting model is a mixed-effects model (or mixed model for short). The distribution of the outcome can be obtained from the convolution of ε_1 and ε_2. Typically, the errors are assumed to be normal.[14,15] However, distributions other than normal have been considered by several authors to make the model more flexible.[16–18] Mixed models are widely applied in medical research. An important area of application is the study of patient-level heterogeneity in human genetics[19] and precision medicine.[20] In these contexts, it is natural to see each patient as a population in and of itself, with unique characteristics that, in response to a given treatment, determine individual variability. But this is tantamount to a population generated as in (1.3). Therefore, mixed models that are able to capture heterogeneity at the sub-population level have particular relevance in similar contexts.

Our contribution to the literature of mixed models is three-fold. First, we introduce a mixed-effects model based on the generalized Laplace distribution which includes the Gaussian and Laplace distributions as special cases. Our mixed models can be seen also as a full extension of Zhang and Davidian's[18] mixed models with flexible random effects distributions. Secondly, we discuss model choice within the newly introduced class. Third, we demonstrate the flexibility of our proposed models with different applications to clinical and biological data.

The rest of the paper is organized as follows. In Section 2, we introduce some notation and definitions. In Section 3, we consider generalized Laplace convolutions and discuss in detail four special cases. In Section 4, we discuss inference, including parameter and standard errors estimation, and model selection. In Section 5, we study finite-sample properties and comparative performance of our methods, while in Section 6 we report the results of two applications, namely an analysis of repeated measurements from a clinical trial on Crohn's disease patients, and a growth curve modeling study. We conclude with final remarks in Section 7. All the methods are implemented in the `R`[21] package `nlmm`, for which we give a brief tutorial in Supplemental Material.

2 Notation and definitions

We consider data from two-level nested cluster designs. Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^\top$ be a multivariate $n_i \times 1$ continuous random vector, and x_{ij} and z_{ij} be, respectively, $p \times 1$ and $q \times 1$ vectors of covariates for the jth observation, $j = 1, \ldots, n_i$, in cluster i, $i = 1, \ldots, M$. Also, let X_i and Z_i be, respectively, $n_i \times p$ and $n_i \times q$ design matrices for cluster i. The total sample size is denoted by $N = \sum_i^M n_i$. Observations in different clusters are assumed to be independent. The $n \times n$ identity matrix will be denoted by I_n. The n-variate normal distribution with location parameter μ and variance-covariance matrix Σ will be denoted by $\mathcal{N}_n(\mu, \Sigma)$, while the n-dimensional t-distribution with location parameter μ, scale matrix Σ and degrees of freedom δ will be denoted by $t_n(\mu, \Sigma, \delta)$ (we will make use of the multivariate t-distribution only in the simulation study).

We now introduce the generalized Laplace distribution[2] and its multivariate extension.[22] In particular, we present their symmetric formulations. The original models, which incorporate an additional parameter for asymmetry, are not relevant to the ensuing discussion as we focus our attention on symmetric distributions only.

Definition 1. *A random variable Y is said to follow a generalized (symmetric) Laplace (GL) distribution, $Y \sim \mathcal{GL}(\mu, \sigma, \alpha)$, if its density is given by*

$$f_{GL}(y) = \frac{1}{\sqrt{\frac{\pi}{2}}\Gamma(1/\alpha)\sigma^{1/\alpha+1/2}} \left(\frac{|y-\mu|}{\sqrt{2}}\right)^{\omega} B_{\omega}\left(\frac{\sqrt{2}|y-\mu|}{\sigma}\right), \tag{2.1}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, $0 < \alpha \leq 1$, $\omega = \frac{1}{\alpha} - \frac{1}{2}$ and B_u is the modified Bessel function of the third kind with index u.

Remark 1. *If $Y \sim \mathcal{GL}(\mu, \sigma, \alpha)$, then $\mathrm{E}(Y) = \mu$ and $\mathrm{var}(Y) = \frac{1}{\alpha}\sigma^2$.*

The GL distribution defined above admits the representation

$$Y \overset{d}{=} \mu + \sigma\sqrt{W}U \tag{2.2}$$

where W is standard gamma with shape $1/\alpha$, $W \sim \mathcal{G}(1/\alpha, 1)$, and $U \sim \mathcal{N}(0,1)$, independent from W.

It follows that if $W \sim \mathcal{G}(1,1)$, then (2.2) reduces to (1.3). Thus, for $\alpha = 1$ the GL distribution coincides with the Laplace distribution (1.2). Moreover, if $Y \sim \mathcal{GL}(\mu_\alpha, \sigma_\alpha, \alpha)$ where $\lim_{\alpha\to 0^+} \mu_\alpha/\alpha = \mu_0$ and $\lim_{\alpha\to 0^+} \sigma_\alpha^2/\alpha = \sigma_0^2$, then $Y \xrightarrow{d} \mathcal{N}(\mu_0, \sigma_0)$. The parameter α controls the behavior of the tails and kurtosis. Although the GL distribution is defined also for $\alpha > 1$, we restrict the space of α to the unit interval since the behavior of $f_{GL}(y)$ for $y \to 0^+$ is troublesome when $\alpha > 1$.[2]

The density (2.1) is a particular case of the generalized Laplace distribution.[2] The generalized (asymmetric) Laplace distribution has been extended to the multivariate case.[22] Again, here we consider its symmetric formulation.

Definition 2. *An n-dimensional random variable $Y = (Y_1, Y_2, \ldots, Y_n)^\top$ is said to follow a multivariate generalized (symmetric) Laplace distribution if it has density*

$$f_{GL}(y) = \frac{2}{(2\pi)^{n/2}\Gamma(1/\alpha)|\Sigma|^{1/2}} \left(\frac{Q(y,\mu)}{\sqrt{2}}\right)^{\omega} B_{\omega}\left(\sqrt{2}Q(y,\mu)\right), \tag{2.3}$$

where $\mu \in \mathbb{R}^n$, Σ is a positive-definite $n \times n$ matrix, $0 < \alpha \leq 1$, $\omega = \frac{1}{\alpha} - \frac{n}{2}$, and $Q(y,\mu) = \sqrt{(y-\mu)^\top\Sigma^{-1}(y-\mu)}$. This distribution is denoted by $\mathcal{GL}_n(\mu, \Sigma, \alpha)$.

Remark 2. *If $Y \sim \mathcal{GL}_n(\mu, \Sigma, \alpha)$, then $\mathrm{E}(Y) = \mu$ and $\mathrm{var}(Y) = \frac{1}{\alpha}\Sigma$.[22] For a diagonal matrix $\Sigma = \mathrm{diag}(\varsigma_1, \ldots, \varsigma_q)$, the coordinates of the multivariate GL are uncorrelated, but not necessarily independent.*

As in the univariate case, we restrict the space of α to the unit interval. The multivariate GL has a number of properties that are desirable in our modeling framework. First of all, if $n = 1$ we obtain the one-dimensional GL introduced in (2.1). Secondly, a linear combination of the coordinates of the multivariate GL is still GL.[22] Namely, if $Y \sim \mathcal{GL}_n(0, \Sigma, \alpha)$ and z is an $n \times 1$ real vector with at least one non-zero element, then $z^\top Y \sim \mathcal{GL}(0, \sigma, \alpha)$ where $\sigma = z^\top\Sigma z$. Finally, the multivariate GL distribution admits the representation

$$Y \overset{d}{=} \mu + \sqrt{W}U, \tag{2.4}$$

where $W \sim \mathcal{G}(1/\alpha, 1)$ and $U \sim \mathcal{N}_n(0, \Sigma)$.

It follows that the multivariate GL converges to a multivariate normal when $\alpha \to 0^+$. To see this, express (2.4) as $Y \stackrel{d}{=} \mu + U_0 C_\alpha \sqrt{W}$, where $U_0 \sim MVN_n(0, I_n)$, $C_\alpha/\sqrt{\alpha} \to C_0$, and $C_0^\top C_0 = \Sigma_0$ is a positive-definite matrix. Then it is straightforward to verify that $C_\alpha\sqrt{W}$ converges in probability to C_0, and therefore Y converges in distribution to $\mathcal{N}_n(\mu, \Sigma_0)$. A proof in the univariate case can be found in the book by Kotz and colleagues.[2]

Also, $\mathcal{GL}_n(\mu, \Sigma, 1) \stackrel{d}{=} \mathcal{L}_n(\mu, \Sigma)$. The latter denotes the n-variate Laplace distribution and is a multivariate generalization of (1.2). The multivariate Laplace distribution is defined as follows.

Definition 3. *An n-dimensional random variable $Y = (Y_1, Y_2, \ldots, Y_n)^\top$ is said to follow a multivariate Laplace distribution, $Y \sim \mathcal{L}_n(\mu, \Sigma)$, if its density is given by*

$$f_L(y) = K \cdot \left\{ \frac{(y-\mu)^\top \Sigma^{-1}(y-\mu)}{2} \right\}^{\omega/2} B_\omega \left\{ \sqrt{2(y-\mu)^\top \Sigma^{-1}(y-\mu)} \right\}, \tag{2.5}$$

where $K = \dfrac{2}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}$, $\mu \in \mathbb{R}^n$, Σ is an $n \times n$ positive-definite symmetric matrix, and $\omega = (2-n)/2$.

Remark 3. *If $Y \sim \mathcal{L}_n(0, \Sigma)$, then $\mathrm{var}(Y) = \Sigma$.[2] For a diagonal matrix $\Sigma = \mathrm{diag}(\varsigma_1, \ldots, \varsigma_q)$, the coordinates of the multivariate Laplace are uncorrelated, but not independent. Therefore, the joint distribution of n independent univariate Laplace variates does not have the properties of the multivariate Laplace with diagonal variance-covariance matrix.*

For $n = 1$, the multivariate Laplace density defined in (2.5) reduces to the univariate density (1.2) with $\sigma = \Sigma^{1/2}$. Moreover, as in the case of a multivariate GL, a linear combination of the coordinates of the multivariate Laplace is still a Laplace.[2] Indeed, if we assume $Y \sim \mathcal{L}_q(0, \Sigma)$ and z is an $n \times 1$ real vector, then $z^\top Y \sim \mathcal{L}(0, \sigma)$, where $\sigma = \sqrt{z^\top \Sigma z}$.

3 Generalized Laplace mixed-effects models

3.1 The general model

We first consider a general model where ε_1 and ε_2 have densities as in (2.3). Subsequently, we consider four special cases which include the multivariate normal and Laplace distributions. The subscripts 1 and 2 indicate, respectively, which of the two random variables plays the role of a random effect and which one is considered to be the noise. Here, the former may in general be associated with a vector of covariates and may represent an inferential quantity of interest; the latter is treated as a nuisance.

We define the following generalized Laplace mixed-effects (GLME) model

$$Y_i = X_i\beta + Z_i\varepsilon_{1i} + \varepsilon_{2i}, \tag{3.1}$$

where $\varepsilon_{1i} \sim \mathcal{GL}_q(0, \Sigma_1, \alpha_1)$ and $\varepsilon_{2i} \sim \mathcal{GL}_{n_i}(0, \Sigma_{2i}, \alpha_2)$, for $0 < \alpha_a \leq 1$, $a = 1, 2$, with $\varepsilon_{1i} \perp \varepsilon_{2i}$. The location parameter is modeled as a linear function of a $p \times 1$ dimensional vector of regression coefficients, $\beta \in \mathbb{R}^p$. We assume that the $q \times 1$ vector of random effects ε_{1i} has $q \times q$ variance-covariance matrix $\Psi_1 = \frac{1}{\alpha_1}\Sigma_1$ for all $i = 1, \ldots, M$. The errors ε_{2i} are assumed to have variance-covariance $\Psi_{2i} = \frac{1}{\alpha_2}\Sigma_{2i}$, $i = 1, \ldots, M$, which allows for heteroscedasticity (but not for residual correlation). Possible models for Σ_1 and Σ_{2i} are discussed in Section 4.3.

To our knowledge, the use of the generalized Laplace distribution in mixed-effects models is novel and there are no proposals comparable to ours. Yet, it is worth briefly discussing the proposal by

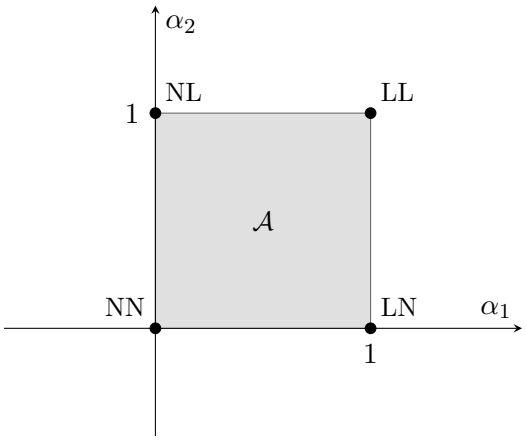

Figure 2: Model space $\mathcal{A}$ with Normal-Normal (NN), Normal-Laplace (NL), Laplace-Normal (LN), and Laplace-Laplace convolutions at the boundary of the space.

Yavuz and Arslan,[23] brought up to our attention by a referee as possibly related to our GLME model. Here, we point out fundamental differences showing that this is not the case. In their model, called Laplace linear mixed model, Yavuz and Arslan considered a multivariate formulation of the power exponential (PE) distribution[24,25] whose density is given by

$$f_{PE}(y) = \frac{n\Gamma(n/2)}{\pi^{n/2}2^{1+\frac{n}{2\gamma}}\Gamma(1+\frac{n}{2\gamma})|\Lambda|^{1/2}} \exp\left\{-\frac{1}{2}\left(\sqrt{(y-\mu)^\top\Lambda^{-1}(y-\mu)}\right)^\gamma\right\},$$

where Λ is a positive-definite variance-covariance matrix and $\gamma > 0$ is a parameter related to kurtosis. In particular, Yavuz and Arslan set γ equal to $1/2$ to obtain what they called "multivariate Laplace". However, the PE distribution defined above with $\gamma = 1/2$ is not related to the multivariate Laplace distribution (2.5).[26]

The reason that probably led Yavuz and Arslan to give the Laplace appellative to their model resides in the connection between the multivariate PE and GL distributions that was reported by Arslan.[27] By setting $\alpha \equiv \frac{2}{n+1}$ and $\Sigma \equiv 8\Lambda$, the density f_{GL} in (2.3) reduces to the density f_{PE} above with $\gamma = 1/2$. Unfortunately, in this formulation α is no longer a free parameter, while, worryingly, the shape of the distribution is dictated by the dimension of the support. Another drawback of Yavuz and Arslan's approach is that random effects and errors in each cluster are constrained to have a common shape (see their eq. 5), which is equivalent to setting the shape parameter in the GL distribution equal to $\frac{2}{n_i+q+1} \leq \frac{2}{3}$, jointly for (Y_i, ε_{1i}). In conclusion, not only can Yavuz and Arslan's model never be Laplace, but it is also a heterogeneous mix of cluster-specific power exponentials, each gradually morphing into a normal as soon as $n_i + q$ increases.

In contrast, the parameter α in our GLME model is unknown and estimated from the data (although it can be constrained if so desired). This entails a continuum of distributions ranging from the normal to the Laplace, independently for the random effects and the error. We illustrate this graphically in Figure 2, which depicts $\mathcal{A} = (0, 1] \times (0, 1]$, the parameter space of $\alpha = (\alpha_1, \alpha_2)^\top$. Each point in $\mathcal{A}$ represents a model in the GLME family. In particular, note that the corners on the boundary of $\mathcal{A}$ represent four special cases of model (3.1), which we discuss in the next section.

3.2 Special cases

Four special cases arise from (3.1) and these are schematically shown in Figure 2. For ease of presentation, the letters ν and λ are used to denote normal and Laplace variates, respectively.

With a slight abuse of notation, we will write $\alpha_1 = 0^+$ or $\alpha_2 = 0^+$ whenever $\alpha_1 \to 0^+$ or $\alpha_2 \to 0^+$, respectively.

3.2.1 Normal-Normal (NN)

The first special case of (3.1) for $\alpha = (0^+, 0^+)^\top$ is the Normal-Normal convolution (NN)

$$Y_i = X_i\beta + Z_i\nu_{1i} + \nu_{2i}, \tag{3.2}$$

where $\nu_{1i} \sim \mathcal{N}_q(0, \Sigma_1)$ and $\nu_{2i} \sim \mathcal{N}_{n_i}(0, \Sigma_{2i})$. Model (3.2) is known as a linear mixed effects (LME) model.[14,15] There is a vast number of applications of LME models, especially for the analysis of clustered data in the social, life and physical sciences.

3.2.2 Normal-Laplace (NL)

The second special case of (3.1) for $\alpha = (0^+, 1)^\top$ consists of a normal and a Laplace components, that is

$$Y_i = X_i\beta + Z_i\nu_{1i} + \lambda_{2i}, \tag{3.3}$$

where $\nu_{1i} \sim \mathcal{N}_q(0, \Sigma_1)$ and $\lambda_{2i} \sim \mathcal{L}_{n_i}(0, \Sigma_{2i})$. The Normal-Laplace (NL) convolution[28] arises from a Brownian motion whose starting value is normally distributed and whose stopping hazard rate is constant. An extension of the NL model to skewed forms can be obtained by letting λ_{2i} follow an asymmetric Laplace distribution.[28] Applications of the NL convolution can be found in finance.[29,30] See also the double Pareto-lognormal distribution, associated with $\exp(Y)$, which has applications in modeling size distributions.[31]

Model (3.3) is a median regression model with normal random effects, a special case of the linear quantile mixed models (LQMMs).[7,32] LQMMs have been used in a wide range of research areas, including marine biology,[33–35] environmental science,[36] cardiovascular disease,[37,38] physical activity,[39,40] and ophthalmology.[41,42]

3.2.3 Laplace-Normal (LN)

The third special case of (3.1) for $\alpha = (1, 0^+)^\top$ is the Laplace-Normal (LN) convolution

$$Y_i = X_i\beta + Z_i\lambda_{1i} + \nu_{2i}, \tag{3.4}$$

where $\lambda_{1i} \sim \mathcal{L}_q(0, \Sigma_1)$ and $\nu_{2i} \sim \mathcal{N}_{n_i}(0, \Sigma_{2i})$. As compared to the NL convolution, the Laplace component in the LN convolution is associated with the random effects, not with the error term. The LN model appears in robust meta-analysis.[15]

3.2.4 Laplace-Laplace (LL)

The fourth and last convolution, which represents the special case of (3.1) for $\alpha = (1, 1)^\top$, consists of two Laplace variates, i.e.

$$Y_i = X_i\beta + Z_i\lambda_{1i} + \lambda_{2i}, \tag{3.5}$$

where $\lambda_{1i} \sim \mathcal{L}_q(0, \Sigma_1)$ and $\lambda_{2i} \sim \mathcal{L}_{n_i}(0, \Sigma_{2i})$. The Laplace-Laplace (LL) model (3.5) is a median regression model with 'robust' random effects, another special case of LQMMs.[32]

3.3 Some properties

The marginal densities of the convolutions (3.2), (3.3), (3.3), and (3.5) are symmetric, unimodal, twice differentiable and have continuous first and second derivatives (the NN and NL are also smooth). Also, they are log-concave since both the normal (1.1) and Laplace (1.2) densities are log-concave.[43] The analytic expressions of the marginal densities are given elsewhere[44] and are shown in the right plot of Figure 1. As compared to the NN density, the NL (LN) and LL densities are leptokurtic and have more weight in the tails, with the NL density sitting between the NN and LL distributions.

4 Inference

In this section, we introduce inferential methods for GLME models, all of which are implemented in the `R` package `nlmm`.[45] A brief tutorial on `nlmm` is given in Supplementary Material.

4.1 Estimation

There are several approaches to mixed effects model estimation,[14,15] each approach having its own advantages and disadvantages. One approach is to work with the marginal likelihood of Y_i. Starting from the scale-mixture representation of the GL distribution (2.4), the model's likelihood takes the form of a multivariate normal likelihood weighted by a gamma distribution. Indeed, model (3.1) can be represented as a convolution of two mixtures, namely as

$$Y_i = X_i\beta + \sqrt{W_{1i}}U_{1i} + \sqrt{W_{2i}}U_{2i}, \tag{4.1}$$

where the independent coordinates of the vector $W_i = (W_{1i}, W_{2i})^\top$ each have a standard gamma distribution with shape $1/\alpha_a$, $a = 1, 2$, $U_{1i} \sim \mathcal{N}_{n_i}(0, Z_i\Sigma_1 Z_i^\top)$, and $U_{2i} \sim \mathcal{N}_{n_i}(0, \Sigma_{2i})$. We note that conditional on $W_i = (w_{1i}, w_{2i})^\top$, Y_i is normal, i.e. $Y_i|W_i \sim \mathcal{N}_{n_i}(X_i\beta, \Omega_i)$, where $\Omega_i = w_{1i}Z_i\Sigma_1 Z_i^\top + w_{2i}\Sigma_{2i}$. Based on (4.1), the distribution of Y can be found as the marginal distribution $g(y) = \int g(y|w) \cdot h(w)\,\mathrm{d}w = \mathrm{E}_w\{g(y|w)\}$. The latter integral does not seem to have an immediate closed-form solution, except for $\alpha = (0^+, 0^+)^\top$.

To estimate the parameters of interest, we resort to numerical integration. Alternatively, one can use an expectation-maximization (EM) algorithm.[46] The disadvantage of quadrature methods is that they are notoriously inefficient if the dimension of the integral is large. However, the EM often results in an intractable form of the expectation step. As discussed by Geraci[44] for the four special cases in Section 3.2, the E-step can be approximated via Monte Carlo sampling, whereby the unobservable variable w is sampled from the conditional density $g(w|y)$. While the computational efficiency of Monte Carlo EM (MCEM) estimation does not depend as much on dimensionality as quadrature methods do, convergence can be slower for MCEM than for quadrature-based methods.[46] Indeed, in a simulation study with NL models and moderately-sized datasets,[47] computational times were reported in the region of 1 second using numerical integration, against a whopping 42 minutes for a (supposedly fast) stochastic approximation of the EM algorithm.[48]

Before we proceed, we need to re-parameterize (4.1). Let $\Sigma_1 = \sigma^2\tilde{\Sigma}_1$ and $\Sigma_{2i} = \sigma^2\tilde{\Sigma}_{2i}$. Since $\tilde{\Sigma}_{2i}$ is positive-definite, it admits an invertible square root, say $\Lambda_i \equiv \tilde{\Sigma}_{2i}^{-1/2}$. Let $\tilde{Y}_i = \Lambda_i^\top Y_i$, $\tilde{X}_i = \Lambda_i^\top X_i$, and $\tilde{Z}_i = \Lambda_i^\top Z_i$. Then

$$\tilde{Y}_i = \tilde{X}_i\beta + \sqrt{W_{1i}}\tilde{U}_{1i} + \sqrt{W_{2i}}\tilde{U}_{2i}, \tag{4.2}$$

where $\tilde{U}_{1i} \sim \mathcal{N}_{n_i}(0, \sigma^2\tilde{Z}_i\tilde{\Sigma}_1\tilde{Z}_i^\top)$ and $\tilde{U}_{2i} \sim \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i})$. For estimation purposes, the relative precision matrix $\tilde{\Sigma}_1^{-1} = \sigma^2\Sigma_1^{-1}$ is parameterized in terms of an unrestricted r-dimensional vector,

$1 \leq r \leq q(q+1)/2$, of non-redundant parameters $\xi_1 \in \mathbb{R}^r$. The parameter ξ_1 is defined to be the vector of non-zero elements of the upper triangle of the matrix logarithm of C, where C is the $q \times q$ matrix obtained from the Cholesky decomposition $\tilde{\Sigma}_1^{-1} = C^\top C$.[14] Let $\xi_2 \in \mathbb{R}^s$ be the unrestricted s-dimensional vector of parameters in the error variance function (see Section 4.3 for more details). Moreover, we define $\phi = \log(\sigma)$, and $\tau_a = \log\{\alpha_a/(1-\alpha_a)\}$, $a = 1, 2$. The parameter to be estimated is then $\theta = \left(\beta^\top, \xi_1^\top, \xi_2^\top, \phi, \tau_1, \tau_2\right)^\top \in \mathbb{R}^d$, where $d = p + r + s + 3$.

Let $w_{\mathbf{k}} = (w_{k_1}, w_{k_2})$ and $\pi_{\mathbf{k}} = (\pi_{k_1}, \pi_{k_2})$ be, respectively, nodes and weights of the 2-dimensional Gaussian quadrature rule resulting from two standard gamma distributions with shape parameters $1/\alpha_1$ and $1/\alpha_2$. Each element of the 2-dimensional index $\mathbf{k} = (k_1, k_2)$ runs from 1 to K, where K is the number of integration points for a 1-dimensional rule. This entails a total of K^2 integration points. The log-integrated likelihood for model (4.1) is then given by

$$\ell_{GQ}(\theta) = \sum_{i=1}^{M} \log|\Lambda_i| + \sum_{i=1}^{M} \log \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \frac{|\Omega_{i\mathbf{k}}|^{-1/2}}{(2\pi\sigma^2)^{n_i/2}} \exp\left\{-\frac{1}{2\sigma^2}(\tilde{y}_i - \tilde{X}_i\beta)^\top \Omega_{i\mathbf{k}}^{-1}(\tilde{y}_i - \tilde{X}_i\beta)\right\} \pi_{k_1}\pi_{k_2}, \tag{4.3}$$

where $\Omega_{i\mathbf{k}} = w_{k_1}\tilde{Z}_i\tilde{\Sigma}_1\tilde{Z}_i^\top + w_{k_2}I_{n_i}$. Our maximum likelihood (ML) estimator is given by

$$\hat{\theta} = \arg\max_{\theta} \ell_{GQ}(\theta). \tag{4.4}$$

Estimation is carried out iteratively. At iteration t, the estimate $\hat{\theta}^{(t)}$ is updated according to a general-purpose optimization algorithm (e.g., Nelder-Mead). The starting values $\hat{\beta}^{(0)}$, $\hat{\xi}_1^{(0)}$, $\hat{\xi}_2^{(0)}$, and $\hat{\phi}^{(0)}$ can be obtained from analogous LME models. In preliminary numerical studies (not shown), we experimented different starting values $\hat{\tau}_1^{(0)}$ and $\hat{\tau}_2^{(0)}$ and observed sensitivity of the final estimates of θ. Therefore, we recommend using a multi-start strategy with starting values $\hat{\tau}_a^{(0)} \in \{-6.9067, 0, 6.9067\}$ (which correspond to the logit of 0.001, 0.5, and 0.999), for each $a = 1, 2$, giving a grid with 9 possible combinations. These starting values are used by default in the `nlmmControl` function of the `nlmm` package. However, when the number of clusters is small, or the number of parameters is large, or both, it is advisable to use a fine sequence of additional starting values, which includes those used by default. Increasing the number of integration points should also be considered. Finally, we point out that the `nlmm` package implements a Cholesky decomposition to obtain the square-root inverse of $\Omega_{i\mathbf{k}}$. Thus, poor scaling of the input may cause numerical instability.

The variance-covariance of $\hat{\theta}$ can be obtained by first calculating a numerical approximation to the Hessian matrix of (4.3) (evaluated at the parameter's estimate) and then calculating its Moore-Penrose generalized inverse. Estimates of the random effects for the ith cluster can be recovered by using the best linear predictor[32]

$$Z_i\hat{\varepsilon}_{1i} = \mathrm{E}(Z_i\varepsilon_{1i}) + \{\mathrm{cov}(Z_i\varepsilon_{1i}, y_i)\}(Z_i\Psi_1 Z_i^\top + \Psi_{2i})^{-1}\{y_i - \mathrm{E}(y_i)\}. \tag{4.5}$$

The first term of (4.5) is zero and, by model's assumption, $\mathrm{cov}(Z_i\varepsilon_1, y_i) = Z_i\Psi_1 Z_i^\top$. Hence, by plugging in the ML estimates of θ in (4.5), we obtain the estimated best linear predictions

$$\hat{\varepsilon}_{1i} = \frac{1}{\hat{\alpha}_1}\hat{\Sigma}_1 Z_i^\top \left(\frac{1}{\hat{\alpha}_1} Z_i\hat{\Sigma}_1 Z_i^\top + \frac{1}{\hat{\alpha}_2}\hat{\Sigma}_{2i}\right)^{-1} \left(y_i - X_i\hat{\beta}\right), \quad i = 1, \ldots, M. \tag{4.6}$$

4.2 Likelihood ratio testing for model selection

In Section 3.2, we have seen that model (3.1) takes on different special forms involving the normal and Laplace distributions if we constrain the parameters α_1 and α_2 to particular values. In Table 1,

we defined eight particular scenarios where either or both α-parameters are constrained to 0 or 1 under specific null hypotheses of interest. For example, the first hypothesis corresponds to testing whether an LME model fits the data sufficiently well, against the alternative hypothesis that a generalized Laplace model is required. Of course, the list in Table 1 is not exhaustive. For instance, one might want to test $H_0 : \alpha_1 = 0, \alpha_2 = 0.5$. More in general, we might be interested in testing the null hypothesis that α_1 or α_2 or both are equal to some real constants between 0 and 1. This effectively is equivalent to testing between models with different random-effects and error distributions. We consider the likelihood ratio test (LRT) statistic

$$D = -2\{l(\hat{\theta}_{\alpha_0}) - l(\hat{\theta})\}, \tag{4.7}$$

where $\hat{\theta}_{\alpha_0}$ is the ML estimate of θ from a model with constraints on one or both α-parameters, while $\hat{\theta}$ is the ML estimate from the unconstrained model. Let θ_0 denote the true value of θ conditionally on H_0. If the constraints on the α-parameters are in the interior of the unit interval, the null distribution of (4.7) is a chi-squared distribution. However, the null hypotheses in Table 1 place the parameter on the boundary of the parameter space, and therefore the LRT is not regular.[49]

Table 1: Special cases of null and alternative hypotheses for the shape parameters α_1 and α_2 and corresponding null models.

Null model	H_0	H_1
(1) Normal-Normal	$\alpha_1 = \alpha_2 = 0$	$0 < \alpha_1 \leq 1$ and $0 < \alpha_2 \leq 1$
(2) Normal-Laplace	$\alpha_1 = 1$ and $\alpha_2 = 0$	$0 < \alpha_1 < 1$ and $0 < \alpha_2 \leq 1$
(3) Laplace-Normal	$\alpha_1 = 0$ and $\alpha_2 = 1$	$0 < \alpha_1 \leq 1$ and $0 \leq \alpha_2 < 1$
(4) Laplace-Laplace	$\alpha_1 = \alpha_2 = 1$	$0 \leq \alpha_1 < 1$ and $0 \leq \alpha_2 < 1$
(5) Normal random effects	$\alpha_1 = 0$	$0 < \alpha_1 \leq 1$
(6) Normal errors	$\alpha_2 = 0$	$0 < \alpha_2 \leq 1$
(7) Laplace random effects	$\alpha_1 = 1$	$0 \leq \alpha_1 < 1$
(8) Laplace errors	$\alpha_2 = 1$	$0 \leq \alpha_2 < 1$

We reparameterize θ as $\tilde{\theta}$ so that the hypotheses can be expressed as $H_0 : G\tilde{\theta} = 0$ vs $H_1 : G\tilde{\theta} > 0$, where G has g rows, corresponding to the number of constrained shape parameters under H_0. If we test for a normal model, then $\tilde{\theta} = \theta$. Let $J(\tilde{\theta})$ denote the average information matrix as a function of $\tilde{\theta}$. Also, let $w = (w_0, \ldots, w_g)^\top$ be a $(g+1)$-dimensional vector of weights that depend on the positive-definite matrix $GJ(\tilde{\theta}_0)^{-1}G^\top$. Such weights can be computed as outlined elsewhere.[49,50] Then, D is distributed according to the following $\bar{\chi}^2$-type law[49]

$$\Pr(D \leq d) = \sum_{j=0}^{g} w_j \Pr\left(\chi_j^2 \leq d\right), \tag{4.8}$$

where χ_j^2 is chi-squared with j degrees of freedom. Expression (4.8) can be used to compute p-values based on (4.7), with the log-likelihood approximated by (4.3). The information matrix can be evaluated at the (consistent) constrained ML estimate via numerical differentiation of the log-integrated likelihood.

Finally, we note that a correction for multiplicity is warranted when using the same data to evaluate several hypotheses corresponding to different null models (e.g., NN and NL).[51]

4.3 Modeling of variance-covariance

In this section, we briefly introduce some alternative models for Σ_1 and Σ_{2i}. The reader is referred to the book by Pinheiro and Bates[14] for more details.

There are different possible structures for Σ_1. The simplest is a multiple of the identity matrix, with constant diagonal elements and zero off-diagonal elements. Other structures include, for example, diagonal (variance components), compound symmetric (constant diagonal and constant off-diagonal elements), and the more general symmetric positive-definite matrix. These are all available in the `nlmm`,[45] `nlme`,[52] `lme4`,[53] and `lqmm`[54] packages, as well as in SAS procedures for mixed effects models.

To model the residual heteroscedasticity Σ_{2i}, we consider variance structures of the type[14]

$$\text{var}(\varepsilon_{2ij}|\varepsilon_{1i}) = \sigma^2 h^2(v_{ij}, \xi_2), \tag{4.9}$$

where v_{ij} is a vector of variance covariates, $\xi_2 \in \mathbb{R}^s$ is a vector of unrestricted variance parameters, and $h(\cdot)$ is the variance function, assumed to be continuous in ξ_2. For example, the `nlmm` and `nlme` packages support several models for h including exponential, power, and fixed-weights variance functions.

5 Simulation study

We carried out a simulation study organized into two parts, (a) and (b). In part (a), we assessed: bias and mean squared error (MSE) of the GLME estimator; coverage of 95% confidence intervals for $\hat{\beta}$; and rejections rates of the LRT in Section 4.2. The goal was to evaluate finite-sample properties of the proposed estimator and related tests. In part (b), we compared bias, MSE and coverage for the GLME estimator with those for corresponding estimators from analogous standard and heavy-tailed[16] LME models. The goal was to understand the advantages and disadvantages of our proposal relative to existing alternatives. The `R` code to run the simulation is included in Supplementary Material.

5.1 Part (a): finite-sample properties

We generated data from two scenarios: the NN model

$$Y_i = X_i\beta + Z_i\nu_{1i} + \nu_{2i},$$

and the LL model

$$Y_i = X_i\beta + Z_i\lambda_{1i} + \lambda_{2i},$$

as defined in (3.2) and (3.5), respectively. The matrix X_i had generic vector $x_{ij} = (1, x_{1,ij}, x_{2,ij})^\top$, where $x_{1,ij} = \gamma_i + \zeta_{ij}$, $\gamma_i \sim N(0,1)$, $\zeta_{ij} \sim N(0,1)$ and $x_{2,ij} \sim \text{Binom}(1, 0.5)$. Note that, realistically, covariates may, too, show intraclass correlation. This is accounted for in the generation of $x_{1,ij}$. The matrix Z_i had generic vector $z_{ij} = (1, x_{1,ij})^\top$. The parameters were set as follows: $\beta = (1, 2, 0)^\top$, $\Sigma_1 = \begin{bmatrix} 2 & 0.8 \\ 0.8 & 1 \end{bmatrix}$, $\Sigma_{2i} = \sigma^2 I_{n_i}$, and $\sigma = 1$. (In this setting, the random effects have a Pearson's correlation equal to 0.56.) We considered a balanced design for all possible combinations of $n \in \{5, 10\}$ repetitions and $M \in \{20, 50\}$ clusters. In total, there were 8 simulation cases (4 sample sizes for each of the 2 scenarios). For each case, we independently replicated $R = 500$ datasets

Table 2: Mean square error of the estimator for the generalized Laplace mixed-effects model with data generated under the Normal-Normal and Laplace-Laplace scenarios.

Sample size (n, M)	β_0	β_1	β_2	ς_{11}	ς_{12}	ς_{22}	ψ_2	α_1	α_2
Normal-Normal scenario									
$(5, 20)$	0.150	0.070	0.062	0.694	0.216	0.184	0.035	0.120	0.033
$(10, 20)$	0.128	0.059	0.028	0.552	0.187	0.137	0.013	0.115	0.003
$(5, 100)$	0.029	0.014	0.012	0.157	0.048	0.039	0.007	0.028	0.002
$(10, 100)$	0.024	0.013	0.005	0.106	0.034	0.031	0.003	0.021	0.001
Laplace-Laplace scenario									
$(5, 20)$	0.102	0.052	0.032	0.867	0.270	0.273	0.109	0.382	0.108
$(10, 20)$	0.081	0.041	0.011	0.898	0.251	0.220	0.096	0.279	0.062
$(5, 100)$	0.017	0.008	0.006	0.186	0.055	0.056	0.023	0.069	0.025
$(10, 100)$	0.015	0.007	0.002	0.158	0.050	0.047	0.040	0.055	0.031

Table 3: Coverage of 95% confidence intervals for the generalized Laplace mixed-effects model with data generated under the Normal-Normal (NN) and Laplace-Laplace (LL) scenarios.

	Normal-Normal scenario			*Laplace-Laplace scenario*		
Sample size (n, M)	β_0	β_1	β_2	β_0	β_1	β_2
$(5, 20)$	0.914	0.912	0.944	0.938	0.930	0.944
$(10, 20)$	0.914	0.930	0.930	0.946	0.932	0.954
$(5, 100)$	0.950	0.940	0.948	0.968	0.950	0.948
$(10, 100)$	0.938	0.940	0.936	0.960	0.962	0.938

and fitted the GLME model (3.1) to each dataset. For estimation, we used the multi-start strategy described in Section 4.1 with $K = 8$ quadrature points per integral.

We then estimated element-wise bias and MSE for the estimators of the fixed effects β, variance-covariance of the random effects $\Psi_1 = \frac{1}{\alpha_1}\Sigma_1 \equiv \begin{bmatrix} \varsigma_{11} & \varsigma_{12} \\ \varsigma_{12} & \varsigma_{22} \end{bmatrix}$, variance of the error $\psi_2 = \sigma^2/\alpha_2$, and shape parameter α, averaged over all replications. We also computed element-wise 95% confidence intervals for β using standard errors based on the numerically-approximated Hessian of the log-integrated likelihood (as detailed in Section 4.1). Coverage probabilities were estimated by the proportions (over 500 replications) of confidence intervals containing the true parameter. (Note that the third element of β is 0, hence one minus the coverage gives the observed significance level of the corresponding Wald test.) Finally, we calculated the rejections rates at the 5% significance level of the LRT for H_0: $\alpha_1 = \alpha_2 = 0$ vs H_1: $0 < \alpha_1 \leq 1$ and $0 < \alpha_2 \leq 1$. Note that the rejection rate is an estimate of the level of significance of the test under the NN scenario, but an estimate of its power under the LL scenario.

The results are reported in Supplementary Table 1 and Supplementary Figure 1 for the bias, and in Table 2 for the MSE. Bias and MSE for $\hat{\beta}$ and $\hat{\psi}_2$ were small at all considered sample sizes and in both scenarios. Comparatively, the estimators of Ψ_1 and α showed larger MSE values, though these decreased rapidly with increasing number of clusters M. Bias was, in contrast, conspicuous for $\hat{\alpha}$ under the LL scenario but, again, it decreased with increasing M.

The confidence intervals for β (Table 3) had coverage close to the nominal 95% level at all considered sample sizes and in both scenarios. Similarly, rejection rates of the LRT were reasonably close to the nominal 5% (Table 4) under the true null hypothesis (NN scenario), though slightly

conservative at larger values of M. On the other hand, the test was rather powerful. Even with as little as $M = 20$ clusters, the power was above 90% under the LL scenario.

Table 4: Rejection rates of the likelihood ratio test for the Normal-Normal hypothesis (1) in Table 1 at the nominal 5% significance level for the Normal-Normal and Laplace-Laplace scenarios.

Sample size (n, M)	*Normal-Normal scenario*	*Laplace-Laplace scenario*
$(5, 20)$	0.042	0.911
$(10, 20)$	0.040	0.998
$(5, 100)$	0.018	1.000
$(10, 100)$	0.022	1.000

5.2 Part (b): relative performance

We generated data from two scenarios: the LL model

$$Y_i = X_i\beta + Z_i\lambda_{1i} + \lambda_{2i},$$

as defined in (3.5), and the model (which we label 'TT') with multivariate t-distributions for both random effects and errors, i.e.

$$Y_i = X_i\beta + Z_i\varepsilon_{1i} + \varepsilon_{2i},$$

where $\varepsilon_{1i} \sim t_q(0, \Sigma_1, \delta_1)$ and $\varepsilon_{2i} \sim t_{n_i}(0, \Sigma_2, \delta_2)$. We set $\delta_1 = \delta_2 = 4$. All other simulation settings were the same as in the previous section. In addition to our proposed GLME model (3.1), we fitted a standard LME model and an LME model with t-distributed random effects and errors (tLME).[16] The former is available in the R package `nlme`, while the latter is implemented in the R package `heavy`.[55] Note that the t-distribution has heavier tails than the normal distribution and, thus, represents an alternative to the Laplace distribution.

In this part of the simulation study, we compared average bias and MSE for the estimators of β and Ψ_1 (in the TT model, $\Psi_1 = \frac{\delta_1}{\delta_1 - 2}\Sigma_1$), as well as coverage probabilities for β.

The results are reported in Supplementary Table 2 and Supplementary Figure 2 for the bias, and in Table 5 for the MSE. Our proposed GLME estimator outperformed the LME estimator at all considered sample sizes and in both scenarios. This was expected since LME cannot handle data from heavy-tailed distributions. This is particularly evident from the very large values of the bias and MSE for $\hat{\Psi}_1$, with out-of-control estimates under the TT scenario, which signal convergence issues in LME. Although the very same LME estimates were used as starting values in GLME fitting, our estimator was able to yield acceptable results. Surprisingly, bias and MSE from tLME were also out of control in both scenarios, not only for $\hat{\Psi}_1$, but also for $\hat{\beta}$. Finally, coverage of the confidence intervals for β was close to the nominal 95% for GLME and LME, but off the mark for tLME (Table 6).

We are not able to explain why tLME failed so dramatically, even under its own TT scenario, but we cannot exclude possible bugs in the software. Whatever the explanation, we can conclude that GLME estimation offers a reliable, robust, and superior alternative to existing tools for mixed-effects modeling when data originate from particular heavy-tailed distributions.

Table 5: Mean square error of the estimator for the generalized Laplace mixed-effects (GLME) model, linear mixed-effects (LME) model, and heavy-tailed LME (tLME) model with data generated under the Laplace-Laplace and t-t scenarios.

Sample size (n, M)	β_0	β_1	β_2	ς_{11}	ς_{12}	ς_{22}
Laplace-Laplace scenario						
				GLME		
$(5, 20)$	0.102	0.052	0.032	0.867	0.270	0.273
$(10, 20)$	0.081	0.041	0.011	0.898	0.251	0.220
$(5, 100)$	0.017	0.008	0.006	0.186	0.055	0.056
$(10, 100)$	0.015	0.007	0.002	0.158	0.050	0.047
				LME		
$(5, 20)$	0.138	0.066	0.057	291.645	364.672	436.508
$(10, 20)$	0.110	0.056	0.024	39.578	70.326	112.517
$(5, 100)$	0.026	0.012	0.011	0.873	6.082	9.955
$(10, 100)$	0.023	0.011	0.005	0.516	4.863	8.347
				tLME		
$(5, 20)$	49850.301	1982.035	0.017	7.98×10^{10}	7.52×10^{9}	1.65×10^{9}
$(10, 20)$	4447.144	23.154	0.008	2.80×10^{10}	4.12×10^{8}	7.46×10^{6}
$(5, 100)$	1576.992	3.127	0.003	1.54×10^{8}	2.91×10^{5}	1.14×10^{3}
$(10, 100)$	537.904	2.169	0.001	2.06×10^{7}	4.23×10^{4}	417.90
t-t scenario						
				GLME		
$(5, 20)$	0.220	0.097	0.085	7.714	1.701	1.583
$(10, 20)$	0.173	0.076	0.031	6.081	1.494	1.489
$(5, 100)$	0.044	0.018	0.014	4.015	0.781	0.966
$(10, 100)$	0.035	0.016	0.006	4.076	0.730	0.924
				LME		
$(5, 20)$	0.265	0.114	0.113	1.10×10^{16}	1.12×10^{16}	1.13×10^{16}
$(10, 20)$	0.217	0.094	0.051	76.51	126.95	3.43×10^{3}
$(5, 100)$	0.057	0.023	0.023	32.66	51.27	103.85
$(10, 100)$	0.045	0.021	0.009	12.35	19.30	57.87
				tLME		
$(5, 20)$	8.77×10^{4}	877.470	0.033	1.54×10^{11}	1.28×10^{9}	4.68×10^{7}
$(10, 20)$	1.11×10^{4}	61.823	0.020	2.74×10^{9}	1.89×10^{7}	4.78×10^{5}
$(5, 100)$	5.88×10^{3}	9.098	0.006	1.92×10^{9}	1.60×10^{6}	4.77×10^{3}
$(10, 100)$	2.36×10^{3}	5.901	0.004	1.56×10^{8}	3.02×10^{5}	1.14×10^{3}

Table 6: Coverage of 95% confidence intervals for the generalized Laplace mixed-effects (GLME) model, linear mixed-effects (LME) model, and heavy-tailed LME (tLME) model with data generated under the Laplace-Laplace and t-t scenarios.

	Laplace-Laplace scenario			*t-t scenario*		
Sample size (n, M)	β_0	β_1	β_2	β_0	β_1	β_2
			GLME			
$(5, 20)$	0.938	0.930	0.944	0.914	0.924	0.934
$(10, 20)$	0.946	0.932	0.954	0.934	0.932	0.932
$(5, 100)$	0.968	0.950	0.948	0.934	0.960	0.956
$(10, 100)$	0.960	0.962	0.938	0.940	0.942	0.938
			LME			
$(5, 20)$	0.946	0.942	0.946	0.936	0.932	0.949
$(10, 20)$	0.950	0.940	0.954	0.958	0.944	0.928
$(5, 100)$	0.958	0.954	0.948	0.940	0.954	0.950
$(10, 100)$	0.960	0.960	0.952	0.948	0.956	0.964
			tLME			
$(5, 20)$	0.994	0.982	1.000	0.988	0.982	1.000
$(10, 20)$	0.998	1.000	1.000	1.000	1.000	1.000
$(5, 100)$	0.758	0.864	1.000	0.774	0.868	1.000
$(10, 100)$	0.609	0.711	1.000	0.646	0.702	1.000

6 Examples

In this section, we present two real-data analyses that are relevant to medical researchers. In the first example, we analyze repeated measurements from a clinical trial on Crohn's disease patients. In the second example, we revisit a classical biological dataset with observations from a weight gain experiment in rats.

6.1 Repeated measurements in clinical trials

Ten Crohn's disease patients with endoscopic recurrence were followed over time.[56] Colonoscopy was performed and surrogate markers of disease activity were collected on four occasions. One of the goals of this trial was to assess the association between fecal calprotectin (FC – mg/kg) and endoscopic score (ES – Rutgeerts). The data were analyzed using a log-linear median regression model under the assumption of independence between measurements.[56] Here, we take the multilevel structure into account and analyze the data using the following three models: the NN model

$$Y_i = X_i\beta + \nu_{1i} + \nu_{2i}, \quad i = 1, \ldots, 10,$$

the LL model

$$Y_i = X_i\beta + \lambda_{1i} + \lambda_{2i}, \quad i = 1, \ldots, 10,$$

and the GL model

$$Y_i = X_i\beta + \varepsilon_{1i} + \varepsilon_{2i}, \quad i = 1, \ldots, 10,$$

where the response Y_i is a 4×1 vector of log-FC measurements taken on patient $i = 1, \ldots, 10$, X_i is a design matrix with generic vector $x_{ij} = (1, w_{ij})^\top$ and w_{ij} is the ES measurement on occasion j, $\beta = (\beta_0, \beta_1)^\top$, $\nu_{1i} \sim \mathcal{N}(0, \psi_1)$, $\nu_{2i} \sim \mathcal{N}_4(0, \psi_2^2 I_4)$, $\lambda_{1i} \sim \mathcal{L}(0, \psi_1)$, $\lambda_{2i} \sim \mathcal{L}_4(0, \psi_2^2 I_4)$, $\varepsilon_{1i} \sim \mathcal{GL}(0, \sigma_1, \alpha_1)$, and $\varepsilon_{2i} \sim \mathcal{GL}_4(0, \sigma_2, \alpha_2)$.

Here, the parameters of interest are the slope β_1 and the intraclass correlation coefficient (ICC) $\rho = \psi_1^2/(\psi_1^2 + \psi_2^2)$, where $\psi_1^2 = \frac{1}{\alpha_1}\sigma_1^2$ and $\psi_2^2 = \frac{1}{\alpha_2}\sigma_2^2$ in the GL model. Estimation was carried out using the `nlmm` package. Given the relatively small number of clusters, we fitted the GL model using a multi-start strategy as introduced in Section 4.1 with values $\hat{\tau}_a^{(0)} \in \{-6.9067, -2.1972, -0.8473, 0, 0.8473, 2.1972, 6.9067\}$ (which correspond to the logit of 0.001, 0.1, 0.3, 0.5, 0.7, 0.9, and 0.999). In addition, we used $K = 10$ quadrature points.

The results are shown in Table 7. The estimates of the regression coefficients β tallied across models, although the standard errors from the NN model were larger than those compared to the LL and GL models. Moreover, there seems to be a discrepancy between the NN model and the other models in terms of ICC, with the former showing a larger value (16% vs 13%).

As compared to the NN model, the GL model gave a larger value of the log-likelihood, suggesting that the goodness of the fit is improved by using the GL distribution instead the normal distribution. In particular, there is evidence that the random effects follow a Laplace distribution ($\hat{\alpha}_1 = 0.999$), and, in general, that both the random effects and the error term have heavier tails than normal, although the estimate of α_2 was more uncertain than that of α_1. The LRT statistic comparing the fitted NN and GL models was 7.455 with weights $w = (0.5, 0.5, 0)^\top$ and p-value equal to 0.003, which gives further support to GLME regression. In contrast, the likelihood of the LL model was very similar to that of the GL model, while the corresponding LRT had a large p-value (0.218).

Table 7: Association between fecal calprotectin and endoscopic score in Crohn's disease patients. Estimates of the fixed effects (β), random effects parameter (ψ_1^2), intraclass correlation (ρ), and shape parameter (α) from three models. Standard errors (SE) are given for β, ψ^2, and, unless constrained (†), for α. The log-likelihood (ℓ) is reported in brackets.

	β_0	β_1	ψ_1^2	ρ	α_1	α_2
Normal-Normal ($\ell = -18.2$)						
Estimate	3.299	0.906	0.024	0.161	$0.000^\dagger$	$0.000^\dagger$
SE	0.111	0.057	0.128			
Laplace-Laplace ($\ell = -14.8$)						
Estimate	3.263	0.913	0.021	0.126	$1.000^\dagger$	$1.000^\dagger$
SE	0.088	0.042	0.150			
Generalized Laplace ($\ell = -14.4$)						
Estimate	3.266	0.914	0.021	0.126	0.999	0.569
SE	0.096	0.043	0.169		0.001	0.397

6.2 Growth curves

In a weight gain experiment, 30 rats were randomly assigned to three treatment groups: treatment 1, a control (no additive); treatments 2 and 3, which consisted of two different additives (thiouracil and thyroxin, respectively) to the rats drinking water.[57] Weight (grams) of the rats was measured at baseline (week 0) and at weeks 1, 2, 3, and 4. Data on three of the 10 rats from the thyroxin group were subsequently removed due to an accident at the beginning of the study. Supplementary Figure 3 shows estimated intercepts and slopes obtained from rat-specific LS regressions of the type

$$Y_{i,k} = X_i\beta_{i,k} + \varepsilon_{i,k}, \quad \varepsilon_{i,k} \sim \mathcal{N}_{n_i}(0, \Sigma_{i,k}), \tag{6.1}$$

where the response $Y_{i,k}$ is an $n_i \times 1$ vector of weight measurements $Y_{ij,k}$ taken on rat $i = 1, \ldots, M_k$ on occasion $j = 1, \ldots, 5$ conditional on treatment group $k = \{1, 2, 3\}$, X_i is a design matrix with generic vector $x_{ij} = (1, t_j)^\top$, $t_j = j - 1$, $\beta_{i,k} = (\beta_{0i,k}, \beta_{1i,k})^\top$, and $\Sigma_{i,k} = \sigma_{i,k}^2 I_{n_i}$. (Note that $M_1 = M_2 = 10$ and $M_3 = 7$.) It is evident that the weight of rats treated with thiouracil grew slower than the controls', though at baseline the former tended to be heavier than the latter. In contrast, rats in the control and thyroxin groups had, on average, similar intercepts and slopes. The distributions of intercepts and slopes showed the presence of skewness and bimodality. Therefore, some degree of robustness against departures from normality might be needed.

To model the heterogeneity between rats within each treatment group, random intercepts and slopes were included in the following two models: the NN model

$$Y_{i,k} = X_i\beta_k + Z_i\nu_{1i} + \nu_{2i}, \qquad j = 1, \ldots, 5, \quad i = 1, \ldots, M_k, \quad k = 1, 2, 3,$$

and the GL model

$$Y_{i,k} = X_i\beta_k + Z_i\varepsilon_{1i} + \varepsilon_{2i}, \qquad j = 1, \ldots, 5, \quad i = 1, \ldots, M_k, \quad k = 1, 2, 3,$$

where X_i was defined as in (6.1), $\beta_k = (\beta_{0,k}, \beta_{1,k})^\top$, $Z_i = X_i$, and where we assumed $\nu_{1i} \sim \mathcal{N}_2(0, \Psi_1)$, $\nu_{2i} \sim \mathcal{N}_{n_i}(0, \Psi_{2i})$, $\varepsilon_{1i} \sim \mathcal{GL}_2(0, \Sigma_1, \alpha_1)$, and $\varepsilon_{2i} \sim \mathcal{GL}_{n_i}(0, \Sigma_{2i}, \alpha_2)$. The variance-covariance of the random effects is given by Ψ_1, where $\Psi_1 = \frac{1}{\alpha_1}\Sigma_1$ in the GL model. Similarly, the variance-covariance of the error is Ψ_{2i}, where $\Psi_{2i} = \frac{1}{\alpha_2}\Sigma_{2i}$ in the GL model.

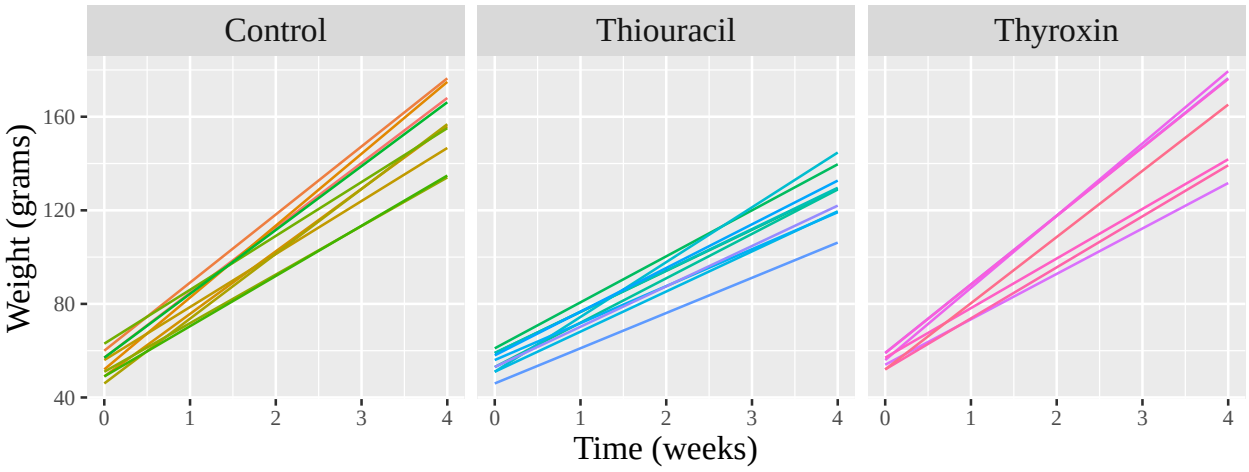

Figure 3: Rats weight gain data. Rat-specific predicted growth trajectories by treatment group using the heteroscedastic generalized Laplace mixed-effects model.

The 2×2 matrix Ψ_1 is symmetric

$$\Psi_1 = \begin{bmatrix} \varsigma_{11} & \varsigma_{12} \\ \varsigma_{12} & \varsigma_{22} \end{bmatrix}.$$

Further, in one specification of the NN and GL models above, we considered homoscedastic errors $\Psi_{2i} = \psi_2^2 I_{n_i}$, while, in another specification, we fitted heteroscedastic models using the `varIdent` function in `nlme`,

$$\text{var}(\varepsilon_{2ij}|\varepsilon_{1i}) \equiv \psi_{2ij}^2 = \psi_2^2 \delta_j^2,$$

where δ_j represents the ratio between the standard deviations at time t_j and time t_1 (baseline), $j = 2, \ldots, 5$, with the identifiability constraint $\delta_1 = 1$.[14] Therefore, the parameter of the variance function as defined in (4.9) is given by $\xi_2 = (\log \delta_2, \log \delta_3, \log \delta_4, \log \delta_5)^\top$.

Estimation was carried out using the `nlmm` package. Given the relatively small number of clusters, we fitted the GL models using a multi-start strategy as introduced in Section 4.1 with values $\hat{\tau}_a^{(0)} \in \{-6.9067, -2.1972, -0.8473, 0, 0.8473, 2.1972, 6.9067\}$ (which correspond to the logit of 0.001, 0.1, 0.3, 0.5, 0.7, 0.9, and 0.999). In addition, we used $K = 10$ quadrature points.

The NN and GL models, either homoscedastic or heteroscedastic, gave similar estimates of the fixed effects as well as of the variances of the random effects (Table 8). The residual standard deviations $\{\hat{\psi}_{2ij},\ j = 1, \ldots, 5\}$ from the NN and GL models were equal to $\{0.486 \times 10^{-3}, 5.847, 5.047, 1.902, 6.414)\}$ and to $\{0.731 \times 10^{-4}, 2.215, 1.876, 1.191, 2.712\}$, respectively. Differences between treatment groups in terms of variability of the rat-specific trajectories and residual heteroscedasticity can be better appreciated in Figure 3. Predictions of rat weight at each time point were obtained from the GL model with heteroscedastic errors, with random effects estimated via the best linear predictor (4.6).

The homoscedastic NN and GL models gave similar values of the log-likelihood (the LRT statistic comparing these two models had a p-value equal to 0.316). Introducing a within-cluster variance function in the model improved the fits, though only moderately, resulting in similar values of the log-likelihood for the heteroscedastic NN and GL models (the LRT statistic comparing these two models had a p-value close to 1). These results were consistent with the estimates of the shape parameters in the GL models, which were not far from 0, thus suggesting that random effects and

error term follow normal distributions (though the large standard errors prevent us from drawing strong conclusions). In summary, the normal mixed model with a time-specific error variance function seems appropriate for these data.

Table 8: Rats weight gain data. Estimates of the fixed effects (β), random effects parameter (ς), and shape parameter (α) from four models. Standard errors (SE) are given for β and, unless constrained (†), for α. The log-likelihood (ℓ) is reported in brackets.

	$\beta_{0,1}$	$\beta_{0,2}$	$\beta_{0,3}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	ς_{11}	ς_{12}	ς_{22}	α_1	α_2
Homoscedastic Normal-Normal ($\ell = -447.5$)											
Estimate	52.880	57.700	52.086	26.480	17.050	27.143	27.550	−1.974	12.202	$0.000^\dagger$	$0.000^\dagger$
SE	2.016	2.016	2.409	1.214	1.214	1.451					
Homoscedastic Generalized Laplace ($\ell = -447.4$)											
Estimate	53.562	57.508	53.293	26.592	17.351	27.098	27.411	−0.536	11.847	0.039	0.319
SE	1.974	2.027	2.349	1.172	1.176	1.426				0.089	0.273
Heteroscedastic Normal-Normal ($\ell = -439.3$)											
Estimate	54.000	54.700	55.571	25.554	17.944	25.748	19.178	1.905	12.129	$0.000^\dagger$	$0.000^\dagger$
SE	1.417	1.417	1.693	1.142	1.142	1.365					
Heteroscedastic Generalized Laplace ($\ell = -440.4$)											
Estimate	54.020	55.444	55.924	26.027	18.359	26.225	20.240	3.294	12.343	0.128	0.164
SE	1.477	1.432	1.669	1.171	1.135	1.487				0.161	0.380

7 Final remarks

In the words of Wilson "No phenomenon is better known perhaps, as a plain matter of fact, than that the frequencies which we actually meet in everyday work in economics, in biometrics, or in vital statistics, very frequently fail to conform at all closely to the so-called normal distribution".[3] Kotz and colleagues[2] echo Wilson's observations on the inadequacy of the normal distribution in many practical applications and give a systematic exposition of the Laplace distribution, an unjustifiably neglected error law which can be "a natural and sometimes superior alternative to the normal law".

Our proposed generalized Laplace mixed models bring together the normal and Laplace distributions showing that these models represent a *family* of sensible alternatives as they flexibly introduce shape parameters in the modeling process. Estimation can be approached in different ways. The algorithm based on numerical quadrature discussed in this paper takes advantage of the scale mixture representation of the Laplace distribution which provides the opportunity for computational simplification. In a simulation study with small to moderate sample sizes, our estimator provided satisfactory results in terms of bias, MSE and coverage, and outperformed a competing mixed-effects model based on heavy-tailed distributions. Moreover, we developed a model selection strategy based on likelihood ratio tests for both regular and non-regular inference which performed satisfactorily. Although we made our models general enough to allow for residual heteroscedasticity, we did not include within-cluster correlation structures.[14] Such a development is mathematically straightforward, but it does require additional computational and programming effort. Moreover, it possible to extend the proposed GLME model to multiple levels of the random effects. If the random effects levels are nested, the estimation procedure based on numerical quadrature as described in Section 4.1 is still feasible, though computationally inefficient. On the other hand, if levels are non-nested, then one must consider alternative estimation approaches.

To reiterate the main point of this study, these convolutions have a large number of potential applications and, as demonstrated using several examples, may provide valuable insight into different aspects of the analysis.

References

1 Geraci M, Cortina-Borja M. The Laplace distribution. Significance. 2018;15(5):10–11.

2 Kotz S, Kozubowski TJ, Podgórski K. The Laplace distribution and generalizations. Boston, MA: Birkhäuser; 2001.

3 Wilson EB. First and second laws of error. Journal of the American Statistical Association. 1923;18(143):841–851.

4 Bassett G, Koenker R. Asymptotic theory of least absolute error regression. Journal of the American Statistical Association. 1978;73(363):618–622.

5 Koenker R, Bassett G. Regression quantiles. Econometrica. 1978;46(1):33–50.

6 Yu KM, Moyeed RA. Bayesian quantile regression. Statistics & Probability Letters. 2001;54(4):437–447.

7 Geraci M, Bottai M. Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics. 2007;8(1):140–154.

8 Farcomeni A. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. Statistics and Computing. 2012;22(1):141–152.

9 Farcomeni A, Viviani S. Longitudinal quantile regression in presence of informative drop-out through longitudinal-survival joint modeling. Statistics in Medicine. 2015;34:1199–1213.

10 Marino MF, Farcomeni A. Linear quantile regression models for longitudinal experiments: An overview. Metron. 2015;73(2):229–247.

11 Portnoy S, Koenker R. The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. Statistical Science. 1997;12(4):279–300.

12 Koenker R, Ng P. A Frisch–Newton algorithm for sparse quantile regression. Acta Mathematicae Applicatae Sinica (English Series). 2005;21(2):225–236.

13 Kozubowski TJ, Nadarajah S. Multitude of Laplace distributions. Statistical Papers. 2010;51(1):127–148.

14 Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS. New York: Springer Verlag; 2000.

15 Demidenko E. Mixed models: Theory and applications with R. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2013.

16 Pinheiro JC, Liu C, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. Journal of Computational and Graphical Statistics. 2001;10(2):249–276.

17 Staudenmayer J, Lake EE, Wand MP. Robustness for general design mixed models using the t-distribution. Statistical Modelling. 2009;9(3):235–255.

18 Zhang D, Davidian M. Linear mixed models with flexible distributions of random effects for longitudinal data. Biometrics. 2001;57(3):795–802.

19 Dandine-Roulland C, Perdry H. The use of the linear mixed model in human genetics. Human Heredity. 2015;80(4):196–206.

20 Raman G, Balk EM, Lai L, Shi J, Chan J, Lutz JS, et al. Evaluation of person-level heterogeneity of treatment effects in published multiperson N-of-1 studies: systematic review and reanalysis. BMJ Open. 2018;8(5):e017641.

21 R Core Team. R: A Language and environment for statistical computing. Vienna, Austria; 2019. Http://www.R-project.org.

22 Kozubowski TJ, Podgórski K, Rychlik I. Multivariate generalized Laplace distribution and related random fields. Journal of Multivariate Analysis. 2013;113:59–72.

23 Yavuz FG, Arslan O. Linear mixed model with Laplace distribution (LLMM). Statistical Papers. 2018;59(1):271–289.

24 Gómez E, Gomez-Viilegas MA, Marín JM. A multivariate generalization of the power exponential family of distributions. Communications in Statistics - Theory and Methods. 1998;27(3):589–600.

25 Ernst MD. A multivariate generalized Laplace distribution. Computational Statistics. 1998;13(2):227–232.

26 Kozubowski TJ, Podgórski K. A multivariate and asymmetric generalization of Laplace distribution. Computational Statistics. 2000;15(4):531–540.

27 Arslan O. An alternative multivariate skew Laplace distribution: Properties and estimation. Statistical Papers. 2010;51(4):865–887.

28 Reed WJ. The normal-Laplace distribution and its relatives. In: Balakrishnan N, Castillo E, Sarabia JM, editors. Advances in Distribution Theory, Order Statistics, and Inference. Boston, MA: Birkhäuser; 2006. p. 61–74.

29 Reed WJ. Brownian-Laplace motion and its use in financial modelling. Communications in Statistics – Theory and Methods. 2007;36(3):473–484.

30 Meintanis SG, Tsionas E. Testing for the generalized normal-Laplace distribution with applications. Computational Statistics & Data Analysis. 2010;54(12):3174–3180.

31 Reed WJ, Jorgensen M. The double Pareto-lognormal distribution—A new parametric model for size distributions. Communications in Statistics – Theory and Methods. 2004;33(8):1733–1753.

32 Geraci M, Bottai M. Linear quantile mixed models. Statistics and Computing. 2014;24(3):461–479.

33 Muir PR, Wallace CC, Done T, Aguirre JD. Limited scope for latitudinal extension of reef corals. Science. 2015;348(6239):1135–1138.

34 Duffy LM, Olson RJ, Lennert-Cody CE, Galván-Magaña F, Bocanegra-Castillo N, Kuhnert PM. Foraging ecology of silky sharks, *Carcharhinus falciformis*, captured by the tuna purse-seine fishery in the eastern Pacific Ocean. Marine Biology. 2015;162(3):571–593.

35 Barneche DR, Kulbicki M, Floeter SR, Friedlander AM, Allen AP. Energetic and ecological constraints on population density of reef fishes. Proceedings of the Royal Society B: Biological Sciences. 2016;283(1823):1–8.

36 Fornaroli R, Cabrini R, Sartori L, Marazzi F, Vracevic D, Mezzanotte V, et al. Predicting the constraint effect of environmental characteristics on macroinvertebrate density and diversity using quantile regression mixed model. Hydrobiologia. 2015;742(1):153–167.

37 Degerud E, Løland KH, Nygård O, Midttun Ø, Ueland PM, Seifert R, et al. Vitamin D status was not associated with 'one-year' progression of coronary artery disease, assessed by coronary angiography in statin-treated patients. European Journal of Preventive Cardiology. 2014;22(5):594–602.

38 Blankenberg S, Salomaa V, Makarova N, Ojeda F, Wild P, Lackner KJ, et al. Troponin I and cardiovascular risk prediction in the general population: the BiomarCaRE consortium. European Heart Journal. 2016;37(30):2428–2437.

39 Ng SW, Howard AG, Wang HJ, Su C, Zhang B. The physical activity transition among adults in China: 1991-2011. Obesity Reviews. 2014;15(S1):27–36.

40 Beets MW, Weaver RG, Turner-McGrievy G, Moore JB, Webster C, Brazendale K, et al. Are we there yet? Compliance with physical activity standards in YMCA afterschool programs. Childhood Obesity. 2016;12(4):237–246.

41 Patel DE, Cumberland PM, Walters BC, Russell-Eggitt I, Cortina-Borja M, Rahi JS. Study of optimal perimetric testing in children (OPTIC): Normative visual field values in children. Ophthalmology. 2015;122(8):1711–1717.

42 Patel DE, Geraci M, Cortina-Borja M. Modelling normative kinetic perimetry isopters using mixed-effects quantile regression. Journal of Vision. 2016;16(7):1–6.

43 Prékopa A. Logarithmic concave measures and functions. Acta Scientiarum Mathematicarum. 1973;34(1):334–343.

44 Geraci M. Mixed-effects models using the normal and the Laplace distributions: A 2×2 convolution scheme for applied research. ArXiv e-prints. 2017;1712.07216.

45 Geraci M. nlmm: Generalized Laplace mixed-effects models; 2019. R package version 1.0.0. https://github.com/marco-geraci/nlmm.

46 McLachlan G, Krishnan T. The EM algorithm and extensions. Hoboken, NJ: John Wiley & Sons; 2008.

47 Geraci M. Letter to the Editor. Statistics and Its Interface. 2019;12(1):71–75.

48 Galarza CE, Lachos VH, Bandyopadhyay D. Quantile regression in linear mixed models: A stochastic approximation EM approach. Statistics and Its Interface. 2017;10(3):471–482.

49 Silvapulle MJ, Sen PK. Constrained statistical inference: Inequality, order, and shape restrictions. New York, NY: John Wiley & Sons; 2004.

50 Shapiro A. Towards a unified theory of inequality constrained testing in multivariate analysis. International Statistical Review. 1988;56(1):49–62.

51 Farcomeni A. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. Statistical Methods in Medical Research. 2008;17(4):347–388.

52 Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and nonlinear mixed effects models; 2014. R package version 3.1-117. http://CRAN.R-project.org/package=nlme.

53 Douglas B, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 2015;67(1):1–48.

54 Geraci M. Linear quantile mixed models: The lqmm package for Laplace quantile regression. Journal of Statistical Software. 2014;57(13):1–29.

55 Osorio F. heavy: Robust estimation using heavy-tailed distributions; 2018. R package version 0.38.19. https://CRAN.R-project.org/package=heavy.

56 Sorrentino D, Paviotti A, Terrosu G, Avellini C, Geraci M, Zarifi D. Low-dose maintenance therapy with infliximab prevents postsurgical recurrence of Crohn's disease. Clinical Gastroenterology and Hepatology. 2010;8(7):591–599.

57 Box GEP. Problems in the analysis of growth and wear curves. Biometrics. 1950;6(4):362–389.