Robust Henderson III Estimators of Variance Components in the Nested Error Model

Betsabé Pérez, Daniel Peña, and Isabel Molina

Universidad Carlos III de Madrid, Department of Statistics,
Postal address: C/Madrid 126, 28903 Getafe, Spain
{betsabe.perez,daniel.pena,isabel.molina}@uc3m.es

Summary. This work deals with robust estimation of variance components under a nested error model. Traditional estimation methods include Maximum Likelihood (ML), Restricted Maximum Likelihood (REML) and Henderson method III (H3). However, when outliers are present, these methods deliver estimators with poor properties. Robust modifications of the ML and REML have been proposed in the literature (see for example, Fellner [3], Richardson and Welsh [14] and Richardson [13]). In this work we explore some robust alternatives based on the idea of Henderson method III. The work is organized as follows. In section 2, we introduce the nested error model. In section 3, we describe the traditional methods for estimating variance components. In section 4, several robustified versions of the H3 estimators of the variance components are presented. In section 5, we present some results on diagnostics methods. In section 6 we perform a Monte Carlo study to compare the new robust estimation methods with the non-robust alternatives.

Keywords: Henderson method III, linear mixed models, nested error model, outliers, robust estimation, variance components

1 Introduction

In the last decades, linear mixed models (Laird and Ware [8]) have received considerable attention in the literature from a practical and theoretical point of view (e.g. McCulloch and Searle [10], Verbeke and Molenberghs [17] and Jiang [7]). These models are frequently used in small area estimation or to analyze repeated measures data, because they model flexibly the within-subject correlation often present in these type of data. However, there are many other fields of application of these models, such as clinical trials (Vangeneugden *et al.* [16]), air pollution studies (Wellenius *et al.* [18]), etc. Despite the many applications in which these models are used, only few works have been done on model diagnostics, an important step to validate the model. Christensen et al. [2] studied case deletion diagnostics. Banerjee and Frees [1] developed influence diagnostics. Galpin and Zewotir [5] extended some results of the ordinary linear regression influence diagnostics to the linear mixed models context such as residuals, leverages and outliers when the variance components are known. However, in practice the variance components need to be estimated from sample data. If sample data are contaminated, then the estimation might be affected and this will in turn affect all diagnostic tools.

Here we focus on a particular linear mixed model with only one random factor, called nested error model. For this model, we propose several robust alternatives to the H3 estimators of variance components. Section 2 describes the data structure and the model. Section 3 summarizes the most common methods for estimation. Section 4 introduces our proposed robust estimators. Section 5 describes diagnostic tools for these models and finally, Section 6 reports the results of a simulation study that compares the robustness properties of the proposed estimators with those of the traditional non-robust ones. Finally, Section 7 gives some concluding remarks.

2 The Model

In this section we introduce the nested error model and describe some of its properties. Consider that the sample observations come from D different populations groups, with n_d observations coming from d-th group, $d = 1, \ldots, D$ and $n = \sum_{d=1}^{D} n_d$ being the total sample size. Let us denote y_{dj} the value of the study variable for j-th sample unit from d-th group and $\mathbf{x}_{dj}$ a (column) vector containing the values of p auxiliary variables for the same unit. We consider the model

$$y_{dj} = \mathbf{x}_{dj}^T \boldsymbol{\beta} + u_d + e_{dj} \quad j = 1, \ldots, n_d \quad d = 1, \ldots, D, \tag{1}$$

where $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed parameters, u_d is the random effect of d-th group and e_{dj} is the model error. Random effects and errors are supposed to be independent with distributions

$$u_d \overset{iid}{\sim} N(0, \sigma_u^2) \quad \text{and} \quad e_{dj} \overset{iid}{\sim} N(0, \sigma_e^2).$$

Stacking the model elements y_{dj}, $\mathbf{x}_{dj}^T$ and e_{dj} in columns, we can express the model in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad \mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_D), \quad \mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_n). \tag{2}$$

where $\mathbf{u} = (u_1, \ldots, u_D)^T$ and $\mathbf{Z}$ is the $n \times D$ design matrix associated with $\mathbf{u}$, containing in its columns the indicators of the groups $d = 1, \ldots, D$.

The expectation and covariance matrix of $\mathbf{y}$ are given by

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{and} \quad Var(\mathbf{y}) = \sigma_u^2 \mathbf{Z}\mathbf{Z}^T + \sigma_e^2 \mathbf{I}_n := \mathbf{V}.$$

Let us define the vector of variance components $\boldsymbol{\theta} = (\sigma_u^2, \sigma_e^2)^T$. When $\boldsymbol{\theta}$ is known, Henderson [6] obtained the Best Linear Unbiased Estimator (BLUE) of $\boldsymbol{\beta}$ and the Best Linear Unbiased Predictor (BLUP) of $\mathbf{u}$, which are given respectively by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}, \tag{3}$$

$$\tilde{\mathbf{u}} = \sigma_u^2\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \tag{4}$$

3 Estimation of Variance Components

The estimator of $\boldsymbol{\beta}$ and the predictor of $\mathbf{u}$ in (3) and (4) respectively depend on $\boldsymbol{\theta}$, which in practice is unknown and needs to be estimated from sample data. The empirical versions of (3) and (4), (EBLUE and EBLUP respectively) are obtained by replacing a suitable estimator $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)^T$ for $\boldsymbol{\theta}$ in (3) and (4) and are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}, \tag{5}$$

$$\hat{\mathbf{u}} = \hat{\sigma}_u^2\mathbf{Z}^T\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{6}$$

where $\hat{\mathbf{V}} = \hat{\sigma}_u^2\mathbf{Z}\mathbf{Z}^T + \hat{\sigma}_e^2\mathbf{I}_n$.

Next we describe the ML, REML and H3 methods to estimate variance components.

Maximum likelihood

Maximum likelihood estimation is usually done by assumming that $\mathbf{y}$ has a multivariate normal distribution. Under this assumption, the likelihood is given by

$$f(\boldsymbol{\theta}|\mathbf{y}) = (2\pi)^{-\frac{n}{2}}|\mathbf{V}|^{-1/2}exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}.$$

The log-likelihood is

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = \ln(f(\boldsymbol{\theta}|\mathbf{y})) = c - \frac{1}{2}[\ln|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})],$$

where c is denotes a constant. Using the relations

$$\frac{\partial \ln|\mathbf{V}|}{\partial\boldsymbol{\theta}} = tr\left\{\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\boldsymbol{\theta}}\right\} \quad \text{and} \quad \frac{\partial\mathbf{V}^{-1}}{\partial\boldsymbol{\theta}} = -\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\boldsymbol{\theta}}\mathbf{V}^{-1},$$

the first order partial derivatives of ℓ with respect to $\boldsymbol{\beta}$, σ_u^2 and σ_e^2 are

$$\begin{aligned}
\frac{\partial\ell(\boldsymbol{\theta}|\mathbf{y})}{\partial\boldsymbol{\beta}} &= \mathbf{X}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),\\
\frac{\partial\ell(\boldsymbol{\theta}|\mathbf{y})}{\partial\sigma_u^2} &= -\frac{1}{2}tr\left\{\mathbf{V}^{-1}\mathbf{Z}\mathbf{Z}^T\right\} + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}\mathbf{Z}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),\\
\frac{\partial\ell(\boldsymbol{\theta}|\mathbf{y})}{\partial\sigma_e^2} &= -\frac{1}{2}tr\{\mathbf{V}^{-1}\} + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),
\end{aligned}$$

and equating to zero we obtain the equations

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} = \mathbf{X}\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}, \tag{7}$$

$$tr\{\mathbf{V}^{-1}\mathbf{Z}\mathbf{Z}^T\} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}\mathbf{Z}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{8}$$

$$tr\{\mathbf{V}^{-1}\} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{9}$$

From (7), the ML estimating equation for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}$$

Equations (8) and (9) do not have analytic solution and need to be solved using numerical methods such as Newton-Raphson or Fisher-Scoring.

Restricted maximum likelihood

REML approach starts by transforming $\mathbf{y}$ into two independent vectors, $\mathbf{y}_1 = K_1\mathbf{y}$ and $\mathbf{y}_2 = K_2\mathbf{y}$. The distribution of $\mathbf{y}_1$ does not depend on $\boldsymbol{\beta}$ and satisfies $E(\mathbf{y}_1) = \mathbf{0}$, which means that $K_1\mathbf{X} = \mathbf{0}$. On the other hand, $\mathbf{y}_2$ is independent of $\mathbf{y}_1$, which means that $K_1\mathbf{V}K_2^T = \mathbf{0}$. The matrix K_1 is chosen to have maximum rank, i.e. $n - p$, so the rank of K_2 is p. The likelihood function of $\mathbf{y}$ is the product of the likelihoods of $\mathbf{y}_1$ and $\mathbf{y}_2$. The variance components coming from the REML approach are the ML estimators of these parameters based on $\mathbf{y}_1$, see Patterson and Thompson [11]. Similarly to the ML case, the obtained equations do not have analytic solutions and need to be solved using iterative techniques. This method takes into account the degrees of freedom of estimation of $\boldsymbol{\beta}$ when estimating the variance components and for this reason it gives less biased estimators.

Henderson method III

ML and REML estimators of $\boldsymbol{\theta}$ are typically obtained under the assumption that the vector $\mathbf{y}$ has a multivariate normal distribution. In many circumstances, however, this assumption does not hold. An alternative method which does not rely on the normality assumption and provides explicit solutions to the variance components estimators is the Henderson method III. This method works as follows. First, consider a general linear mixed model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\boldsymbol{\beta}$ might contain fixed and random effects. Let us split $\boldsymbol{\beta}$ into two subvectors $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and rewrite the model as

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}. \tag{10}$$

If we treat $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ as fixed, the total and residual sum of squares are respectively

$$SST = \mathbf{y}^T\mathbf{y} \quad \text{and} \quad SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mathbf{y}^T\mathbf{P}\mathbf{y},$$

where $\mathbf{P} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. The sum of squares of the regression is

$$SSR(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = SST - SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \mathbf{y}^T\mathbf{Q}\mathbf{y}, \tag{11}$$

where $\mathbf{Q} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Secondly, consider the reduced model

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \varepsilon, \tag{12}$$

considering $\boldsymbol{\beta}_1$ as fixed. The residual sum of squares is given by

$$SSE(\boldsymbol{\beta}_1) = \mathbf{y}^T\mathbf{P}_1\mathbf{y},$$

where $\mathbf{P}_1 = \mathbf{I}_n - \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T$. The sum of squares of the regression is

$$SSR(\boldsymbol{\beta}_1) = SST - SSE(\boldsymbol{\beta}_1) = \mathbf{y}^T\mathbf{Q}_1\mathbf{y},$$

where $\mathbf{Q}_1 = \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T$. The reduction in sum of squares due to introducing $\mathbf{X}_2$ in the model with only $\mathbf{X}_1$ is

$$SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SSR(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - SSR(\boldsymbol{\beta}_1). \tag{13}$$

Now consider model (2) and rewrite it as (10) taking $\boldsymbol{\beta}_1 = \boldsymbol{\beta}$, $\boldsymbol{\beta}_2 = \mathbf{u}$, $\mathbf{X}_1 = \mathbf{X}$ and $\mathbf{X}_2 = \mathbf{Z}$. This method calculates the expectations of (11) and (13) and equates the sum of squares to their expectations obtaining two equations. Solving for σ_e^2 and σ_u^2 in the resulting equations, we obtain unbiased estimators for σ_e^2 and σ_u^2 (for more details see Searle *et al.* [15], chapter 5). Let $\hat{\mathbf{e}}$ and $\hat{\boldsymbol{\varepsilon}}$ be the vectors of residuals of the submodels (10) and (12), respectively. If $\text{rank}(\mathbf{X}) = p$ and $\text{rank}(\mathbf{X}|\mathbf{Z}) = p + D$, then the Henderson III estimators of the variance components are given by

$$\hat{\sigma}_e^2 = \frac{\sum_{d=1}^{D}\sum_{j=1}^{n_d}\hat{e}_{dj}^2}{n - p - D}, \quad \hat{\sigma}_u^2 = \frac{\sum_{d=1}^{D}\sum_{j=1}^{n_d}\hat{\varepsilon}_{dj}^2 - \hat{\sigma}_e^2(n-p)}{tr\left\{\mathbf{Z}^T[\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{Z}\right\}}, \tag{14}$$

where $\hat{e}_{dj}$ is the residual corresponding to observation $(\mathbf{x}_{dj}^T, y_{dj})$ in model (10) and $\hat{\varepsilon}_{dj}$ is the residual for the same observation but obtained from model (12).

4 Robust Estimation of Variance Components

In this section we introduce some new robust estimators of variance components based on Henderson method III. We have chosen this method for three reasons; first, because it provides explicit formulas of the estimators, fact which will help to decrease the computational time; second, it does not need the normality assumption; third, the estimation procedure consists simply of solving two standard regression problems. Let us rewrite the estimators as

$$\hat{\sigma}_e^2 = \frac{n[\sum_{d=1}^{D}\sum_{j=1}^{n_d}\hat{e}_{dj}^2/n]}{n-p-D}, \quad \hat{\sigma}_u^2 = \frac{n[\sum_{d=1}^{D}\sum_{j=1}^{n_d}\hat{\varepsilon}_{dj}^2/n] - \hat{\sigma}_e^2(n-p)}{tr\,\{\mathbf{Z}^T[\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{Z}\}}, \quad (15)$$

These two estimators contain in the numerator sample means of squared residuals obtained from models (10) and (12) respectively. Then a small fraction of outliers, even a single observation, might seriously affect these estimators. To avoid this problem, we propose to use robust methods to fit the two models (10) and (12) and then, replacing the means of squared residuals in (15) by other more robust functions of residuals. Model (12) is a standard linear regression model, which can be robustly fitted using any method available in the literature such as L_1 estimation, M estimation or the fast method of Peña and Yohai [12]. Model (10) is a model with a categorical variable that distributes the observations into groups, which can be robustly fitted using an adaptation of the principal sensibility components method of Peña and Yohai [12] to the grouped data structure, or by the M-S estimation of Maronna and Yohai [9]. These fitting methods would then provide better residuals e_{dj} and ε_{dj}, which are in turn used to find robust estimators of the variance components similar to (15). Below we describe different estimators obtained using robust functions of these new residuals obtained using the robust fit of models (10) and (12).

MADH3 estimators

In (15), we substitute the mean of the squared residuals by the square of the normalized median of absolute deviations (MAD), given by

$$MAD = 1.481 \cdot \text{Med}(|\hat{\xi}_{dj}|, \hat{\xi}_{dj} \neq 0),$$

where $\hat{\xi}_{dj}$ is the residual of observation $(\mathbf{x}_{dj}^T, y_{dj})$ under the corresponding fitted model, either (10) or (12). Then, our first robust proposal for the estimation of the variance components is given by

$$\hat{\sigma}_{e,MADH3}^2 = \frac{n[1.481 \cdot \text{Med}_i(|\hat{e}_{dj}|, \hat{e}_{dj} \neq 0)]^2}{n-p-D} \quad (16)$$

$$\hat{\sigma}_{u,MADH3}^2 = \frac{n[1.481 \cdot \text{Med}_i(|\hat{\varepsilon}_{dj}|, \hat{\varepsilon}_{dj} \neq 0)]^2 - \hat{\sigma}_{e,MADH3}^2(n-p)}{tr\,\{\mathbf{Z}^T[\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{Z}\}} \quad (17)$$

TH3 estimators

Trimming consists of giving zero weight to a percentage of extreme cases. In this case, instead of this, in the two equations given in (15) we trim the residuals that are outside the interval (b_1, b_2) with

$$b_1 = q_1 - k(q_3 - q_1) \quad \text{and} \quad b_2 = q_3 + k(q_3 - q_1). \quad (18)$$

Here, q_1 and q_2 are the first and third sample quartiles of residuals and k is a constant. Based on results obtained from different simulation studies, we propose to use the constant $k = 2$, just slightly smaller than that one used as outer frontier in the box-plot for detecting outliers.

RH3 estimators

Instead of replacing extreme residuals by zero as in the previous proposal, we can smooth the residuals appearing in (15) according to an appropriate smoothing function. Here we consider the Tukey's biweight function, given by

$$\Psi(x) = x[1-(x/k)^2]^2, \quad \text{if } |x| \le k. \tag{19}$$

5 Model Diagnostics

In this section we describe some diagnostics tools for the nested error model. Considering that $\boldsymbol{\theta}$ is known, the vector of predicted values is defined as

$$\tilde{\mathbf{y}} = (\mathbf{I} - \mathbf{R})\mathbf{y},$$

where

$$\mathbf{R} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}. \tag{20}$$

This relation evokes the definition of the hat matrix as

$$\mathbf{H} = \mathbf{I} - \mathbf{R}.$$

The diagonal elements $(1 - r_{dj})$ of matrix $\mathbf{H}$ are measures of the leverage effect of the observations and are called *leverages*. Thus, Galpin and Zewotir [5] proposed the use of the r_{dj}s to identify influential observations. If r_{dj} approaches zero, this indicates that the corresponding observation $(\mathbf{x}_{dj}^T, y_{dj})$ has a large leverage effect.

Due to the data structure in nested error models, it seems more relevant to study the leverage effect of full groups instead of isolated observations. Here we define the leverage effect of group d as

$$h_d = \overline{\mathbf{x}}_d^T(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\overline{\mathbf{x}}_d, \quad d = 1, \ldots, D, \tag{21}$$

where $\overline{\mathbf{x}}_d = n_d^{-1}\sum_{j=1}^{n_d}\mathbf{x}_{dj}$.

Concerning residuals, the i-th internally studentized residual is defined as

$$t_{dj} = \frac{e_{dj}}{\sqrt{\mathrm{var}(e_{dj})}} = \frac{e_{dj}}{\sigma_e\sqrt{r_{dj}}} \tag{22}$$

In practice, the variance components involved in (20), (21) and (22) need to be estimated. When there are outliers, these might affect the estimators of variance components, and these in turn will change the distribution of standarized residuals. Better versions of these diagnostic tools can be obtained using the robust variance components estimators introduced in Section 4.

6 Monte Carlo Simulation

In this section we describe a Monte Carlo simulation study that compares the robust estimators of the variance components with the traditional ones. For this, we generated data coming from $D = 10$ groups. The group sample sizes n_d, $d = 1, \ldots, D$ were respectively 20, 20, 30, 30, 40, 40, 50, 50, 60 and 60, with a total sample size of $n = 400$. We considered $p = 4$ auxiliary variables, and they were generated from normal distributions with means and standard deviations coming from a real data set from the Australian Agricultural and Grazing Industries Survey. Thus, the values of the four auxiliary variables were generated respectively as $X_1 \sim N(3.3, 0.6)$, $X_2 \sim N(1.7, 1.2)$, $X_3 \sim N(1.7, 1.6)$ and $X_4 \sim N(2.4, 2.6)$.

The simulation study is based on $L = 500$ iterations. In each iteration, we generated group effects as $u_d \overset{iid}{\sim} N(0, \sigma_u^2)$ with $\sigma_u^2 = 0.25$. Similarly, we generated errors as $e_{dj} \overset{iid}{\sim} N(0, \sigma_e^2)$ with $\sigma_e^2 = 0.25$. Then we generated the model responses y_{dj}, $j = 1, \ldots, n_d$, $d = 1, \ldots, D$, from model (1). Observe that in principle there is no contamination. Finally, we introduced contamination according to three different scenarios:

A. *No contamination.*
B. *Groups with a mean shift:* A subset $\mathcal{D}_c \subseteq \{1, 2, \ldots, D\}$ of groups was selected for contamination. For each selected group $d \in \mathcal{D}_c$, half of the observations were replaced by $c_{d1} = \bar{y}_d + k\, s_{Y,d}$ and the other half by $c_{d2} = \bar{y}_d - k\, s_{Y,d}$ with $k = 5$, where $\bar{y}_d$ and $s_{Y,d}$ are respectively the mean and the standard deviation of the outcome for the clean data in d-th group. This increases the between group variability σ_u^2.
C. *Groups with high variability:* A small percentage of contaminated observations was introduced in each selected group $d \in \mathcal{D}_c$, similarly as described in Scenario B. This increases the within group variability σ_e^2.

After each iteration, we fitted the two models (10) and (12) using the procedure of Peña and Yohai [12], using in the first model an adaptation of this method for grouped data. Then, we calculated the traditional estimators H3, ML and REML, and the proposed robust estimators, MADH3, TH3 and RH3. After the $L = 500$ iterations, we computed their empirical bias and mean squared error (MSE). Table 1 reports the resulting empirical bias and percent MSE of each estimator under Scenario A, without contamination. Observe in that table that in absence of outlying observations, the traditional non-robust estimators, H3, ML and REML, provide the minimum MSE, but the robust alternatives TH3 and RH3 are not too far away from them. However, under Scenario B with full groups contaminated with a mean shift (Tables 2 and 3), the estimators ML, REML and H3 of σ_u^2 increase considerably their MSE. The estimator TH3 achieves the minimum MSE, followed by RH3. Under Scenario C with contamination introduced to make the within cluster variability increase (Tables 4 and 5), now the estimators ML, REML, and H3 of σ_e^2 increase considerably their MSE whereas the robust estimator TH3 resists quite well.

Table 1. Theoretical values $\sigma_u^2 = \sigma_e^2 = 0.25$. Scenario 0: No contamination

Method	Estimators		Bias		MSE $\times 10^2$	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
H3	0.24	0.25	-0.0081	0.0014	1.43	0.03
ML	0.22	0.25	-0.0298	-0.0011	1.16	0.03
REML	0.25	0.25	-0.0046	0.0014	1.32	0.03
MADH3	0.25	0.25	0.0041	0.0018	2.33	0.09
TH3	0.23	0.25	-0.0189	-0.0019	1.04	0.04
RH3	0.24	0.23	-0.0136	-0.0179	1.25	0.06

Table 2. Theoretical values $\sigma_u^2 = \sigma_e^2 = 0.25$. Scenario B: One outlying group

Method	Estimators		Bias		MSE $\times 10^2$	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
H3	1.28	0.24	1.0286	-0.0095	123.73	0.04
ML	1.15	0.24	0.9000	-0.0120	123.27	0.04
REML	1.28	0.24	1.0285	-0.0096	123.38	0.04
MADH3	0.44	0.23	0.1884	-0.0169	7.84	0.10
TH3	0.24	0.24	-0.0089	-0.0142	1.25	0.05
RH3	0.46	0.22	0.2106	-0.0277	6.04	0.10

Table 3. Theoretical values $\sigma_u^2 = \sigma_e^2 = 0.25$. Scenario B: Two outlying groups

Method	Estimators		Bias		MSE $\times 10^2$	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
H3	2.79	0.23	2.5375	-0.0242	715.98	0.08
ML	2.13	0.22	1.8807	-0.0266	495.49	0.10
REML	2.37	0.23	2.1179	-0.0242	500.14	0.08
MADH3	1.10	0.21	0.8529	-0.0437	91.67	0.25
TH3	0.27	0.22	0.0227	-0.0319	2.13	0.13
RH3	0.76	0.21	0.5088	-0.0412	31.52	0.19

Table 4. Theoretical values $\sigma_u^2 = \sigma_e^2 = 0.25$. Scenario C: 10% of atypical observations shared among groups

Method	Estimators		Bias		MSE $\times 10^2$	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
H3	0.23	0.60	-0.0175	0.3512	1.47	12.58
ML	0.21	0.60	-0.0397	0.3450	1.23	12.15
REML	0.24	0.60	-0.0144	0.3512	1.35	12.58
MADH3	0.28	0.27	0.0253	0.0198	2.78	0.14
TH3	0.24	0.25	-0.0073	-0.0012	1.17	0.04
RH3	0.22	0.30	-0.0266	0.0487	1.22	0.26

Table 5. Theoretical values $\sigma_u^2 = \sigma_e^2 = 0.25$. Scenario C: 20% of atypical observations shared among groups

Method	Estimators		Bias		MSE $\times 10^2$	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
H3	0.22	0.93	-0.0268	0.6814	1.50	47.19
ML	0.20	0.92	-0.0489	0.6719	1.32	45.89
REML	0.23	0.93	-0.0236	0.6814	1.39	47.19
MADH3	0.30	0.29	0.0473	0.0406	3.48	0.29
TH3	0.25	0.25	0.0045	0.0003	1.27	0.04
RH3	0.21	0.37	-0.0400	0.1151	1.18	1.35

7 Discussion

In this work we present three robust versions of H3 estimators called MADH3, TH3 and RH3 estimators. These robust estimators are obtained by first, fitting in a robust way the two models (10) and (12), and then replacing the means of squared residuals in H3 estimators by other robust functions of the residuals coming from those robust fittings. In simulations we have analyzed the robustness of our proposed estimators against two different kind of contamination scenarios: when the between groups variability is increased by including a mean shift in some groups, and when the within group variability is increased by introducing given percentages of outliers shared among the clusters. The new robust estimator TH3 gets the best results in these simulations, achieving great efficiency under both types of contamination but preserving at the same time good efficiency when there is not contamination.

Acknowledgement. This work is supported by the grant SEJ2007-64500 from the Spanish Ministerio de Educación y Ciencia and the Collaborative Project 217565, Call identifier FP7-SSH-2007-1, from the European Commision.

References

1. Banerjee, M., Frees, E.W.: Influence diagnostics for linear longitudinal models. J. Amer. Statist. Assoc. 92, 999–1005 (1997)
2. Christensen, R., Pearson, L.M., Johnson, W.: Case-deletion diagnostics for mixed models. Technometrics 34, 38–45 (1992)
3. Fellner, W.H.: Robust estimation of variance components. Technometrics 28, 51–60 (1986)
4. Galpin, J.S., Zewotir, T.: Influence diagnostics for linear mixed models. J. Data Sci. 3, 153–177 (2005)
5. Galpin, J.S., Zewotir, T.: A unified approach on residuals, leverages and outliers in the linear mixed models. TEST 16, 58–75 (2007)
6. Henderson, C.R.: Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423–447 (1975)
7. Jiang, J.: Linear and Generalized Linear Models and their Applications. Springer Series in Statistics. Springer, Heidelberg (2007)
8. Laird, N.M., Ware, J.H.: Random-effects models for longitudinal data. Biometrics 38, 963–974 (1982)
9. Maronna, R.A., Yohai, V.J.: Robust regression with both continuous and categorical predictors. J. Statist. Plan Infer. 89, 197–214 (2000)
10. McCulloch, C., Searle, S.: Generalized, Linear and Mixed Models. Wiley, New York (2001)
11. Patterson, H.D., Thompson, R.: Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545–554 (1971)
12. Peña, D., Yohai, V.J.: A fast procedure for outlier diagnostics in large regression problems. J. Amer. Statist. Assoc. 94, 434–445 (1999)
13. Richardson, A.M.: Bounded influence estimation in the mixed linear model. J. Amer. Statist. Assoc. 92, 154–161 (1997)
14. Richardson, A.M., Welsh, A.H.: Robust restricted maximum likelihood in mixed linear models. Biometrics 51, 1429–1439 (1995)
15. Searle, S.R., Casella, G., McCulloch, C.E.: Variance Components. Wiley Series in Probability and Mathematical Statistics. Wiley, New York (1992)
16. Vangeneugden, T., Laenen, A., Geys, H., Renard, D., Molenberghs, G.: Applying linear mixed models to estimate reliability in clinical trial data with repeated measurements. Contr. Clin. Trials 25, 13–30 (2004)
17. Verbeke, G., Molenberghs, G.: Linear Mixed Models for Longitudinal Data. Springer, Heidelberg (2009)
18. Wellenius, G.A., Yeh, G.Y., Coull, B.A., Suh, H.H., Phillips, R.S., Mittlemann, M.A.: Effects of ambient air pollution on functional status in patients with chronic congestive heart failure: a repeated-measures study. Environ. Health 6(26) (2007)