AStA Advances in Statistical Analysis (2024) 108:127–157
https://doi.org/10.1007/s10182-023-00473-z

ORIGINAL PAPER

Robust estimation of fixed effect parameters and variances of linear mixed models: the minimum density power divergence approach

Giovanni Saraceno[1,3] · **Abhik Ghosh**[2] · **Ayanendranath Basu**[2] · **Claudio Agostinelli**[1]

Received: 17 May 2022 / Accepted: 2 March 2023 / Published online: 29 March 2023

Abstract

Many real-life data sets can be analyzed using linear mixed models (LMMs). Since these are ordinarily based on normality assumptions, under small deviations from the model the inference can be highly unstable when the associated parameters are estimated by classical methods. On the other hand, the density power divergence (DPD) family, which measures the discrepancy between two probability density functions, has been successfully used to build robust estimators with high stability associated with minimal loss in efficiency. Here, we develop the minimum DPD estimator (MDPDE) for independent but non-identically distributed observations for LMMs according to the variance components model. We prove that the theoretical properties hold, including consistency and asymptotic normality of the estimators. The influence function and sensitivity measures are computed to explore the robustness properties. As a data-based choice of the MDPDE tuning parameter α is very important, we propose two candidates as "optimal" choices, where optimality is in the sense of choosing the strongest downweighting that is necessary for the particular data set. We conduct a simulation study comparing the proposed MDPDE, for different values of α, with S-estimators, M-estimators and the classical maximum likelihood estimator, considering different levels of contamination. Finally, we illustrate the performance of our proposal on a real-data example.

Keywords Linear mixed models · Minimum density power divergence estimator · Robustness

Mathematics Subject Classification 62G35 · 62G05

✉ Giovanni Saraceno
gsaracen@buffalo.edu

Extended author information available on the last page of the article

1 Introduction

A major interest in statistics concerns the estimation of averages and their variation. The most commonly used method for this purpose is, probably, the *Linear Model* (LM). In this model, to give an example from a two-way layout, the expected value (mean) μ_{ij} of an observation y_{ij} may be expressed as a linear combination of unknown parameters such as $\mu_{ij} = \mu + \alpha_i + \beta_j$, where μ, α_i and β_j are the constants which we are interested in estimating. The linearity in the parameters means that we can write a linear model in the form $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon_i}$, where $\boldsymbol{\beta}$ is the vector of unknown parameters, and the $\boldsymbol{X}_i$s are known matrices. This formulation is the same as that used in case of linear regression model. In the present work, we consider the linear mixed models (LMMs), in which some (unknown) parameters are not treated as constants but as random variables. Random terms come into play when some items cannot be considered as fixed quantities, although their distributions are of interest. Hence, they are the tools to generalize the results to the entire population under study. The types of data that may be appropriately analyzed by LMMs include (i) *Clustered data*, where the dependent variable is measured once for each subject (the unit of analysis) and the units of analysis are grouped into, or nested within, clusters; (ii) *Repeated-measures data*, where the dependent variable is measured more than once on the same unit of analysis across levels of a factor, which may be time or other experimental conditions; (iii) *Longitudinal data*, where the dependent variable is measured at several points in time for each unit of analysis. For a general review of LMs and LMMs, see McCulloch and Searle (2001).

The standard methods used to estimate the parameters in LMMs are methods of maximum likelihood and restricted maximum likelihood. Generally, LMMs are based on normality assumptions, and it is well-known that these classical methods are not robust and can be greatly affected by the presence of small deviations from the assumptions. Furthermore, outlier detection for modern large data sets can be very challenging and, in any case, robust techniques cannot be replaced by the application of classical methods on outlier deleted data.

To answer the need for robust estimation in linear mixed models, a few methods have been proposed. The initial attempts were based on weighted versions of the log-likelihood function (see Huggins 1993a, b; Huggins and Staudte 1994; Stahel and Welsh 1994; Richardson and Welsh 1995; Richardson 1997; Welsh and Richardson 1997). Another attempt, discussed in Welsh and Richardson (1997), of robustifying linear mixed models consists of replacing the Gaussian distribution by the Student's t distribution (see also Lange et al. 1989; Pinheiro et al. 2001). However, this modification of the error distribution is intractable and complicated to implement. In Copt and Victoria-Feser (2006), a multivariate high breakdown point S-estimator, namely the CVFS-estimator, has been adapted to the linear mixed models setup, while the estimator given by Koller (2013), namely the SMDM-estimator, attempts to achieve robustness by a robustification of the score equations. Robust estimators have been proposed, more generally, for generalized linear mixed models by Yau and Kuk (2002) and Sinha (2004).

The density power divergence (DPD) (Basu et al. (1998)), which measures the discrepancy between two probability density functions, has been successfully used to build robust estimators for independent and identically distributed observations. In Ghosh and Basu (2013), the construction of the DPD and the corresponding minimum DPD estimator (MDPDE) has been extended to the case of independent but non-identically distributed data. This approach and theory covers the linear regression model and has later been extended to more general parametric regression models (Ghosh and Basu 2016, 2019; Castilla et al. 2018, 2021; Ghosh 2019, etc.). This MDPDE has become widely popular in recent times due to its good (asymptotic) efficiency along with high robustness, easy computability and direct interpretation as an intuitive generalization of the maximum likelihood estimator (MLE).

In the present work, our aim is to propose a robust estimator of the fixed effect parameters and variances of random effects under the linear mixed model set up. In particular, we are going to consider independent random effects according to the standard variance components models. This is done by an application of the definition and properties of the MDPDE, as formulated by Ghosh and Basu (2013), to the linear mixed models scenario. However, in this case, the observations have a complicated, non-identically distributed structure, which makes this adaptation non-trivial, in the computation of the estimator as well as in the theoretical derivations. We show that under appropriate conditions on the model matrices, the asymptotic and robustness properties hold, and the resulting estimator outperforms the competing robust estimators both in the presence and in the absence of contamination in sample data. Furthermore, the calculation of the sensitivity measures allows us to find "optimal" values for the tuning parameter α, where optimality is in terms of the right amount of data-specific downweighting needed to achieve robustness with minimal loss of efficiency. This is a valuable new contribution to the literature of DPD-based inference since, in contrast with the previous knowledge of α as trade-off between robustness and efficiency, our results lead to a small positive value of α as an optimum choice in practical applications. Furthermore, the estimation of random effects has been also considered, which is an important aspect in the study of LMMs. In particular, we provide a closed form to compute the random effects estimates based on the minimum DPD estimation. Large-scale numerical explorations, including an extensive simulation study, are provided to substantiate the theory developed and justify our claims of the superiority of the MDPDE in the proposed application domain of LMMs. Finally, our proposal is applied successfully (and robustly) to analyze two real-life data, one on orthodontic measures and another on foveal and extrafoveal vision acuity.

The rest of the paper is organized as follows. The MDPDE for non-homogeneous observations described in Ghosh and Basu (2013) is briefly presented in Sect. 2. In Sect. 3, we define the proposed estimator in case of linear mixed models, considering the estimation of fixed effect parameters and variances of random effects, and the prediction of random terms. The asymptotic and robustness properties of our procedures are considered in Sect. 4 together with the computational aspects and the case of balanced data. Section 5 reports the organization and results of the simulation study we conducted, comparing the performance of the MDPDE to the most recent methods, exploring also the case of contaminated data. Section 6 provides the

application of the proposed estimator to the real-data example on orthodontic measures. Concluding remarks are presented in Sect. 7. The proof of the main theorem is reported in Appendix A, while the Supplementary Material contains the derivation of equivariance properties, some additional theoretical and Monte Carlo results, an example which shows the robustness of predicted random terms, and the application to the real-data example on foveal and extrafoveal vision acuity.

2 The MDPDE for independent non-homogeneous observations

The density power divergence family was first introduced by Basu et al. (1998) as a measure of discrepancy between two probability density functions. The authors used this measure to robustly estimate the model parameters under the usual setup of independent and identically distributed data. The density power divergence measure $d_\alpha(g,f)$ between two probability densities g and f is defined, in terms of a single tuning parameter $\alpha \geq 0$, as

$$d_\alpha(g,f) = \int \left\{ f^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) f^\alpha g + \frac{1}{\alpha} g^{1+\alpha} \right\} \qquad \text{if } \alpha > 0, \tag{1}$$

$$d_0(g,f) = \int g \ln\left(\frac{g}{f}\right) \quad \text{if } \alpha = 0, \tag{2}$$

where ln denotes the natural logarithm. Basu et al. (1998) demonstrated that the tuning parameter α controls the trade-off between efficiency and robustness of the resulting estimator. With increasing α, the estimator acquires greater stability with a slight loss in efficiency. Since the divergence is not defined for $\alpha = 0$, $d_0(g,f)$ in Equation (2) represents the divergence obtained in the limit of (1) as $\alpha \to 0$, which corresponds to a version of the Kullback–Leibler divergence. On the other hand, $\alpha = 1$ generates the squared L_2 distance.

Let G be the true data generating distribution and g the corresponding density function. To model g, consider the parametric family of densities $\mathcal{F}_\theta = \{f_\theta : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$. The minimizer of $d_\alpha(g,f_\theta)$ over $\boldsymbol{\theta} \in \Theta$, whenever it exists, is the minimum DPD functional at the distribution point G. Note that the third term of the divergence $d_\alpha(g,f_\theta)$ is independent of $\boldsymbol{\theta}$; hence, it can be discarded from the objective function as it has no role in the minimization process. Consider a sequence of independent and identically distributed (i.i.d) observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ from the true distribution G. Using the empirical distribution function G_n in place of G, the MDPDE of $\boldsymbol{\theta}$ can be obtained by minimizing

$$\int f_\theta^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^{n} f_\theta^\alpha(\boldsymbol{Y}_i)$$

over $\boldsymbol{\theta} \in \Theta$. In the above equation, the empirical distribution function is used to approximate its theoretical version (or, alternatively, the sample mean is used to

approximate the population mean). Note that, it is valid in case of continuous densities also. See Basu et al. (2011) for more details and examples.

Ghosh and Basu (2013) generalized the above concept of robust minimum DPD estimation to the more general case of independent non-homogeneous observations, i.e., they considered the case where the observed data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are independent but for each i, $\boldsymbol{Y}_i \sim g_i$ where $g_1, \ldots, g_n$ are possibly different densities with respect to some common dominating measure. We model g_i by the family $\mathcal{F}_{i,\boldsymbol{\theta}} = \{f_i(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ for $i \in \{1, \ldots, n\}$. While the distributions $f_i(\cdot, \boldsymbol{\theta})$ can be distinct, they share the same parameter vector $\boldsymbol{\theta}$. Ghosh and Basu (2013) proposed to minimize the average divergence between the data points and the model densities which leads to the minimization of the objective function

$$H_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left[\int f_i(\boldsymbol{y}, \boldsymbol{\theta})^{1+\alpha} dy - \left(1 + \frac{1}{\alpha}\right) f_i(\boldsymbol{Y}_i, \boldsymbol{\theta})^{\alpha}\right] = \frac{1}{n}\sum_{i=1}^{n} H_i(\boldsymbol{Y}_i, \boldsymbol{\theta}), \quad (3)$$

where $H_i(\boldsymbol{Y}_i, \boldsymbol{\theta})$ is the indicated term within the square brackets in the above equation. Differentiating the above expression with respect to $\boldsymbol{\theta}$, we get the estimating equations of the MDPDE for non-homogeneous observations. Note that the estimating equation is unbiased when each g_i belongs to the model family $\mathcal{F}_{i,\boldsymbol{\theta}}$, respectively. When $\alpha \to 0$, the corresponding objective function reduces to $-\sum_{i=1}^{n} \ln(f_i(\boldsymbol{Y}_i, \boldsymbol{\theta}))/n$, which is the negative of the log-likelihood function. In Section SM–1 of the Supplementary Material, we report Assumptions (A1)–(A7) which are used to prove the asymptotic normality of the MDPDE (Ghosh and Basu 2013).

3 The MDPDE for linear mixed models

The general formulation of a LMM may be expressed as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\mathcal{U}} + \boldsymbol{\epsilon}, \quad (4)$$

where $\boldsymbol{Y} \in \mathbb{R}^d$ is the response vector, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known design matrices, $\boldsymbol{\beta}$ is the parameter vector for fixed effects, $\boldsymbol{\mathcal{U}}$ is the vector of random effects, and $\boldsymbol{\epsilon}$ is the random error vector. The vector $\boldsymbol{\mathcal{U}}$ is assumed to be a random variable, in particular $\boldsymbol{\mathcal{U}} \sim N_q(\boldsymbol{0}, \boldsymbol{D})$, then $\mathbb{E}(\boldsymbol{Y}|\boldsymbol{\mathcal{U}} = \boldsymbol{u}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$. Notice that we will use $\boldsymbol{\mathcal{U}}$ to indicate the random variable and $\boldsymbol{u}$ for its realization. Finally, assume that $\boldsymbol{\epsilon} \sim N_d(\boldsymbol{0}, \boldsymbol{R})$ and $\boldsymbol{\epsilon}$ and $\boldsymbol{\mathcal{U}}$ are independent of each other, then

$$\boldsymbol{Y} \sim N_d\left(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^{\top} + \boldsymbol{R}\right). \quad (5)$$

The estimation of the matrices $\boldsymbol{D}$ and $\boldsymbol{R}$ involves a large amount of parameters; indeed, we need to assume some additional structure to the mixed model. According to the variance components model, the levels of any random effect are assumed to be independent with the same variance. Different effects are assumed independent with possibly different variances. Let the model have r random factors $\boldsymbol{\mathcal{U}}_j$ with q_j levels, $j \in \{1, \ldots, r\}$, with $q = \sum_{j=1}^{r} q_j$. As stated in Christensen (2011), a "natural generalization" of model 5 is to partition the vector $\boldsymbol{\mathcal{U}} = [\boldsymbol{\mathcal{U}}_1, \ldots, \boldsymbol{\mathcal{U}}_r]$ and matrix

$\boldsymbol{Z} = [\boldsymbol{Z}_1 \dots \boldsymbol{Z}_r]$, such that $\boldsymbol{\mathcal{U}}_j$ are independent between each other, i.e., $\mathbb{C}ov(\boldsymbol{\mathcal{U}}_i, \boldsymbol{\mathcal{U}}_j) = 0$ for $i \neq j$, and $\mathbb{C}ov(\boldsymbol{\mathcal{U}}_j) = \boldsymbol{D}_j = \sigma_j^2 \boldsymbol{I}_{q_j}$, $j = 1, \dots, r$. Finally, assume $\boldsymbol{\epsilon}$ and $\boldsymbol{\mathcal{U}}_j$ independent for all j and $\boldsymbol{R} = \sigma_0^2 \boldsymbol{I}_d$, where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

Then, $\boldsymbol{Y} \sim N_d(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V})$ where

$$\boldsymbol{V} = \sigma_0^2 \boldsymbol{I}_d + \sum_{j=1}^{r} \boldsymbol{Z}_j \boldsymbol{Z}_j^\top \sigma_j^2 = \sigma_0^2 \left(\boldsymbol{I}_d + \sum_{j=1}^{r} \boldsymbol{Z}_j \boldsymbol{Z}_j^\top \gamma_j \right), \quad \text{with } \gamma_j = \frac{\sigma_j^2}{\sigma_0^2}.$$

In case of multiple measurements on each of a collection of observational units $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$, where n_i denotes the size of each group and $\sum_{i=1}^{n} n_i = N$, the LMM in equation (4) can be written for each observation as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \sum_{j=1}^{r} \boldsymbol{Z}_{ij} \boldsymbol{\mathcal{U}}_j + \boldsymbol{\epsilon}_i, \qquad i \in \{1, \dots, n\}, \tag{6}$$

where $\boldsymbol{Y}_i$ $(n_i \times 1)$ is the response vector for group i, $\boldsymbol{X}_i$ $(n_i \times k)$ and $\boldsymbol{Z}_{ij}$ $(n_i \times q_j)$ are the model matrices, $\boldsymbol{\beta}$ $(k \times 1)$ is the vector of unknown parameters for fixed effects, and $\boldsymbol{\epsilon}_i$ $(n_i \times 1)$ is the error term. Then, assuming $\boldsymbol{\epsilon}_i \sim \sigma_0^2 \boldsymbol{I}_{n_i}$ and independent with respect to $\boldsymbol{\mathcal{U}}_j$ for all i, j,

$$\boldsymbol{Y}_i \sim N_{n_i}\left(\boldsymbol{X}_i \boldsymbol{\beta}, \boldsymbol{V}_i\right), \qquad i \in \{1, \dots, n\}.$$

where $\boldsymbol{V}_i = \sigma_0^2 (\boldsymbol{I}_{n_i} + \sum_{j=1}^{r} \boldsymbol{Z}_{ij} \boldsymbol{Z}_{ij}^\top \gamma_j)$.

Remark Note that the covariance structure of the linear mixed models considered here also includes the case of standard random intercept and random slope models. The real-data application presented in Sect. 6 provides an example. Indeed, Christensen (Christensen 2011) also considered this model for the discussion of all the variance components estimation methods. Also note that all the methods and results described in this paper can be routinely derived for any other (low-dimensional) parametric structure specified for the variance components in the LMM as per the need, but such situations (beyond the structure considered here) would rarely occur in practice.

3.1 Robust parameter estimation

In this setting, we can obtain the MDPDE for the p-dimensional parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top; \sigma_j^2, j \in \{0, \dots, r\})^\top$, with $p = k + r + 1$, by minimizing the objective function given in Equation (3) with $f_i \equiv N_{n_i}(\boldsymbol{X}_i \boldsymbol{\beta}, \boldsymbol{V}_i)$. Upon simplification, the objective function is given by

$$H_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left[\eta_{i\alpha} (\alpha + 1)^{-\frac{n_i}{2}} - \left(1 + \frac{1}{\alpha}\right) w_i(\boldsymbol{\theta}) \right] \tag{7}$$

where

$$\eta_{i\alpha} = (2\pi)^{-\frac{n_i\alpha}{2}} |\boldsymbol{V}_i|^{-\frac{\alpha}{2}}$$

and

$$w_i(\boldsymbol{\theta}) = \eta_{i\alpha} \exp\left\{-\frac{\alpha}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\}. \tag{8}$$

Differentiating the above equation with respect to $\boldsymbol{\beta}$, we get the corresponding estimating equation for the MDPDE of $\boldsymbol{\beta}$ as

$$\frac{\partial H_n}{\partial \boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\left[-(1+\alpha)w_i(\boldsymbol{\theta})\boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right] = 0. \tag{9}$$

Let $\boldsymbol{U}_{ij}$ denote the partial derivative of the matrix $\boldsymbol{V}_i$ with respect to σ_j^2. We have that $\boldsymbol{U}_{i0} = \boldsymbol{I}_{n_i}$, and $\boldsymbol{U}_{ij} = \boldsymbol{Z}_{ij}\boldsymbol{Z}_{ij}^\top$, $j \in \{1, \ldots, r\}$. Then, the partial derivative of the objective function with respect to σ_j^2, $j \in \{0, \ldots, r\}$, leads to their MDPDE estimating equations as given by

$$\begin{aligned}\frac{\partial H_n}{\partial \sigma_j^2} =& \frac{1}{n}\sum_{i=1}^{n}\Bigg\{-\frac{\alpha\eta_{i\alpha}\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij})}{2(\alpha+1)^{\frac{n_i}{2}}} + \frac{\alpha}{2}\left(1+\frac{1}{\alpha}\right)w_i(\boldsymbol{\theta}) \\ &\times\left[\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}) - (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right]\Bigg\} = 0,\end{aligned} \tag{10}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of the argument matrix. Note that for the case $\alpha = 0$, the estimating equations (9)–(10) correspond to the MLE score equations. Thus, the MDPDE at $\alpha = 0$ is nothing but the usual MLE under LMMs.

3.2 Robust estimation of the random effects

Finally, the estimates obtained by minimizing the objective function in (7) are used to predict the random realizations $\boldsymbol{u}_i|\boldsymbol{Y}_i$, $i = 1, \ldots, n$. Consider the joint distribution of $(\boldsymbol{Y}_i, \boldsymbol{\mathcal{U}}) \sim g_i(\boldsymbol{y}, \boldsymbol{u})$, and we have that $g_i(\boldsymbol{y}, \boldsymbol{u}) = g_i(\boldsymbol{y}|\boldsymbol{u})g(\boldsymbol{u})$, where $g(\boldsymbol{u})$ is the true density of $\boldsymbol{\mathcal{U}}$. Let $f_i(\boldsymbol{y}, \boldsymbol{u})$ be the parametric density to model $g_i(\boldsymbol{y}, \boldsymbol{u})$, then it can be expressed as $f_i(\boldsymbol{y}, \boldsymbol{u}) = f_i(\boldsymbol{y}|\boldsymbol{u})f(\boldsymbol{u})$, which, by the plug-in principle, can be estimated using

$$f_i(\boldsymbol{y}|\boldsymbol{u}) = N_{n_i}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\boldsymbol{u}_i, \hat{\sigma}_0^2\boldsymbol{I}_{n_i}) \text{ and } f(\boldsymbol{u}) = N_q(\boldsymbol{0}, \hat{\boldsymbol{D}}).$$

Hence, we can estimate the random coefficients by minimizing the density power divergence measure between the densities $g_i(\boldsymbol{y}, \boldsymbol{u})$ and $f_i(\boldsymbol{y}, \boldsymbol{u})$ given by

$$H_n(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) = \frac{1}{n}\sum_{i=1}^{n}\kappa_\alpha\left[(\alpha+1)^{-\frac{n_i+q}{2}} + \left(1+\frac{1}{\alpha}\right)c_i(\boldsymbol{u}_i)\right],$$

where

$$c_i(\boldsymbol{u}_i) = \exp\left\{ -\frac{\alpha}{2\hat{\sigma}_0^2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\boldsymbol{u}_i)^\top(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\boldsymbol{u}_i) - \frac{\alpha}{2}\boldsymbol{u}_i^\top\hat{\boldsymbol{D}}^{-1}\boldsymbol{u}_i \right\}$$

and

$$\kappa_\alpha = \frac{1}{(2\pi)^{\frac{(n_i+q)\alpha}{2}}(\hat{\sigma}_0^2)^{\frac{n_i\alpha}{2}}|\hat{\boldsymbol{D}}|^{\frac{\alpha}{2}}}.$$

Differentiating the above equation with respect to $\boldsymbol{u}_i$, $i = 1, \ldots, n$, we obtain the estimating equations for the MDPDE of $\boldsymbol{u}_i$ as

$$\frac{\partial H_n}{\partial \boldsymbol{u}_i} = \frac{(1+\alpha)\kappa_\alpha c_i(\boldsymbol{u}_i)}{n}\left[(\hat{\sigma}_0^2)^{-1}\boldsymbol{Z}_i^\top(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\boldsymbol{u}_i) - \hat{\boldsymbol{D}}^{-1}\boldsymbol{u}_i\right] = 0. \tag{11}$$

The computational aspects about how equations (9), (10) and (11) are solved numerically are treated in subsection 4.3.

Remark The proposed estimating equations given in equation (11) for the random effects $\boldsymbol{u}_i$ can be simplified to

$$\left[(\hat{\sigma}_0^2)^{-1}\boldsymbol{Z}_i^\top(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\boldsymbol{u}_i) - \hat{\boldsymbol{D}}^{-1}\boldsymbol{u}_i\right] = 0$$

since κ_α does not depend on $\boldsymbol{u}_i$ and $c_i(\boldsymbol{u}_i) > 0\ \forall i$. Then, the estimating equation depends on α only through the estimates of the unknown parameters $\hat{\boldsymbol{\beta}}, \hat{\sigma}_0$ and $\hat{\boldsymbol{D}}$. For $\alpha \to 0$, the estimates of the unknown parameters correspond to the MLE, then the random effects predictions $\hat{u}_i$ correspond to the standard predictions of random effects. Hence, they are the *Best Linear Unbiased Predictors* (BLUPs). Our formula differs from the standard formula since we provide estimates of the random effects for each i-th observation, while they are usually reported for each random effect.

4 Theoretical properties of the MDPDE under the Linear mixed models

4.1 Asymptotic efficiency

In this subsection, we prove that the theorem stated in Ghosh and Basu (2013), about the asymptotic behavior of the MDPDE for non-homogeneous observations, holds for the LMM setup. In particular, we present some conditions on the independent variables and the variance–covariance matrices that are used to derive the asymptotic distribution of the estimator in the LMM application.

We assume that the true densities g_i, $i = 1, \ldots, n$, belong to the model family, i.e., $g_i = f_i(\cdot, \boldsymbol{\theta})$ for some value of $\boldsymbol{\theta} \in \Theta$. Consider the following assumptions.

(MM1) Define $\boldsymbol{X}_i' = \left(\frac{\eta_{i\alpha} \boldsymbol{V}_i^{-1}}{(1+\alpha)^{\frac{n_i}{2}+1}} \right)^{\frac{1}{2}} \boldsymbol{X}_i$, for each i, and $\boldsymbol{X}' = \text{Block-Diag}\,(\boldsymbol{X}_i' : i \in \{1, \ldots, n\})$. Then, the $\boldsymbol{X}'$ matrix satisfies

$$\inf_n \left[\text{min eigenvalue of } \frac{\boldsymbol{X}'^{\top} \boldsymbol{X}'}{n} \right] > 0, \tag{12}$$

and $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are full rank matrices for all i.

(MM2) The values of $\boldsymbol{X}_i$'s are such that, for all j, k, l

$$\sup_{n>1} \max_{1 \le i \le n} |\boldsymbol{X}_{ij}^{\top} \boldsymbol{V}_i^{-\frac{1}{2}}| = O(1), \qquad \sup_{n>1} \max_{1 \le i \le n} |\boldsymbol{X}_{ij}^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{X}_{ik}| = O(1), \tag{13}$$

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} |\boldsymbol{X}_{ij}^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{X}_{ik} \boldsymbol{X}_{il}^{\top} \boldsymbol{V}_i^{-\frac{1}{2}} \mathbf{1}| = O(1), \\ &\frac{1}{n} \sum_{i=1}^{n} |\boldsymbol{X}_{ij}^{\top} \boldsymbol{V}_i^{-\frac{1}{2}}| diag(\boldsymbol{V}_i^{-\frac{1}{2}} \boldsymbol{X}_{ik} \boldsymbol{X}_{il}^{\top} \boldsymbol{V}_i^{-\frac{1}{2}}) \mathbf{1} = O(1), \end{aligned} \tag{14}$$

where $\mathbf{1}(n_i \times 1)$ is a vector of 1's.

(MM3) The matrices $\boldsymbol{V}_i$ and $\boldsymbol{U}_{ij}$ are such that, for all $j, k, l \in \{0, \ldots, r\}$,

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{ij}) = O(1), \qquad \frac{1}{n} \sum_{i=1}^{n} \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{ij}) \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{ik}) = O(1), \\ &\frac{1}{n} \sum_{i=1}^{n} \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{ik} \boldsymbol{V}_i^{-1} \boldsymbol{U}_{ij}) = O(1), \end{aligned} \tag{15}$$

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{ij} \boldsymbol{V}_i^{-1} \boldsymbol{U}_{ik} \boldsymbol{V}_i^{-1} \boldsymbol{U}_{il}) = O(1), \\ &\frac{1}{n} \sum_{i=1}^{n} \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{ij} \boldsymbol{V}_i^{-1} \boldsymbol{U}_{ik}) \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{il}) = O(1), \\ &\frac{1}{n} \sum_{i=1}^{n} \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{ij}) \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{ik}) \text{Tr}(\boldsymbol{V}_i^{-1} \boldsymbol{U}_{il}) = O(1), \end{aligned} \tag{16}$$

where the determinant $|\boldsymbol{V}_i|$ is bounded away from both zero and infinity $\forall i$.

(MM4) Define $\boldsymbol{X}_i^* = \left(\frac{\eta_{i\alpha}^2 \boldsymbol{V}_i^{-1}}{(1+2\alpha)^{\frac{n_i}{2}+1}} \right)^{\frac{1}{2}} \boldsymbol{X}_i$, for each i, and $\boldsymbol{X}^* = \text{Block-Diag}\,(\boldsymbol{X}_i^* : i \in \{1, \ldots, n\})$. Then, the $\boldsymbol{X}^*$ matrix satisfies

$$\max_{1 \le i \le n} \left[\frac{(\boldsymbol{X}^{*\top} \boldsymbol{X}^*)^{-1} \boldsymbol{X}_i^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{X}_i}{n} \right] = O(1). \tag{17}$$

Theorem 1 *Consider the setup of the linear mixed model presented in Sect. 3. Assume that the true data generating density belongs to the model family and that the independent variables satisfy Assumptions* (*MM*1)–(*MM*4) *for a given* (*fixed*) $\alpha \geq 0$. *Then, we have the following results as* $n \to \infty$ *keeping* n_i *fixed for each* i.

(i) *There exists a consistent sequence of roots* $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}_n^\top, \hat{\sigma}_{nj}^2, j \in \{0, \ldots, r\})^\top$ *to the minimum DPD estimating equations given in* (9)–(10).

(ii) *The asymptotic distributions of* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\sigma}_j^2$ *are independent for all* $j \in \{0, \ldots, r\}$.

(iii) *The asymptotic distribution of* $\boldsymbol{\Omega}_n^{-\frac{1}{2}} \boldsymbol{\Psi}_n \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ *is p-dimensional normal with mean zero and covariance matrix* $\boldsymbol{I}_p$. *In particular, the asymptotic distribution of* $(\boldsymbol{X}^{*\top}\boldsymbol{X}^*)^{-\frac{1}{2}}(\boldsymbol{X}'^\top \boldsymbol{X}')(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ *is a k-dimensional normal with mean zero and covariance matrix* $\boldsymbol{I}_k$, *where* $\boldsymbol{X}'$ *and* $\boldsymbol{X}^*$ *are as defined in Assumptions* (*MM*1) *and* (*MM*4), *respectively*.

The derivation of matrices $\boldsymbol{\Omega}_n$ and $\boldsymbol{\Psi}_n$ and the proof of Theorem 1 are presented in Section A.1 of Appendix.

4.2 Influence function

To explore the robustness properties of the coefficient estimates in our treatment of linear mixed models, we derive the influence function of the MDPDEs. Denote the density power divergence functional $T_\alpha = (T_\alpha^{\boldsymbol{\beta}}, T_\alpha^{\boldsymbol{\Sigma}})$ for the parameter vector $\boldsymbol{\theta}^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\Sigma} = (\sigma_0^2, \ldots, \sigma_r^2))$. We continue with the notation of the previous subsections.

The influence function of the estimator $T_\alpha^{\boldsymbol{\beta}}$ with contamination at the direction i_0 at the point $\boldsymbol{t}_{i_0}$ is computed to have the form

$$IF_{i_0}(\boldsymbol{t}_{i_0}, T_\alpha^{\boldsymbol{\beta}}, G_1, \ldots, G_n) = (\boldsymbol{X}'^\top \boldsymbol{X}')^{-1} \boldsymbol{X}_{i_0}^\top \boldsymbol{V}_{i_0}^{-1} (\boldsymbol{t}_{i_0} - \boldsymbol{X}_{i_0}\boldsymbol{\beta}) f_{i_0}(\boldsymbol{t}_{i_0}; \boldsymbol{\theta})^\alpha, \tag{18}$$

and the corresponding influence function for the estimator $T_\alpha^{\boldsymbol{\Sigma}}$ has the form

$$IF_{i_0}(\boldsymbol{t}_{i_0}, T_\alpha^{\boldsymbol{\Sigma}}, G_1, \ldots, G_n) = \left[\sum_{i=1}^{n} \frac{\eta_{i\alpha}}{4(1+\alpha)^{\frac{n_i}{2}+2}} T(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}, \boldsymbol{V}_i^{-1}\boldsymbol{U}_{ik}) \right]^{-1} \tau_{i_0}, \tag{19}$$

where

$$\tau_i = \left(\frac{1}{2} f_i(\boldsymbol{t}_i; \boldsymbol{\theta})^\alpha [\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}) - (\boldsymbol{t}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top \boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}\boldsymbol{V}_i^{-1}(\boldsymbol{t}_i - \boldsymbol{X}_i\boldsymbol{\beta})] - \frac{\eta_{i\alpha}\alpha \mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij})}{2(1+\alpha)^{\frac{n_i}{2}+1}} \right).$$

Note that the influence functions in equations (18) and (19), seen as functions of the point $\boldsymbol{t}$, are bounded for any $\alpha > 0$ since they are proportional to the functions $\boldsymbol{z}e^{-\boldsymbol{z}^\top \boldsymbol{z}}$

and $z^\top z e^{-z^\top z}$, respectively, which are bounded for $z \in \mathbb{R}^{n_i}$. For $\alpha = 0$, the influence functions for $T_\alpha^{\boldsymbol{\beta}}$ and $T_\alpha^{\boldsymbol{\Sigma}}$ are seen to be unbounded; indeed, this case corresponds to the non-robust maximum likelihood estimator. Hence, unlike the MLE, the minimum DPD estimators are B-robust, i.e., their associated influence functions are bounded, for $\alpha > 0$.

The influence function of the estimators $T_\alpha^{\boldsymbol{\beta}}$ and $T_\alpha^{\boldsymbol{\Sigma}}$ with contamination in all the n cases at the contamination points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$, can be derived with similar computations and correspond to the sum of equations (18) and (19), respectively, for $i_0 = 1, \ldots, n$. In this case also the influence functions are bounded for $\alpha > 0$ and unbounded for $\alpha = 0$.

Several summary measures of robustness based on the influence function for i.i.d. observations have been introduced in Hampel (1968, 1974). Following the same approaches, some influence function-based gross summary measures can be defined for the non-homogeneous case. For $\alpha > 0$, the gross-error sensitivity and the self-standardized sensitivity of the estimator $T_\alpha^{\boldsymbol{\beta}}$ in the case of contamination only in the i_0-th direction are given by

$$\begin{aligned}\gamma_{i_0}^u(T_\alpha^{\boldsymbol{\beta}}, G_1, \ldots, G_n) &= \sup_{\boldsymbol{t}} \{||IF_{i_0}(\boldsymbol{t}, T_\alpha^{\boldsymbol{\beta}}, G_1, \ldots, G_n)||\} \\ &= \frac{\left[\lambda_{max}\left((\boldsymbol{X}'^\top \boldsymbol{X}')^{-2} \boldsymbol{X}_{i_0}^\top \boldsymbol{V}_{i_0}^{-1} \boldsymbol{X}_{i_0}\right)\right]^{1/2}}{\sqrt{\alpha}(2\pi)^{\frac{n_i \alpha}{2}} |\boldsymbol{V}_i|^{\frac{\alpha}{2}} e^{1/2}}\end{aligned} \tag{20}$$

and

$$\begin{aligned}&\gamma_{i_0}^s(T_\alpha^{\boldsymbol{\beta}}, G_1, \ldots, G_n) \\ &\quad = \sup_{\boldsymbol{t}} \{IF_{i_0}(\boldsymbol{t}, T_\alpha^{\boldsymbol{\beta}}, G_1, \ldots, G_n)^\top (\boldsymbol{\Psi}_n^{-1} \boldsymbol{\Omega}_n \boldsymbol{\Psi}_n^{-1})^{-1} IF_{i_0}(\boldsymbol{t}, T_\alpha^{\boldsymbol{\beta}}, G_1, \ldots, G_n)\}^{\frac{1}{2}} \\ &\quad = \frac{\left[\lambda_{max}\left((\boldsymbol{X}^{*\top} \boldsymbol{X}^*)^{-1} \boldsymbol{X}_{i_0}^\top \boldsymbol{V}_{i_0}^{-1} \boldsymbol{X}_{i_0}\right)\right]^{1/2}}{n\sqrt{\alpha}(2\pi)^{\frac{n_i \alpha}{2}} |\boldsymbol{V}_i|^{\frac{\alpha}{2}} e^{1/2}},\end{aligned} \tag{21}$$

where $\lambda_{max}(\boldsymbol{A})$ indicates the largest eigenvalue of the matrix $\boldsymbol{A}$, while they are equal to ∞ if $\alpha = 0$. Details of the computations are provided in Section SM–2 of the Supplementary Materials. The sensitivity measures for $T_\alpha^{\boldsymbol{\Sigma}}$ have no compact form and they are not reported.

4.3 Computational aspects

The estimating equations (9)–(10) previously introduced can be solved numerically in order to obtain the estimates of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top; \sigma_j^2, j \in \{0, \ldots, r\})^\top$.

Note that denoting $w_i = w_i(\boldsymbol{\beta}, \sigma_j^2)$, $j = 0, \ldots, r$, as defined in equation (8), the estimating equation for $\boldsymbol{\beta}$ given in equation (9) corresponds to

$$-\sum_{i=1}^{n} w_i \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{Y}_i + \sum_{i=1}^{n} w_i \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \boldsymbol{\beta} = 0.$$

Solving for $\boldsymbol{\beta}$, we get

$$\boldsymbol{\beta} = \left(\sum_{i=1}^{n} w_i \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{n} w_i \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{Y}_i \right),$$

so that, in an iterative fixed point algorithm, the successive iterates have the relation

$$\boldsymbol{\beta}^{(k+1)} = \left(\sum_{i=1}^{n} w_i(\boldsymbol{\beta}^{(k)}, \sigma_j^{(k)}) \boldsymbol{X}_i^\top \boldsymbol{V}_i (\sigma_j^{(k)})^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{n} w_i(\boldsymbol{\beta}^{(k)}, \sigma_j^{(k)}) \boldsymbol{X}_i^\top \boldsymbol{V}_i (\sigma_j^{(k)})^{-1} \boldsymbol{Y}_i \right).$$

The estimating equation (10) for the variance components cannot be written in a similar closed form, then it is solved numerically by a quasi-Newton method which allows box constraints, that is, each variable can be given a lower and/or upper bound.

Finally, the random coefficients can be predicted through the obtained estimates. Notice that the quantity κ_α can be removed to solve equation (11) since it does not depend on $\boldsymbol{u}_i$, as well as $c_i(\boldsymbol{u}_i)$ since $c_i(\boldsymbol{u}_i) > 0$ for all i. Hence, equation (11) is simplified to

$$(\hat{\sigma}_0^2)^{-1} \boldsymbol{Z}_i^\top (\boldsymbol{Y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}) - \left((\hat{\sigma}_0^2)^{-1} \boldsymbol{Z}_i^\top \boldsymbol{Z}_i + \hat{\boldsymbol{D}}^{-1} \right) \boldsymbol{u}_i = 0.$$

Then,

$$\hat{u}_i = \left((\hat{\sigma}_0^2)^{-1} \boldsymbol{Z}_i^\top \boldsymbol{Z}_i + \hat{\boldsymbol{D}}^{-1} \right)^{-1} (\hat{\sigma}_0^2)^{-1} \boldsymbol{Z}_i^\top (\boldsymbol{Y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}). \tag{22}$$

This formulation provides a closed form to predict the realizations $\boldsymbol{u}_i$, $i = 1, \ldots, n$. It is worth noting that even if these estimates may appear not robust, in the sense that an outliers $(\boldsymbol{Y}_i, \boldsymbol{X}_i)$ can affect the predicted $\boldsymbol{u}_i$, they are based on the robust estimates $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_0^2$ and $\hat{\boldsymbol{D}}$.

4.4 An example: the balanced data case

Consider the model defined by Equation (6). Here, we study the simplest case in which $n_i = p$, for all $i \in \{1, \ldots, n\}$, and the associated random effects covariates $(\boldsymbol{Z}_{ij})$ are also the same for all i. In this case, the covariance matrix of $\boldsymbol{Y}_i$ is the same for all i and is denoted by $\boldsymbol{V}$ having the form

$$\boldsymbol{V} = \boldsymbol{V}_i = \sigma_0^2 \left(\boldsymbol{I}_p + \sum_{j=1}^{r} \boldsymbol{U}_j \gamma_j \right),$$

where $\gamma_j = \sigma_j^2 / \sigma_0^2$ and $\boldsymbol{U}_j = \boldsymbol{U}_{ij}$ as it is independent of i.

The influence function of the functional $T_\alpha^{\boldsymbol{\beta}}$ with contamination in the direction i_0, given in Equation (18), can be written as

$$IF_{i_0}(\boldsymbol{t}_{i_0}, T_\alpha^{\boldsymbol{\beta}}, G_1, \dots, G_n) = (1+\alpha)^{\frac{p}{2}+1}\left(\sum_{i=1}^{n} \boldsymbol{X}_i^\top \boldsymbol{V}^{-1}\boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_{i_0}^\top \boldsymbol{V}^{-1}(\boldsymbol{t}_{i_0} - \boldsymbol{X}_{i_0}\boldsymbol{\beta})w_{i_0}.$$

Using this expression, the gross-error sensitivity for the functional $T_\alpha^{\boldsymbol{\beta}}$ is given by

$$\gamma_{i_0}^u(T_\alpha^{\boldsymbol{\beta}}) = \frac{(1+\alpha)^{\frac{p}{2}+1}}{\sqrt{\alpha}}\left[\lambda_{max}\left(\left[\sum_{i=1}^{n} \boldsymbol{X}_i^\top \boldsymbol{V}^{-1}\boldsymbol{X}_i\right]^{-2} \boldsymbol{X}_{i_0}^\top \boldsymbol{V}^{-1}\boldsymbol{X}_{i_0}\right)\right]^{1/2} e^{-1/2}. \tag{23}$$

Similarly, the self-standardized sensitivity of the functional $T_\alpha^{\boldsymbol{\beta}}$ can be written as

$$\gamma_{i_0}^s(T_\alpha^{\boldsymbol{\beta}}) = \frac{(1+\alpha)^{\frac{p+2}{4}}}{n\sqrt{\alpha}}\left[\lambda_{max}\left(\left[\sum_{i=1}^{n} \boldsymbol{X}_i^\top \boldsymbol{V}^{-1}\boldsymbol{X}_i\right]^{-1} \boldsymbol{X}_{i_0}^\top \boldsymbol{V}_{i_0}^{-1}\boldsymbol{X}_{i_0}\right)\right]^{1/2} (2e)^{-1/2}. \tag{24}$$

The simpler form of the sensitivity measures allows us to assess their performance with respect to the tuning parameter α. Indeed, the function $\frac{(1+\alpha)^{\frac{p}{2}+1}}{\sqrt{\alpha}}$ in the gross-error sensitivity (23) has a minimum for the value $\alpha^* = \frac{1}{p+1}$, suggesting that this value of the parameter α gives the most robust estimator. Similarly, the function $\frac{(1+\alpha)^{\frac{p+2}{4}}}{\sqrt{\alpha}}$ in the self-standardized sensitivity (24) has a minimum for the value $\bar{\alpha} = \frac{2}{p}$. The existence of such minimum values for α is in contrast with the previously held knowledge about this parameter. It was introduced as a trade-off between efficiency and robustness, instead here we show that, passing some threshold, with increasing α , we lose both efficiency and robustness. Finally, the proposed optimal values α^* and $\bar{\alpha}$ depend only on the dimension of observations and constitute valuable choices in practical situations.

In the following, we present a simple example for which we will compute the theoretical quantities introduced above. This example in linear mixed models has been chosen for its similarity to the case of longitudinal data; it is often also named as LMM with random intercept and random slope.

Notice that the MDPDE satisfies the equivariance properties. In particular, the regression equivariance allows us to assume, without loss of generality, any suitable value for the parameter $\boldsymbol{\beta}$ while proving the asymptotic properties with the following example or for the Monte Carlo studies. Since the MDPDE is also scale and affine equivariant, such estimators do not depends on the choice of the coordinate system for the variables $\boldsymbol{x}$ and on the measurement unit of $\boldsymbol{y}$. The derivation of these properties is reported in Section SM–3 of the Supplementary Material. Furthermore, Section SM–4 reports an illustrative example which shows the robustness achieved by the random effects predictions computed according to equation (22).

We consider $n = 50$ different subjects (groups), and for each of them, we have $p = 10$ measurements taken with respect to the factor u_{i2}, $i = 1, \ldots, n$, with two levels, modeled here as a random effect. The $\boldsymbol{X}$'s model matrices are simulated from a standard normal. In particular, the model is described by

$$\boldsymbol{Y}_i = \beta_0 + \beta_1 \boldsymbol{X}_i + \boldsymbol{\mathcal{U}}_1 + \boldsymbol{\mathcal{U}}_2 \boldsymbol{Z}_{i2} + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{\mathcal{U}}_1 \sim N_p(0, \sigma_1^2 \boldsymbol{I}_p)$, $\boldsymbol{\mathcal{U}}_2 \sim N_2(0, \sigma_2^2 \boldsymbol{I}_2)$ and $\epsilon_i \sim N_p(0, \sigma_0^2 \boldsymbol{I}_p)$, and they are independent. Hence, for this model, $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma_1^2, \sigma_2^2, \sigma_0^2)$, and we take $\boldsymbol{\theta} = (1, 2, 0.25, 0.5, 0.25)$ as the true values of the parameters.

Using the given values, we compute the variance–covariance matrices $\boldsymbol{V}_i$ and the matrices $\boldsymbol{\Psi}_n$ and $\boldsymbol{\Omega}_n$. First, we will look at the *Asymptotic Relative Efficiency* (ARE) of the minimum density power divergence estimators with respect to the fully efficient maximum likelihood estimator. Figure 1 shows the asymptotic relative efficiencies of the estimators of β_1 and σ_2^2 for $\alpha \in [0, 0.6]$. It is easy to see that there is a loss of efficiency which increases with α. However, for small positive values of α, the estimator retains reasonable efficiency. The ARE of the estimators of the other parameters are similar to those displayed here, and are given in Section SM–5 of the Supplementary Materials.

On the other hand, to study the robustness properties, Fig. 2 shows the influence functions of $T_\alpha^{\beta_0}$ and $T_\alpha^{\sigma_1^2}$, with respect to $\alpha = 0, 0.05, \alpha^*, \bar{\alpha}$ where $\alpha^* = 1/(p+1) = 1/11$ and $\bar{\alpha} = 2/p = 0.2$. Here, we have plotted $IF(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n, T_\alpha, G_1, \ldots, G_n)$, the influence function of the estimator T_α, computed with respect to constant vectors $\underline{\boldsymbol{t}} = t(1, \ldots, 1)^\top$ for varying $t \in \mathbb{R}$. Note that except for the case $\alpha = 0$, we can easily see that the influence function is bounded as may also be noted from equations (18) and (19); thus, the estimator will be robust with respect to outliers. The influence function for the estimators of other parameters behaves similarly; these plots are available in Section SM–5 of the Supplementary Materials.

Finally, Fig. 3 shows the gross-error sensitivity and the self-standardized sensitivity of the functional T_α^{β}. Here, we have considered a particular direction $i_0 \in \{1, \ldots, n\}$. Note that in the present case of balanced data, the choice of i_0 does not change the behavior of the sensitivity measures with respect to α.

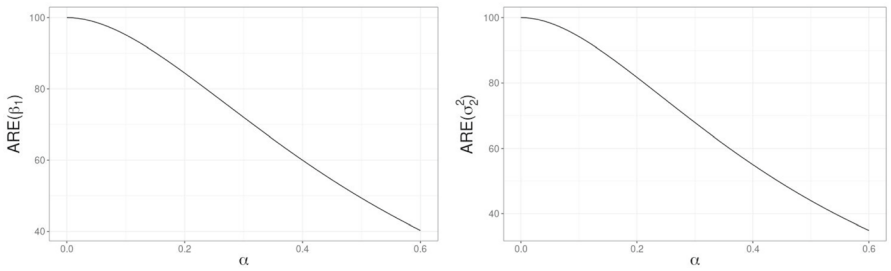
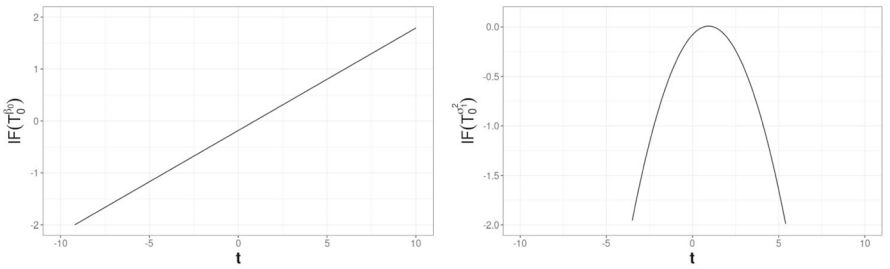
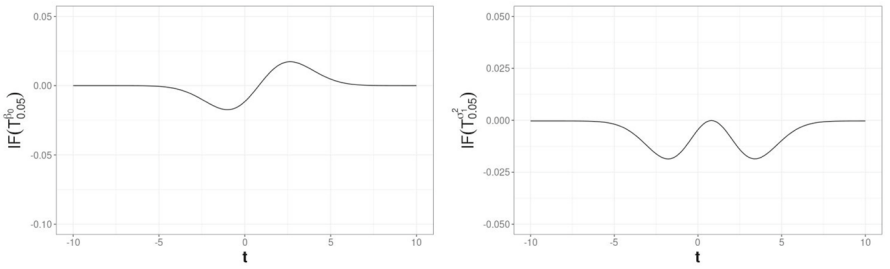


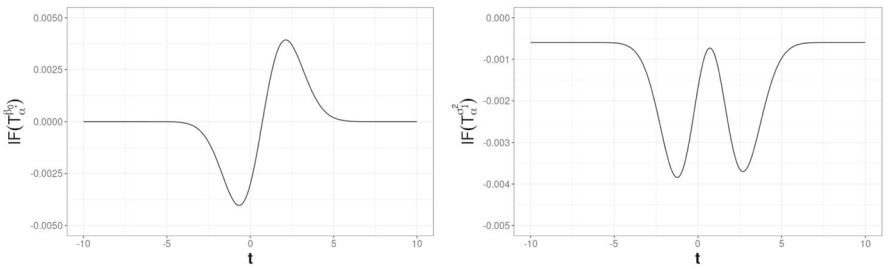
Fig. 1 Asymptotic relative efficiency with respect to α for β_1 and σ_2^2, respectively



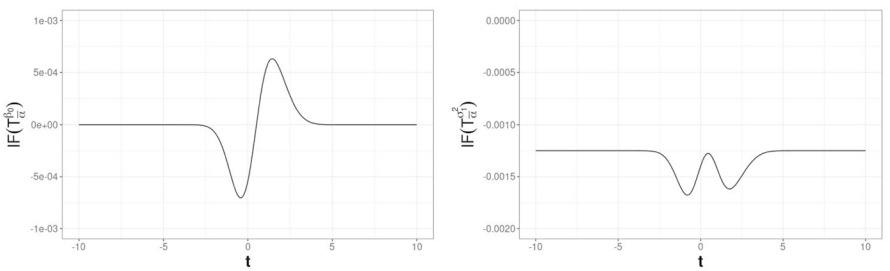
(a) $\alpha = 0$



(b) $\alpha = 0.05$



(c) $\alpha = \alpha^* = \frac{1}{p+1} = \frac{1}{11}$



(d) $\alpha = \bar{\alpha} = \frac{2}{p} = 0.2$

Fig. 2 Influence function for the functionals $T_\alpha^{\beta_0}$(left panel) and $T_\alpha^{\sigma_1^2}$ (right panel), for different values of the tuning parameter α

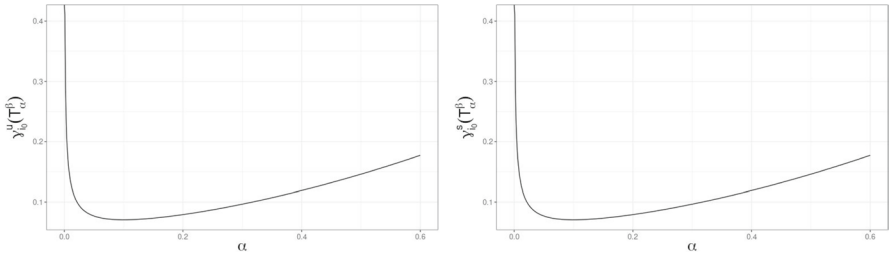


Fig. 3 Gross-error sensitivity (left panel) and self-standardized sensitivity (right panel) of the functional T_α^β with respect to $i_0 = 10$

5 Monte Carlo simulations

5.1 Model setting

We consider a simulation setting as introduced in Agostinelli and Yohai (2016) and reported here in order to facilitate the comparison of the considered estimators.

Consider an LMM for a two-way cross classification with interaction, where the model is given by

$$y_{fgh} = \boldsymbol{x}_{fgh}^\top \boldsymbol{\beta}_0 + a_f + b_g + c_{fg} + e_{fgh},$$

where $f = 1, \ldots, F, g = 1, \ldots, G$, and $h = 1, \ldots, H$. Here, we set $F = 2, G = 2$ and $H = 3$ getting $p = F \times G \times H = 12$. Also $\boldsymbol{x}_{fgh}$ is a $k \times 1$ vector where the last $k - 1$ components are from a standard multivariate normal and the first component is identically equal to 1, and $\boldsymbol{\beta}_0 = (0, 2, 2, 2, 2, 2)^\top$ is a $k \times 1$ vector of the fixed parameters with $k = 6$. The random variables a_f, b_g and c_{fg} are the random effects which are normally distributed with variances σ_a^2, σ_b^2, and σ_c^2. Arranging the y_{fgh} in lexicon order (ordered by h within g within f), we obtain the vector $\boldsymbol{y}$ of dimension p, and in the similar way, the $p \times k$ matrix $\boldsymbol{x}$ obtained arranging $\boldsymbol{x}_{fgh}$. Similarly, we set $\boldsymbol{a} = (a_1, \ldots, a_F)^\top$, $\boldsymbol{b} = (b_1, \ldots, b_G)^\top$ and $\boldsymbol{c} = (c_{11}, \ldots, c_{FG})^\top$, that is, $\boldsymbol{a} \sim N_F(\boldsymbol{0}, \sigma_a^2 \boldsymbol{I}_F)$ and similarly for $\boldsymbol{b}$ and $\boldsymbol{c}$, while $\boldsymbol{e} = (e_{111}, \ldots, e_{FGH})^\top \sim N_p(\boldsymbol{0}, \sigma_e^2 \boldsymbol{I}_p)$. Hence, $\boldsymbol{y}$ is a p multivariate normal with mean $\boldsymbol{\mu} = \boldsymbol{x}\boldsymbol{\beta}$ and variance matrix $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0) = \eta_0(\boldsymbol{V}_0 + \sum_{j=0}^J \gamma_j \boldsymbol{V}_j)$, where $\boldsymbol{V}_0 = \boldsymbol{I}_p$, $\boldsymbol{V}_1 = \boldsymbol{I}_F \otimes \boldsymbol{J}_G \otimes \boldsymbol{J}_H$, $\boldsymbol{V}_2 = \boldsymbol{J}_F \otimes \boldsymbol{I}_G \otimes \boldsymbol{J}_H$, and $\boldsymbol{V}_3 = \boldsymbol{I}_F \otimes \boldsymbol{I}_G \otimes \boldsymbol{J}_H$; $\otimes$ is the Kronecker product and $\boldsymbol{J}_k$ is a $k \times k$ matrix of ones. We took $\sigma_a^2 = \sigma_b^2 = 1/16$ and $\sigma_c^2 = 1/8$. Then, $\boldsymbol{\gamma_0} = (\gamma_{01}, \gamma_{02}, \gamma_{03})^\top = (\sigma_a^2/\sigma_e^2, \sigma_b^2/\sigma_e^2, \sigma_c^2/\sigma_e^2)^\top = (1/4, 1/4, 1/2)^\top$ and $\eta_0 = \sigma_e^2 = 1/4$.

We consider a sample of size $n = 100$ and four levels of contamination $\varepsilon = 0, 5, 10$ and 15%. Hence, $n \times \varepsilon$ observations are contaminated according the following contamination scenarios. Let $\boldsymbol{y}_0$ and $\boldsymbol{x}_0$ indicate the response vector and the fixed effect model matrix for the contaminated observations.

- Complete contamination: $n \times \varepsilon$ elements of the vector $\boldsymbol{y}$ are replaced by observations from $\boldsymbol{y}_0 \sim N_p(\boldsymbol{x}_0\boldsymbol{\beta}_0 + \boldsymbol{\omega}_0, \boldsymbol{\Sigma})$. The matrix $\boldsymbol{x}_0$ is such that the

first column is identically equal to 1, while the last $k-1$ columns are from $N_{p\times(k-1)}(\boldsymbol{\phi}_0, 0.005^2\boldsymbol{I}_{p\times(k-1)})$ where $\boldsymbol{\phi}_0$ indicates a p-vector of constants all equal to ϕ_0 with $\phi_0 = 1, 20$ in the case of low leverage outliers (lev1) or for large leverage outliers (lev20), respectively. $\boldsymbol{\omega}_0$ is a p-vector of constants all equal to ω_0 with $\omega_0 = 0, 1, \ldots, 30$.

- Contamination only on the response $\boldsymbol{y}$: $\boldsymbol{y}_0 \sim N_p(\boldsymbol{x}\boldsymbol{\beta}_0 + \boldsymbol{\omega}_0, \boldsymbol{\Sigma})$ with ω_0 as above.
- Contamination only on the $\boldsymbol{x}$: $n \times \varepsilon$ rows of $\boldsymbol{x}$ are replaced by $\boldsymbol{x}_0$, as defined in the first scenario, with $\phi_0 = 0, 1, 3, 5, 10, 15, 20, 25, 30$. The corresponding elements of $\boldsymbol{y}$ are replaced by observations sampled from $\boldsymbol{y}_0 \sim N_p(\boldsymbol{x}_0\boldsymbol{\beta}_0, \boldsymbol{\Sigma})$.

For each contamination scenario, we compute the CVFS-estimator described in Copt and Victoria-Feser (2006) with Rocke ρ function and with asymptotic rejection probability set to 0.01 as implemented in the `R` R Core Team (2019) package `robustvarComp` (Agostinelli and Yohai (2019)), the SMDM estimator introduced by Koller (2013) as implemented in the `R` package `robustlmm` (Koller (2016)), the composite τ-estimator proposed by Agostinelli and Yohai (2016) and available in the `R` package `robustvarComp`, and our proposed MDPDE with different choices of α. In particular, $\alpha \in \{0, 0.01, 0.1, 0.2, \ldots, 1\}$; note that $\alpha^* = 1/(p+1) = 1/13$ and $\bar{\alpha} = 2/p = 1/6$. For each case, we run 500 Monte Carlo replications.

5.2 Performance Measures

Let $(\boldsymbol{y}, \boldsymbol{x})$ be an observation independent of the sample $(\boldsymbol{y}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_n, \boldsymbol{x}_n)$ used to compute $\widehat{\boldsymbol{\beta}}$ and let $\widehat{\boldsymbol{y}} = \boldsymbol{x}\widehat{\boldsymbol{\beta}}$ be the predicted value of $\boldsymbol{y}$ using $\boldsymbol{x}$. Then, the squared Mahalanobis distance between $\widehat{\boldsymbol{y}}$ and $\boldsymbol{y}$ using the matrix $\boldsymbol{\Sigma}_0$ is

$$\begin{aligned} m(\widehat{\boldsymbol{y}}, \boldsymbol{y}, \boldsymbol{\Sigma}_0) &= (\widehat{\boldsymbol{y}} - \boldsymbol{y})^\top \boldsymbol{\Sigma}_0^{-1} (\widehat{\boldsymbol{y}} - \boldsymbol{y}) \\ &= (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \boldsymbol{x}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{x} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0). \end{aligned}$$

Since $\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0$ is independent of $\boldsymbol{x}$ and has covariance matrix $\boldsymbol{\Sigma}_0$, putting $\boldsymbol{A} = \mathbb{E}(\boldsymbol{x}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{x})$, we have

$$\begin{aligned} \mathbb{E}\left[m(\widehat{\boldsymbol{y}}, \boldsymbol{y}, \boldsymbol{\Sigma}_0)\right] &= \mathbb{E}\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \boldsymbol{A} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] + \operatorname{trace}\left[\boldsymbol{\Sigma}_0^{-1} (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0)(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0)^\top\right] \\ &= \mathbb{E}\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \boldsymbol{A} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] + p. \end{aligned}$$

Then, to evaluate an estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ by its prediction performance, we can use

$$\mathbb{E}\left[m(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0, \boldsymbol{A})\right] = \mathbb{E}\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \boldsymbol{A} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right].$$

Let N be the number of replications in the simulation study, and let $\widehat{\boldsymbol{\beta}}_j$, $1 \le j \le N$ be the value of $\widehat{\boldsymbol{\beta}}$ at the j-th replication, then we can estimate $\mathbb{E}\left[m(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0, \boldsymbol{A})\right]$ by the mean square Mahalanobis distance as

$$\text{MSMD} = \frac{1}{N}\sum_{j=1}^{N} m(\widehat{\boldsymbol{\beta}}_j, \boldsymbol{\beta}_0, \boldsymbol{A}).$$

It is easy to prove that as in this case, $\boldsymbol{x}$ is a $p \times k$ matrix where the cells are independent $N(0, 1)$ random variables, then $\boldsymbol{A} = \text{trace}(\boldsymbol{\Sigma}_0^{-1})\boldsymbol{I}_k$.

Given two p-dimensional covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_0$, one way to measure how close $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_0$ are is through the use of the Kullback–Leibler divergence between two multivariate normal distributions with the same mean and covariance matrices equal to $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_0$, given by

$$\text{KLD}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_0) = \text{trace}\left(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}\right) - \log\left(\det(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1})\right) - p.$$

Since $(\eta_0, \boldsymbol{\gamma}_0)$ determines $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0)$, the covariance matrix of $\boldsymbol{y}$ given $\boldsymbol{x}$ for the particular LMM considered in our simulation (as described in Sect. 5.1), one way to measure the performance of an estimator $(\widehat{\eta}, \widehat{\boldsymbol{\gamma}})$ of $(\eta_0, \boldsymbol{\gamma}_0)$ is by $\mathbb{E}\left[\text{KLD}(\boldsymbol{\Sigma}(\widehat{\eta}, \widehat{\boldsymbol{\gamma}}), \boldsymbol{\Sigma}_0)\right]$. Let $(\widehat{\eta}_j, \widehat{\boldsymbol{\gamma}}_j)$, $1 \le j \le N$, be the value of $(\widehat{\eta}, \widehat{\boldsymbol{\gamma}})$ at the j-th replication, then we can estimate $\mathbb{E}\left[\text{KLD}(\boldsymbol{\Sigma}(\widehat{\eta}, \widehat{\boldsymbol{\gamma}}), \boldsymbol{\Sigma}_0)\right]$ by the mean Kullback–Leibler divergence

$$\text{MKLD} = \frac{1}{N}\sum_{j=1}^{N} \text{KLD}(\boldsymbol{\Sigma}(\widehat{\eta}_j, \widehat{\boldsymbol{\gamma}}_j), \boldsymbol{\Sigma}_0).$$

5.3 Results

We begin with the performance of the estimators in the absence of contamination. Table 1 shows the relative efficiency of the CVFS-estimator, the SMDM-estimator, the Composite τ-estimator and the MDPDE for different values of α with respect to maximum likelihood. The efficiency of the estimators of $\boldsymbol{\beta}$ has been measured by the

Table 1 Relative efficiency for the SMDM-estimator, CVFS-estimator, Composite τ -estimator and MDPDE for different values of α with respect to the maximum likelihood computed by the MSMD for the fixed terms $\boldsymbol{\beta}$ and by the MKLD for the random terms

Method	(α)	MSMD EFF.	MKLD EFF.
SMDM	–	0.956	0.147
CVFS	–	0.706	0.453
Composite τ	–	0.807	0.833
MDPDE	0.01	0.999	0.996
	α^*	0.960	0.945
	0.1	0.937	0.915
	$\bar{\alpha}$	0.853	0.814
	0.2	0.805	0.760
	0.3	0.658	0.603
	0.4	0.519	0.470
	0.5	0.400	0.361
	0.6	0.302	0.273

MSMD ratio, while the MKLD ratio was used for the efficiency of an estimator of (η, γ).

The MDPDEs exhibit a high relative efficiency, even greater than the competitor estimators, for small values of α, while the efficiency decreases with increasing α. Note that the MDPDEs are far more successful in retaining the efficiency of the estimators of the random components. For very small values of α , the MDPDEs dominate either competitor (at least up to $\alpha = \alpha^*$ for SMDM, and at least up to $\alpha = 0.2$ for CVFS) in terms of both (MSMD and MKLD) efficiency measures. As the value of α increases, the MSMD efficiency of the MDPDE eventually lags behind its competitors, but in terms of MKLD efficiency, it beats the competitors at least up to $\alpha = 0.4$, except for the Composite τ-estimator that has a higher efficiency than the MDPDE for $\alpha > \bar{\alpha}$. On the whole, it is clear that under pure data, a properly chosen member of the MDPDE class can perform competitively, if not better, compared to the SMDM, the CVFS and the Composite τ -estimators.

Now, we consider the contamination settings. At first, Tables 2 and 3 report the maximum values of MSMD and MKLD over the values of ω_0 and leverage ϕ_0 considered for the complete contamination and separated contamination, respectively, of the MDPDE for different values of α compared to the CVFS-, the SMDM- and Composite τ-estimators.

Small values of α, as expected, provide much higher maximum values with respect to the other estimators in Table 2. However, for slightly larger values of α, the MDPDEs are extremely competitive with the existing estimators. It may be easily observed that in case of complete contamination, the MDPDE at $\bar{\alpha}$ clearly beats the competitors (CVFS, SMDM and Composite τ) over both performance measures at both leverage values (except in case of complete contamination at MSMD, lev20, where its performance measure is equal to that of CVFS). In this example, the MDPDE at $\alpha = 0.2$ fares even better. In case of contamination on the $\boldsymbol{y}$, the

Table 2 Complete contamination. Maximum values of MSMD and MKLD for the CVFS-, SMDM-, Composite τ-estimators and for the MDPDE at different values of α under 10% of outlier contamination

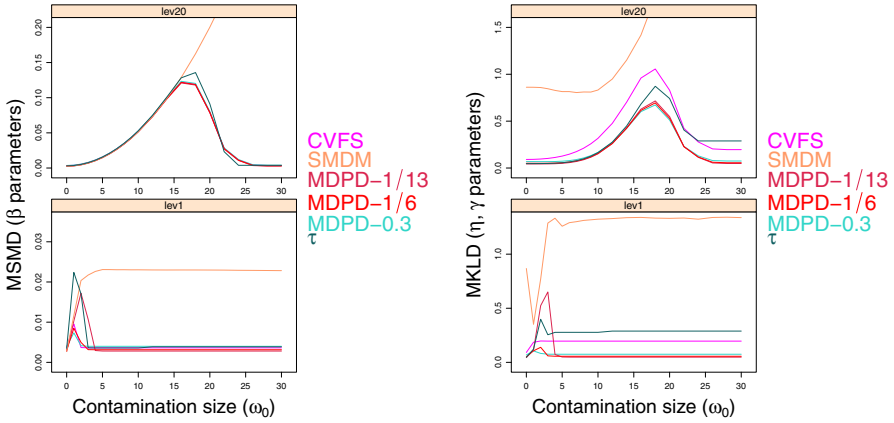
		MSMD		MKLD	
Method	(α)	lev1	lev20	lev1	lev20
CVFS	–	0.010	0.122	0.197	1.057
SMDM	–	0.023	0.450	1.341	9.125
Composite τ	–	0.022	0.136	0.398	0.872
MDPDE	0	9.007	9.005	3.508e22	1.605e25
	0.01	1.114	0.120	106.185	0.732
	$\alpha^*(\frac{1}{13})$	0.017	0.121	0.650	0.716
	0.1	0.012	0.121	0.387	0.710
	$\bar{\alpha}(\frac{1}{6})$	0.008	0.122	0.139	0.695
	0.2	0.008	0.122	0.105	0.688
	0.3	0.007	0.123	0.106	0.673
	0.4	0.007	0.125	0.116	0.665
	0.5	0.008	0.127	0.137	0.662
	0.6	0.010	0.130	0.170	0.666

Table 3 Contamination on $\boldsymbol{x}$ and $\boldsymbol{y}$ separately. Maximum values of MSMD and MKLD for the CVFS-, SMDM-, Composite τ-estimators and for MDPDE considering different values of α under 10% of outlier contamination

Method	(α)	MSMD		MKLD	
		x	y	x	y
CVFS	–	0.004	0.008	0.092	0.210
SMDM	–	0.003	0.021	0.820	1.667
Composite τ	–	0.004	0.010	0.042	0.305
MDPDE	0	0.003	8.998	1.012	203.248
	0.01	0.003	1.134	0.035	109.762
	$\alpha^*(\frac{1}{13})$	0.003	0.016	0.036	0.609
	0.1	0.003	0.011	0.037	0.389
	$\tilde{\alpha}(\frac{1}{6})$	0.003	0.008	0.042	0.134
	0.2	0.003	0.007	0.046	0.109
	0.3	0.004	0.007	0.060	0.108
	0.4	0.005	0.007	0.081	0.120
	0.5	0.007	0.008	0.108	0.145
	0.6	0.009	0.009	0.145	0.185

estimators show similar performance, while, when only the $\boldsymbol{x}$ is contaminated, the MSMD values are almost equal for all the considered estimators suggesting that the estimation of the fixed effects parameter $\boldsymbol{\beta}$ is not affected. However, as in the previous case, the MDPDE for values of α close to zero show larger MKLD values, but slightly increasing α the MDPDE outperforms the competitors.

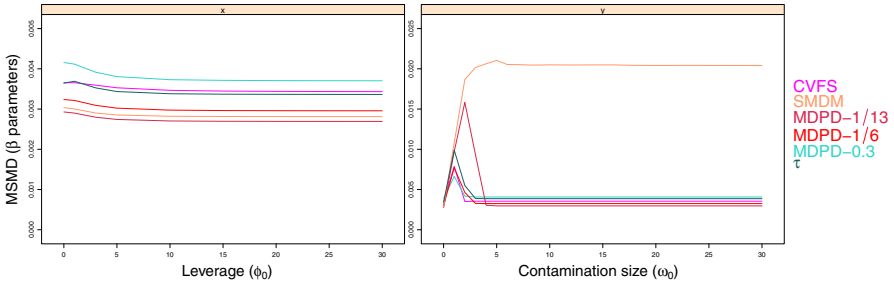
Figures 4a and 4b display the MSMD and MKLD as function of ω_0, comparing the CVFS-, SMDM- and Composite τ-estimators with the MDPDEs for three chosen values of α, under 10% of outlier contamination. In particular, we choose



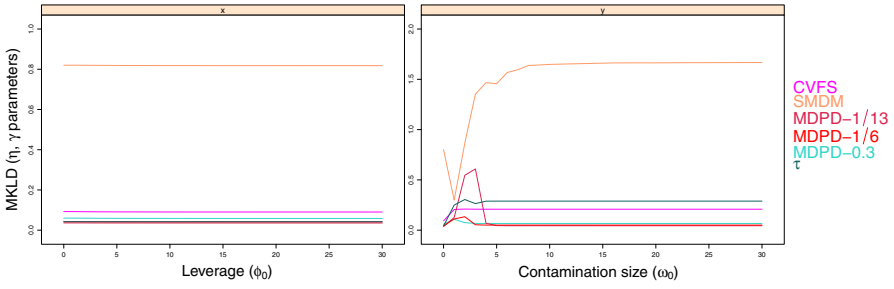
(a) MSMD performance of the estimators of $\boldsymbol{\beta}$ (b) MKLD performance of the estimators of $(\eta, \boldsymbol{\gamma})$

Fig. 4 Complete contamination. Performance of the MDPD-estimators of $\boldsymbol{\beta}$ and $(\eta, \boldsymbol{\gamma})$ for $\alpha = \frac{1}{13}, \frac{1}{6}, 0.3$, compared to the CVFS-, SMDM- and Composite τ-estimators, under 10% outlier contamination

α^* and $\bar{\alpha}$ since they are the values suggested by theory, and $\alpha = 0.3$ since it shows the lowest (or very close to the lowest) maximum values of MSMD and MKLD. We can see that most of the MDPDEs outperform the CVFS- and SMDM-estimators, especially in case of leverage 20 (lev20), where the SMDM-estimator shows an unbounded behavior. On the other hand, in the case of leverage 1 (lev1), even if the CVFS-estimator presents lower maximum value of MSMD and MKLD for very small values of ω_0, the MDPDEs show a better performance when ω_0 increases. In fact, the MDPDE at $\alpha = 0.3$ is competitive or better than CVFS at all values of ω_0. Finally, Fig. 5a and 5b shows the MSMD and MKLD as function of ω_0 if the response $\boldsymbol{y}$ is contaminated and as function of ϕ_0 in case of contamination on the $\boldsymbol{x}$. The CVFS-, SMDM- and Composite τ-estimators are compared with the MDPDEs for three chosen values of α, under 10% of outlier contamination. These results confirm that the MDPDEs, at least for one value of α, outperform the competitor estimators. Complete results of the simulation study, for all the considered estimators and levels of contamination, are reported in Section SM–6 of the Supplementary Materials.



(a) MSMD performance of the estimators of $\boldsymbol{\beta}$



(b) MKLD performance of the estimators of $(\eta, \boldsymbol{\gamma})$

Fig. 5 Contamination on $\boldsymbol{x}$ and $\boldsymbol{y}$. Performance of the MDPD-estimators of $\boldsymbol{\beta}$ and $(\eta, \boldsymbol{\gamma})$ for $\alpha = \frac{1}{13}, \frac{1}{6}, 0.3$, compared to the CVFS-, SMDM- and Composite τ-estimators, under 10% outlier contamination

6 Real-data example: Orthodontic Distance Growth

Let us now present an application of the proposed estimation method to a real-data example on orthodontic measures, while Section SM–8 of the Supplementary Materials reports the analysis of the real-life data on foveal and extrafoveal vision acuity (and crowding) studying their interrelationships with one's reading performances. We compare the estimates obtained by the minimum DPD method with those obtained using the classical (non-robust) restricted MLE, computed using the `lmer` function in `R`, as well as the robust competitors, that is, the SMDM-estimator, the CVFS-estimator and the Composite τ-estimator. A very important consideration in real situations is the selection of an "optimum" value of α that applies to the given data set. We will consider the values α^* and $\bar{\alpha}$, derived from theoretical computations, since they are the suggested optimal values.

We consider an orthodontic study conducted by Potthoff and Roy (1964) where the distance (in millimeters) between the pituitary and the pterygomaxillary fissure has been measured on 16 boys and 11 girls at 8, 10, 12, and 14 years. The data set is available as part of the `R` package `nlme` (Pinheiro et al. 2022). Figure 6 displays the measurements for each individual in the study together with the least-squares fit of the simple linear regression model.

From Fig. 6, it is possible to see that the data set possibly contains some outliers. In particular, the measurements for subject `M09` have more variability around the fitted line with two possible within-subject outliers and the slope for subject `M13` is larger than the others indicating a possible outlier at the level of random effects. Finally, subject `M10` could be also considered as outlying observation for the large distance values since the first measurement. Overall, the intercept and the slope seem to vary with the subject, and the responses for the girls show less variation around the fitted lines than for boys.

According to the mentioned features, the orthodontic distance growth with respect to age can be modeled using the linear mixed model of the form

$$y_{ij} = \beta_0 + \beta_1 I_i(F) + (\beta_2 + \beta_3 I_i(F))t_j + u_{i1} + u_{i2} t_j + \epsilon_{ij}$$

for $i = 1, \ldots, 27$ and $j = 1, \ldots, 4$, where y_{ij} denotes the distance for subject i at age t_j; β_0 and β_1 represent the fixed effect intercept for boys and girls, respectively; β_2 and β_3 represent the fixed effect slope for boys and girls, respectively; $I_i(F)$ indicates an indicator function for the girls group; $(\boldsymbol{\mathcal{U}}_1, \boldsymbol{\mathcal{U}}_2)$ is the vector of random terms and $(\boldsymbol{u}_{i1}, \boldsymbol{u}_{i2})$ is the vector of realizations for subject i; ϵ_{ij} is the error term. Notice that in this example, the random effect for the age is nested into the random effect for the subject level, so a possible correlation between them has to be considered. According to this, the variance components to be estimated are σ_j^2, $j = 0, 1, 2$ as previously described and the covariance σ_{12} between the random variables $\boldsymbol{\mathcal{U}}_1$ and $\boldsymbol{\mathcal{U}}_2$.

Table 4 and Table 5 report the estimates for the fixed terms and for the variance components, respectively, obtained by the MLE, the SMDM-estimator, the CVFS-estimator, the Composite τ-estimator and MDPDE for $\alpha = \alpha^* = 0.2$ and $\alpha = \bar{\alpha} = 0.5$. Furthermore, we computed the maximum likelihood estimates for the reduced data set obtained removing the observations `M09`, `M10` and `M13`, indicated

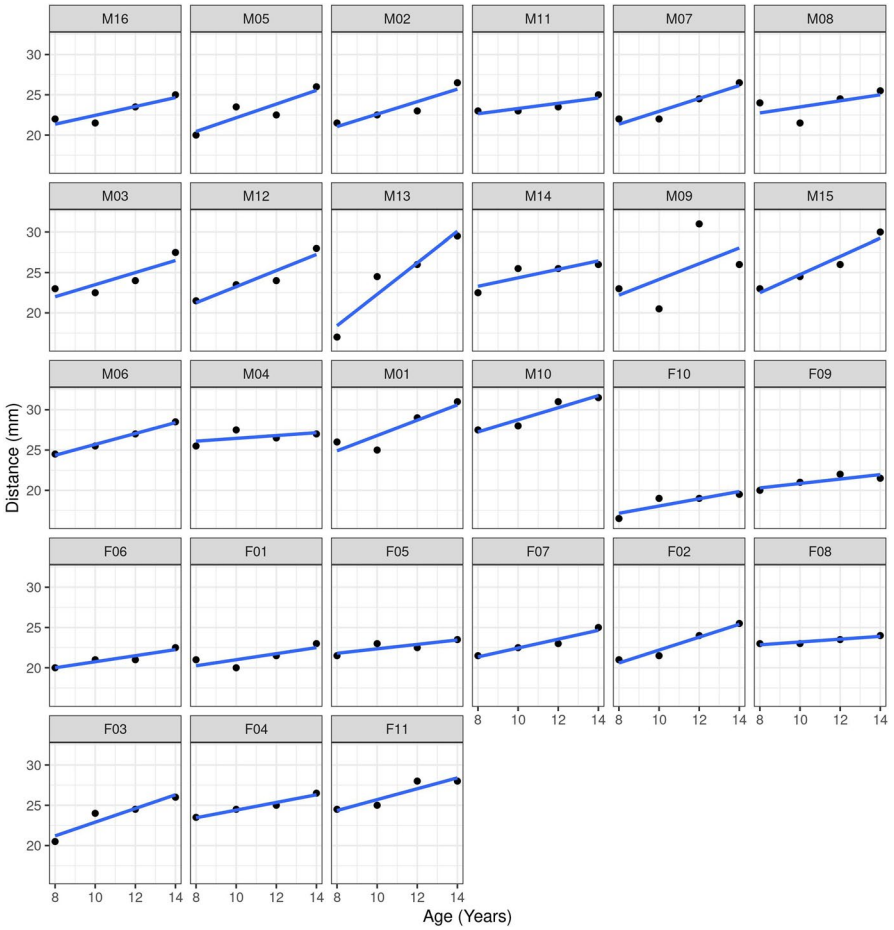


Fig. 6 Orthodontic growth patterns in 11 girls (F) and 16 boys (M) from 8 to 14 years of age. Blue lines correspond to the individual least-squares fit of the simple linear regression model

Table 4 Orthodont data set. Estimates of fixed effects parameters obtained by the MLE (with and without the outlying observations), the SMDM-estimator, the CVFS-estimator, the Composite τ-estimator and MDPDE for $\alpha = \alpha^* = 0.2$ and $\alpha = \bar{\alpha} = 0.5$

	MLE	SMDM	CSVF	Comp. τ	MDPD−0.2	MDPD−0.5	MLE (without)
$\hat{\beta}_0$	16.34	16.91	16.30	17.43	17.20	17.09	17.15
$\hat{\beta}_1$	1.03	0.53	1.32	0.08	0.35	0.74	0.22
$\hat{\beta}_2$	0.78	0.71	0.72	0.67	0.69	0.68	0.68
$\hat{\beta}_3$	−0.30	−0.23	−0.22	−0.20	−0.23	−0.24	−0.20

Table 5 Orthodont data set. Estimates of variance components parameters obtained by the MLE (with and without the outlying observations), the SMDM-estimator, the CVFS-estimator, the Composite τ-estimator and MDPDE for $\alpha = \alpha^* = 0.2$ and $\alpha = \bar{\alpha} = 0.5$

	MLE	SMDM	CSVF	Comp. τ	MDPD−0.2	MDPD−0.5	MLE (without)
$\hat{\sigma}_0^2$	1.72	1.21	1.41	1.09	0.93	0.95	0.88
$\hat{\sigma}_1^2$	5.79	1.08	0.46	2.00	3.51	3.06	3.35
$\hat{\sigma}_{12}$	−0.29	0.03	−0.00	0.02	−0.10	−0.09	−0.13
$\hat{\sigma}_2^2$	0.03	0.00	0.02	0.01	0.02	0.02	0.02

as outliers. The estimates of β_2 and β_3 parameters are quite similar among the considered estimators. The fixed effect parameters differ mainly for the intercept β_0 and β_1, which indicates the effect of gender on the response $\boldsymbol{y}_i$. Robust estimators show values closer to zero than the standard MLE, indicating that the orthodontic distance is less affected by gender. This can be explained by the fact that MLE is sensitive to the presence of outlying observations obtaining a bigger value. Indeed, when such observations are removed, the maximum likelihood estimates are similar to MDPDE. The main differences on the estimation of the random effects terms are both in size (error variance component) and shape (correlation components). The MDPDEs assign less variance to the error term, and the other estimates are in general smaller. As before, the estimates given by the MLE and MDPDE are very close when the outliers are removed. Notice that the SMDM- and CVFS-estimates are quite different with respect to the others, especially for the variance components parameters.

Figure 7 shows the QQ-plots of the estimates of the random terms u_{ij} for $i = 1, \ldots, 27$ and $j = 1, 2$, obtained by the MLE, the SMDM-estimator, the CVFS-estimator, the Composite τ-estimator and MDPDE for $\alpha = \alpha^* = 0.2$ and $\alpha = \bar{\alpha} = 0.5$. The QQ-plots seem to show some structure and confirm the larger

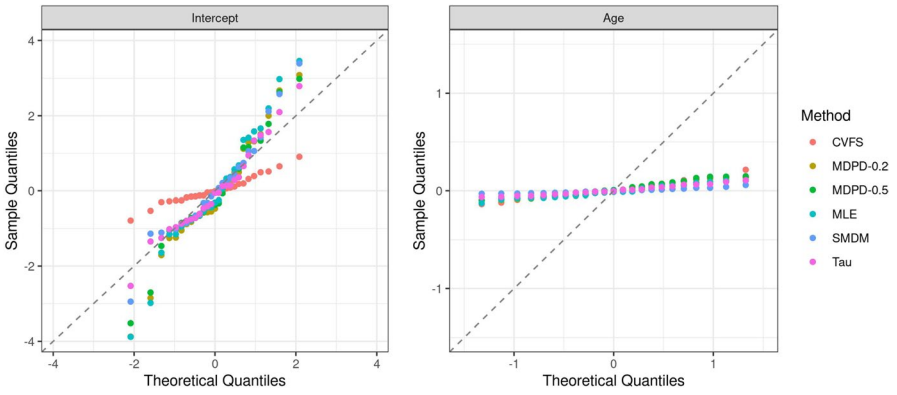


Fig. 7 Orthodont data set. QQ-plots of the random effects estimated by the MLE, the SMDM-estimator, the CVFS-estimator, the Composite τ-estimator and MDPDE for $\alpha = \alpha^* = 0.2$ and $\alpha = \bar{\alpha} = 0.5$

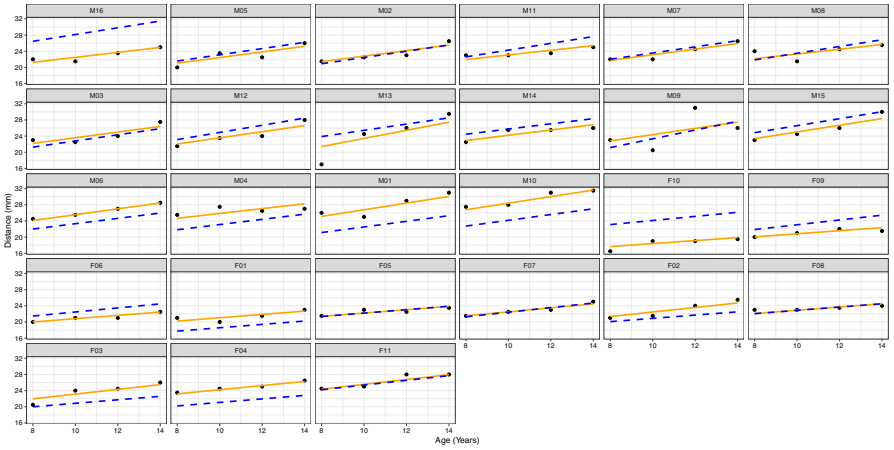


Fig. 8 Orthodontic growth patterns in 11 girls (F) and 16 boys (M) from 8 to 14 years of age. The fitted lines given by the MLE and MDPDE for $\alpha = 0.2$ are represented in blue and orange, respectively

variances estimated by MLE then those given by MDPDE; however, notice that the sample size ($n = 27$) is quite low. Figure 13a and 13b, in Section SM–7 of the Supplementary Materials, shows the QQ-plots of the random effects estimated by different methods separately.

Finally, Fig. 8 shows the fitted lines given by fixed effects estimates and random effects predictions for the MLE and MDPDE with $\alpha = 0.2$, represented in blue and orange, respectively. It is easy to see that the MDPDE achieves a better fit. In particular, notice the case of Subject `M10` previously indicated as possible outlier.

Notice that for the simulation study in Sect. 5.2, the predictions of the response variable have been computed as $\widehat{\boldsymbol{y}} = \boldsymbol{x}\widehat{\boldsymbol{\beta}}$, without considering the random effect estimates, since we were interested in evaluating the performance with respect to fixed effects estimation. In practical situations, such as the one presented in this section, when a linear mixed model is considered, both fixed effects and random effects estimates must be considered, i.e., $\widehat{\boldsymbol{y}}_i = \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{u}}_i$.

7 Conclusions

In this paper, we have developed an estimator based on the density power divergences to deal with the robustness issues in the linear mixed model setup. We demonstrated that the desirable asymptotic properties of the MDPDE, such as consistency and asymptotic normality, hold for the linear mixed model setup. In order to assess the robustness properties, the influence function and sensitivity measures of the estimator were computed. We found that the estimator is B-robust for $\alpha > 0$. From a practical point of view, the choice of the value of the tuning parameter α is fundamental in applications. The behavior of the sensitivity measures suggested two optimal values, denoted by α^* and $\bar{\alpha}$, depending on the dimension p, where the term "optimal" is in the sense of providing minimum sensitivity, and thus producing maximum robustness. The existence of such values is in contrast to the previous

knowledge about the parameter α. Indeed, it was shown that when α continues to increase beyond a certain value, we lose both robustness and efficiency.

The simulation study confirmed how the performance of the minimum density power divergence estimator changes with respect to α. Furthermore, the MDPDE outperforms the competitor estimators; indeed our approach leads to more resistant estimators in the presence of case-wise contamination. Finally, the application of our estimator to a real-life data set indicated that the MDPDE has similar results to the classical maximum likelihood estimator.

We feel that many important extensions of this work are necessary and can be potentially useful. So far, the MDPDE has been implemented only for balanced data (although the theory that we have developed is perfectly general). In future, we propose to extend the implementation to the more general case of groups with possibly different dimensions. The problem of testing of hypothesis also deserves a deeper look in the linear mixed models scenario.

Appendix A

A.1 Proof of Theorem 1

First let us note that under the setup of the linear mixed models introduced in Sect. 3, given a fixed $\alpha \geq 0$, for each i, the matrices $\boldsymbol{\Omega}_i$ and $\boldsymbol{J}_i$ defined in Section SM–1 of Supplementary Material simplify to the forms

$$\boldsymbol{J}^{(i)} = \mathbb{E}_{g_i}[\nabla H_i(\boldsymbol{Y}_i, \boldsymbol{\theta})] = \begin{pmatrix} \boldsymbol{J}_{11}^{(i)} & 0 \\ 0 & \boldsymbol{J}_{22}^{(i)} \end{pmatrix},$$

$$\boldsymbol{\Omega}^{(i)} = \mathbb{V}\mathrm{ar}_{g_i}(\nabla H_i(\boldsymbol{Y}_i, \boldsymbol{\theta})) = \begin{pmatrix} \boldsymbol{\Omega}_{11}^{(i)} & 0 \\ 0 & \boldsymbol{\Omega}_{22}^{(i)} \end{pmatrix},$$

where

$$\boldsymbol{J}_{11}^{(i)} = \frac{4\boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{X}_i}{(1+\alpha)^{\frac{n_i}{2}+1}},$$

$$(j,k)\text{-th element of } \boldsymbol{J}_{22}^{(i)} = \frac{T(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}, \boldsymbol{V}_i^{-1}\boldsymbol{U}_{ik})}{(1+\alpha)^{\frac{n_i}{2}+2}},$$

$$\boldsymbol{\Omega}_{11}^{(i)} = \frac{4\boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1}\boldsymbol{X}_i}{(1+2\alpha)^{\frac{n_i}{2}+1}},$$

$$(j,k)\text{-th element of } \boldsymbol{\Omega}_{22}^{(i)} = \frac{T(4, \boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}, \boldsymbol{V}_i^{-1}\boldsymbol{U}_{ik})}{(1+2\alpha)^{\frac{n_i}{2}+2}} - \frac{\alpha^2 \mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ik})\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij})}{(1+\alpha)^{n_i+2}},$$

for $j,k \in \{0,\ldots,r\}$, with $T(c,\boldsymbol{A},\boldsymbol{B}) = c\alpha^2\mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{B}) + 2\mathrm{Tr}(\boldsymbol{AB})$ for general matrices $\boldsymbol{A},\boldsymbol{B}$ and a constant c ($c = 1$ if not specified). Finally, put

$$\boldsymbol{\Psi}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\eta_{i\alpha}}{4}\boldsymbol{J}^{(i)} \quad \text{and} \quad \boldsymbol{\Omega}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\eta_{i\alpha}^2}{4}\boldsymbol{\Omega}^{(i)}.$$

Also, we assumed that the true data generating density belongs to the model family. Then, the proof of our Theorem 1 is immediate from the results stated by Ghosh and Basu (2013), provided we can show that the general Assumptions (A1)–(A7) are satisfied under the our assumed Conditions (MM1)–(MM4) for the special case of LMMs.

Now, the assumption that the true data generating density belongs to the model family, together with that the model density is normal with mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and variance matrix $\boldsymbol{V}_i$, ensures that Assumptions (A1)–(A3) are directly satisfied. Also, Assumption (A4) follows from Condition (MM1).

Next, we prove Equation (1) of Assumption (A6). For any $j = 1, \ldots, p$, the j-th partial derivative with respect to $\boldsymbol{\beta}$ is given by

$$\nabla_j H_i(\boldsymbol{Y}_i, \boldsymbol{\theta}) = -\alpha\left(1+\frac{1}{\alpha}\right)\eta_{i\alpha}e^{-\frac{\alpha}{2}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})}\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta}).$$

Then, considering $\boldsymbol{Z}_i = \boldsymbol{V}_i^{-\frac{1}{2}}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$,

$$\begin{aligned}
&\frac{(1+\alpha)}{n}\sum_{i=1}^{n}\mathrm{E}_{g_i}\left[\left|\eta_{i\alpha}e^{-\frac{\alpha}{2}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})}\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\right|\right.\\
&\quad\left.\times\ uadI\left(\left|\eta_{i\alpha}e^{-\frac{\alpha}{2}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})}\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\right| > \frac{N}{1+\alpha}\right)\right]\\
&\leq \frac{(1+\alpha)}{n}\sum_{i=1}^{n}\eta_{i\alpha}|\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-\frac{1}{2}}|\mathrm{E}_{g_i}\left[e^{-\frac{\alpha}{2}\boldsymbol{Z}_i^\top\boldsymbol{Z}_i}|\boldsymbol{Z}_i|\right.\\
&\quad\left.\times I\left(e^{-\frac{\alpha}{2}\boldsymbol{Z}_i^\top\boldsymbol{Z}_i}|\boldsymbol{Z}_i| > \frac{N}{(1+\alpha)(\sup\limits_{n>1}\eta_{i\alpha})(\sup\limits_{n>1}\max\limits_{1\leq i\leq n}|\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-\frac{1}{2}}|)}\right)\right]\\
&= E_1\left[e^{-\frac{\alpha}{2}\boldsymbol{Z}_1^\top\boldsymbol{Z}_1}|\boldsymbol{Z}_1|I\left(e^{-\frac{\alpha}{2}\boldsymbol{Z}_1^\top\boldsymbol{Z}_1}|\boldsymbol{Z}_1| > \frac{N}{(1+\alpha)(\sup\limits_{n>1}\eta_{i\alpha})(\sup\limits_{n>1}\max\limits_{1\leq i\leq n}|\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-\frac{1}{2}}|)}\right)\right]\\
&\quad\times\left(\frac{1+\alpha}{n}\sum_{i=1}^{n}\eta_{i\alpha}|\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-\frac{1}{2}}|\right)
\end{aligned}$$

Since the $\sup_{n>1}\max_{1\leq i\leq n}|\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-\frac{1}{2}}| = \mathcal{O}(1)$ by Assumption (MM2) and $\sup_{n>1}\eta_{i\alpha}$ is bounded thanks to the boundness of $|\boldsymbol{V}_i|$ in (MM3), by the dominated convergence theorem (DCT), we have

$$\lim_{N\to\infty} E_1\left[e^{-\frac{\alpha}{2}\mathbf{Z}_1^\top\mathbf{Z}_1}|\mathbf{Z}_1|I\left(e^{-\frac{\alpha}{2}\mathbf{Z}_1^\top\mathbf{Z}_1}|\mathbf{Z}_1| > \frac{N}{(1+\alpha)(\sup_{n>1}\eta_{i\alpha})(\sup_{n>1}\max_{1\le i\le n}|\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-\frac{1}{2}}|)}\right)\right] = 0.$$

Also

$$\sup_{n>1}\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{i\alpha}|\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-\frac{1}{2}}|\right) \le (\sup_{n>1}\eta_{i\alpha})(\sup_{n>1}\max_{1\le i\le n}|\boldsymbol{X}_{ij}^\top\boldsymbol{V}_i^{-\frac{1}{2}}|) = \mathcal{O}(1)$$

and this follows for all $j = 1, \dots, p$. On the other hand, consider the partial derivative with respect to σ_j^2, $j = 0, 1, \dots, r$, given in Equation (10). Hence, denoting with $\boldsymbol{Z}_i = \boldsymbol{V}_i^{-\frac{1}{2}}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{g_i}\left[|\nabla_{\sigma^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| \times I(|\nabla_{\sigma_j^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| > N)\right] \\
&\quad\le \frac{\alpha}{2n}\sum_{i=1}^{n}\frac{\eta_{i\alpha}\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij})}{(1+\alpha)^{\frac{n_i}{2}}}\mathrm{E}_{g_i}[I(|\nabla_{\sigma_j^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| > N)] \\
&\qquad+ \frac{1+\alpha}{2n}\sum_{i=1}^{n}\eta_{i\alpha}\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij})\mathrm{E}_{g_i}[e^{-\frac{\alpha}{2}\mathbf{Z}_i^\top\mathbf{Z}_i}I(|\nabla_{\sigma_j^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| > N)] \\
&\qquad+ \frac{1+\alpha}{2n}\sum_{i=1}^{n}\eta_{i\alpha}\mathrm{E}_{g_i}[|\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i|e^{-\frac{\alpha}{2}\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i}I(|\nabla_{\sigma_j^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| > N)] \\
&\quad= \left(\frac{\alpha}{2n}\sum_{i=1}^{n}\frac{\eta_{i\alpha}\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij})}{(1+\alpha)^{\frac{n_i}{2}}}\right)\mathrm{E}_1[I(|\nabla_{\sigma_j^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| > N)] \\
&\qquad+ \left(\frac{1+\alpha}{2n}\sum_{i=1}^{n}\eta_{i\alpha}\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij})\right)\mathrm{E}_1[e^{-\frac{\alpha}{2}\mathbf{Z}_1^\top\mathbf{Z}_1}I(|\nabla_{\sigma_j^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| > N)]+ \\
&\qquad+ \frac{1+\alpha}{2n}\sum_{i=1}^{n}\eta_{i\alpha}\mathrm{E}_{g_i}[|\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i|e^{-\frac{\alpha}{2}\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i}I(|\nabla_{\sigma_j^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| > N)]
\end{aligned}$$

where in the last term $\boldsymbol{Z}_i = (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$. Note that

$$\mathrm{E}_{g_i}[\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i e^{-\frac{\alpha}{2}\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i}] = \frac{\mathrm{Tr}(\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij})}{(1+\alpha)^{1+\frac{n_i}{2}}}$$

; hence, by Equation (15),

$$\mathrm{E}_{g_i}\left[|\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{U}_{ij}\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i|e^{-\frac{\alpha}{2}\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-1}\boldsymbol{Z}_i}I(|\nabla_{\sigma^2}H_i(\boldsymbol{Y}_i,\boldsymbol{\theta})| > N)\right] \to 0, \quad \text{as } N \to \infty.$$

The expectation in the first two terms goes to zero as $N \to \infty$ by the DCT, as before, and the sums are bounded by Equation (15) in condition (MM3). This holds for all $j = 0, \ldots, r$.

Finally, Assumptions (A5), (A7) and Equation (2) similarly hold using Equations (14) and (16), (17), (13) and (15), respectively.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10182-023-00473-z.

Acknowledgements The research of AG is supported by an INSPIRE Faculty Research Grant from Department of Science and Technology, Government of India, and an internal research grant from Indian Statistical Institute, India.

Author Contributions G.S. designed the methodology, developed the implementation of the proposed method and prepared the visualization of the results as well as the preparation of the original presented work and of the revised version. A.G. formulated the initial idea and the research goal and aims and validated the theoretical and experimental research output. A.B. is responsible of coordinating the research activity. C.A. supervised the planning and the execution of the project and validated the implementation of the method.

Funding Open access funding provided by Università degli Studi di Trento within the CRUI-CARE Agreement.

Declarations

Conflict of interest We have no conflicts of interest to disclose.

References

Agostinelli, C., Yohai, V.J.: robustvarComp: Robust Estimation for Variance Component Models. (2019). R package version 0.1-6

Agostinelli, C., Yohai, V.J.: Composite robust estimators for linear mixed models. J. Am. Stat. Assoc. **111**(516), 1764–1774 (2016)

Basu, A., Harris, I.R., Hjort, N., Jones, M.C.: Robust and efficient estimation by minimizing a density power divergence. Biometrika **85**(3), 549–559 (1998)

Basu, A., Park, C., Shioya, H.: Statistical Inference: The Minimum Distance Approach. CRC Press, Chapman and Hall (2011)

Castilla, E., Ghosh, A., Martin, N., Pardo, L.: New robust statistical procedures for the polytomous logistic regression models. Biometrics **74**(4), 1282–1291 (2018)

Castilla, E., Ghosh, A., Martin, N., Pardo, L.: Robust semiparametric inference for polytomous logistic regression with complex survey design. Adv. Data Anal. Classificat. **15**(3), 701–734 (2021). https://doi.org/10.1007/s11634-020-00430-7
Christensen, R.: Mixed models and variance components. In: Plane Answers to Complex Questions: The Theory of Linear Models, pp. 291–331. Springer, New York, NY (2011)
Copt, S., Victoria-Feser, M.P.: High breakdown inference in the mixed linear model. J. Am. Stat. Assoc. **101**, 292–300 (2006)
Ghosh, A.: Robust inference under the beta regression model with application to health care studies. Stat. Methods Med. Res. **28**(3), 871–888 (2019)
Ghosh, A., Basu, A.: Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. Electr. J. Stat. **7**, 2420–2456 (2013)
Ghosh, A., Basu, A.: Robust estimation in generalized linear models: The density power divergence approach. TEST **25**, 269–290 (2016)
Ghosh, A., Basu, A.: Robust and efficient estimation in the parametric proportional hazards model under random censoring. Stat. Med. **38**(27), 5283–5299 (2019)
Hampel, F.R.: Contributions to the Theory of Robust Estimation. University of California, Berkeley (1968)
Hampel, F.R.: The influence curve and its role in robust estimation. J. Am. Stat. Assoc. **69**(346), 383–393 (1974)
Huggins, R.M.: On the robust analysis of variance components models for pedigree data. Aust. J. Stat. **35**(1), 43–57 (1993)
Huggins, R.M.: A robust approach to the analysis of repeated measures. Biometrics **49**(3), 715–720 (1993)
Huggins, R.M., Staudte, R.G.: Variance components models for dependent cell populations. J. Am. Stat. Assoc. **89**(425), 19–29 (1994)
Koller, M.: Robust estimation of linear mixed models. PhD thesis, ETH Zürich (2013)
Koller, K.: robustlmm: An R package for robust estimation of linear mixed-effects models. J. Stat. Softw. **75**(6), 1–24 (2016)
Lange, K.L., Little, R.J.A., Taylor, J.M.G.: Robust statistical modeling using the t distribution. J. Am. Stat. Assoc. **84**(408), 881–896 (1989)
McCulloch, C.E., Searle, S.R.: Generalized, Linear, and Mixed Models. John Wiley & Sons, Wiley Series in Probability and Statistics (2001)
Pinheiro, J., Bates, D., R Core Team: Nlme: Linear and Nonlinear Mixed Effects Models. (2022). R package version 3.1-157. https://CRAN.R-project.org/package=nlme
Pinheiro, J.C., Liu, C., Wu, Y.N.: Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. J. Comput. Graph. Stat. **10**(2), 249–276 (2001)
Potthoff, R.F., Roy, S.N.: A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika **51**(3/4), 313–326 (1964)
R Core Team: R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2019). R Foundation for Statistical Computing
Richardson, A.M.: Bounded influence estimation in the mixed linear model. J. Am. Stat. Assoc. **92**(437), 154–161 (1997)
Richardson, A.M., Welsh, A.H.: Robust restricted maximum likelihood in mixed linear models. Biometrics **51**(4), 1429–1439 (1995)
Sinha, S.K.: Robust analysis of generalized linear mixed models. J. Am. Stat. Assoc. **99**(466), 451–460 (2004)
Stahel, W.A., Welsh, A.: Approaches to robust estimation in the simplest variance components model. J. Stat. Plan. Infer. **57**(2), 295–319 (1994)
Welsh, A.H., Richardson, A.M.: 13 approaches to the robust estimation of mixed models. Handbook Stat. **15**, 343–384 (1997)
Yau, K.K.W., Kuk, A.Y.C.: Robust estimation in generalized linear mixed models. J. Royal Stat. Soc. **64**(1), 101–117 (2002)

Authors and Affiliations

Giovanni Saraceno[1,3] · Abhik Ghosh[2] · Ayanendranath Basu[2] · Claudio Agostinelli[1]

Abhik Ghosh
abhik.ghosh@isical.ac.in

Ayanendranath Basu
ayanbasu@isical.ac.in

Claudio Agostinelli
claudio.agostinelli@unitn.it

[1] Department of Mathematics, University of Trento, Via Sommarive 14, 38123 Trento, Italy

[2] Interdisciplinary Statistical Research Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road Kolkata, Kolkata 700 108, India

[3] Department of Biostatistics, University at Buffalo, 811 Kimball Tower, Buffalo, NY 14214, USA