



UNIVERSIDAD CARLOS III DE MADRID

working
papers

Working Paper 11-43-(32)
Statistics and Econometrics Series
December, 2011

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

ROBUST HENDERSON III ESTIMATORS OF VARIANCE COMPONENTS IN THE NESTED ERROR MODEL

Betsabé Pérez, Daniel Peña and Isabel Molina

Abstract

Common methods for estimating variance components in Linear Mixed Models include Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML). These methods are based on the strong assumption of multivariate normal distribution and it is well known that they are very sensitive to outlying observations with respect to any of the random components. Several robust alternatives of these methods have been proposed (e.g. Fellner 1986, Richardson and Welsh 1995). In this work we present several robust alternatives based on the Henderson method III which do not rely on the normality assumption and provide explicit solutions for the variance components estimators. These estimators can later be used to derive robust estimators of regression coefficients. Finally, we describe an application of this procedure to small area estimation, in which the main target is the estimation of the means of areas or domains when the within-area sample sizes are small.

Keywords: Henderson method III, Linear mixed model, robust estimators, variance components estimators.

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

Betsabé Pérez, Daniel Peña and Isabel Molina

December 23, 2011

1 Introduction

In the last decades linear mixed models have received considerable attention in the literature from a practical and theoretical point of view (e.g. McCulloch and Searle [14], Verbeke and Molenberghs [23], Demidenko [3]). These models are frequently used in many fields such as small area estimation or longitudinal studies because they model adequately the within-subject correlation typically present in these type of data. Other fields of application include clinical trials (Vangeneugden et al. [22]) and environmental studies (Wellenius et al. [24]). Despite the many different applications of these models, still diagnostic methods are not so well developed. Christensen et al. [2] studied case deletion diagnostics. Banerjee and Frees [1] studied case deletion and subject deletion diagnostics. Galpin and Zewotir [5] and [6] extended some diagnostic tools of ordinary linear regression, such as residuals, leverages and outliers to LMMs when the variances of the random factors are known. In practice, however, these variances are unknown and need to be estimated from sample data. If sample data are contaminated, then the estimation of variance components might be seriously affected and this will in turn affect all diagnostic tools. Given the importance of adequately estimating variance components, we introduce new robust estimators of variance components based on Henderson method III. This method has been chosen for three reasons; first, because it provides explicit formulas for the estimators, avoiding iterative procedures and the need for starting values and reducing the computational time; second, because it does not need any assumption on the shape of the probability of the distribution apart from the existence of first and second order moments; third, the estimation procedure consists simply of solving two standard

regression problems. These estimators can later be used to derive robust estimators of regression coefficients. Finally, we describe an application of this procedure to small area estimation, in which the main target is the estimation of the means of areas or domains when the within-area sample sizes are small.

2 Linear model with random effects

Let us consider sample data that come from D different populations groups. Suppose that there are n_d observations from group d , $d = 1, \dots, D$, where $n = \sum_{d=1}^D n_d$ is the total sample size. 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. The model at individual level is given by

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

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 group effects and errors are supposed to be independent with distributions

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

Observe that under this model, in contrast with model (3.1), the means of the observations are not affected by the group effect u_d since $E(y_{dj}) = \mathbf{x}_{dj}^T \boldsymbol{\beta}$. However, the random group effects induce a (constant) correlation between all pairs of observations in the same group, because $cov(y_{dj}, y_{dk}) = \sigma_u^2$ for $k \neq j$. Still, observations in different groups are uncorrelated. Stacking the elements of the model in columns, we obtain $\mathbf{y} = (y_{11}, y_{12}, \dots, y_{Dn_D})^T$ of size n , $\mathbf{u} = (u_1, u_2, \dots, u_D)^T$ of size D and $\mathbf{e} = (e_{11}, e_{12}, \dots, e_{Dn_D})^T$ of size n . In turn, concatenation of the predictor vectors gives the $n \times p$ matrix $\mathbf{X} = (\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{Dn_D})^T$. Additionally, we define the $n \times D$ block diagonal matrix

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdot & \vdots \\ \vdots & \cdot & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1}_{n_D} \end{pmatrix}$$

where here, $\mathbf{1}_{n_d}$ denotes a vector of ones of size n_d . Then, in matrix notation, the model can be written as

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

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

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

which means that

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_u^2 \mathbf{Z}\mathbf{Z}^T + \sigma_e^2 \mathbf{I}_n)$$

Let us define the vector of variance components $\theta = (\sigma_u^2, \sigma_e^2)^T$. When θ is known, Henderson [32] obtained the Best Linear Unbiased Estimator (BLUE) of $\boldsymbol{\beta}$ and the Best Linear Unbiased Predictor (BLUP) of \mathbf{u} , which are defined respectively as

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

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

3 Estimation of variance components

The estimator (3) and the predictor (4) depend on θ , which in practice is unknown and needs to be estimated from sample data. The empirical versions of (3) and (4), called EBLUE and EBLUP respectively, are obtained by replacing a suitable estimator $\hat{\theta}$ for θ in (3) and (4) and are given by

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

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

where the hat over \mathbf{V} indicates that θ has been replaced by its estimator $\hat{\theta}$.

Traditional methods for estimating variance components include those based on the likelihood, namely maximum likelihood (ML) and restricted/residual ML (REML), and a moments method called Henderson method III, see e.g., Searle et al. [21]. However, when outliers are present, these methods may deliver estimators with poor properties. Below we briefly review each of these methods.

3.1 Maximum likelihood

Maximum likelihood estimation is usually carried out under the assumption that \mathbf{y} has a multivariate normal distribution. Under this assumption, the joint likelihood is given by

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

The joint log-likelihood is

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

where c is denotes a constant. Using the relations

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

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

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

and equating them to zero we obtain the equations

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

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

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

Solving for $\boldsymbol{\beta}$ in (7), we obtain the ML estimating equation for $\boldsymbol{\beta}$,

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

where here \mathbf{V} depends on the ML estimator of $\theta = (\sigma_u^2, \sigma_e^2)^T$. Equations (8) and (9) do not have analytic solution and need to be solved numerically by iterative methods such as Newton-Raphson or Fisher-scoring.

3.2 Restricted maximum likelihood

A criticism of ML estimators of variance components is that they are biased downward, because they do not take into account the loss in degrees of freedom from the estimation of $\boldsymbol{\beta}$. REML method corrects for this problem by transforming \mathbf{y} into two independent vectors, $\mathbf{y}_1 = K_1 \mathbf{y}$ and $\mathbf{y}_2 = K_2 \mathbf{y}$. The probability density function of \mathbf{y}_1 does not depend on $\boldsymbol{\beta}$ and it holds $E(\mathbf{y}_1) = \mathbf{0}$, 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 . Similarly to the ML case, the obtained equations do not have analytic solutions and need to be solved using iterative techniques such as EM algorithm, Fisher-scoring or Newton-Raphson methods.

Jennrich and Schluchter [25] compared the performances of the three algorithms and noted the following: (1) direct comparison of these algorithms in terms of required computational burden is difficult, because this depend to a large degree of how efficiently the algorithms are coded. (2) Newton-Raphson algorithm, with a quadratic convergence rate, generally converges in a small number of iterations, with a higher cost per iteration. (3) EM method has the lowest cost per iteration, but at times requires a large number of iterations. (4) Fisher-scoring algorithm is intermediate in terms of cost per iteration and required number of iterations. However, its cost per iteration is often not much smaller than that of Newton-Raphson algorithm, whereas Fisher-scoring algorithm sometimes requires a considerably larger number of iterations than Newton-Raphson algorithm. Lindstrom and Bates [26] provided arguments favoring the use of Newton-Raphson method.

3.3 Henderson method III

ML and REML estimators of θ are usually based on the assumption that the vector \mathbf{y} has a multivariate normal distribution, although they remain consistent even when normality is not satisfied exactly under some regularity conditions (Jiang, [10]). An alternative method which does not rely on normality and provides explicit formulas for the estimators of the variance components is Henderson method III (H3). This method works as follows. First, consider a 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 $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ and define the full model as

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

The partition in sum of squares of model (10) is given by

$$\begin{aligned} \text{SSR}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= \mathbf{y}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X} \mathbf{y}, \\ \text{SSE}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= \mathbf{e}^T \mathbf{e} = [(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X})\mathbf{y}]^T [(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X})\mathbf{y}], \\ \text{SST}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= \mathbf{y}^T \mathbf{y}, \end{aligned} \quad (11)$$

with their corresponding expected values given by

$$\begin{aligned}
E[\text{SSR}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)] &= \text{tr} \left\{ \begin{pmatrix} \mathbf{X}_1^T \mathbf{X}_1 & \mathbf{X}_1^T \mathbf{X}_2 \\ \mathbf{X}_2^T \mathbf{X}_1 & \mathbf{X}_2^T \mathbf{X}_2 \end{pmatrix} E(\boldsymbol{\beta} \boldsymbol{\beta}^T) \right\} + \text{rank}(\mathbf{X}) \sigma_e^2, \\
E[\text{SSE}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)] &= [n - \text{rank}(\mathbf{X})] \sigma_e^2, \\
E[\text{SST}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)] &= \text{tr} \left\{ \begin{pmatrix} \mathbf{X}_1^T \mathbf{X}_1 & \mathbf{X}_1^T \mathbf{X}_2 \\ \mathbf{X}_2^T \mathbf{X}_1 & \mathbf{X}_2^T \mathbf{X}_2 \end{pmatrix} E(\boldsymbol{\beta} \boldsymbol{\beta}^T) \right\} + n \sigma_e^2.
\end{aligned} \tag{12}$$

Now consider the reduced model with only $\boldsymbol{\beta}_1$,

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\epsilon}. \tag{13}$$

Analogously, the partition in sum of squares of model (13) is given by

$$\begin{aligned}
\text{SSR}(\boldsymbol{\beta}_1) &= \mathbf{y}^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1 \mathbf{y}, \\
\text{SSE}(\boldsymbol{\beta}_1) &= \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = [(\mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1) \mathbf{y}]^T [(\mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1) \mathbf{y}], \\
\text{SST}(\boldsymbol{\beta}_1) &= \mathbf{y}^T \mathbf{y},
\end{aligned} \tag{14}$$

with their corresponding expected values

$$\begin{aligned}
E[\text{SSR}(\boldsymbol{\beta}_1)] &= \text{tr} \left\{ \begin{pmatrix} \mathbf{X}_1^T \mathbf{X}_1 & \mathbf{X}_1^T \mathbf{X}_2 \\ \mathbf{X}_2^T \mathbf{X}_1 & \mathbf{X}_2^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2 \end{pmatrix} E(\boldsymbol{\beta} \boldsymbol{\beta}^T) \right\} + \text{rank}(\mathbf{X}_1) \sigma_e^2, \\
E[\text{SSE}(\boldsymbol{\beta}_1)] &= \text{tr} \{ \mathbf{X}^T [\mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T]^T [\mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T] \mathbf{X} E(\boldsymbol{\beta} \boldsymbol{\beta}^T) \} \\
&\quad + [n - \text{rank}(\mathbf{X})] \sigma_e^2, \\
E[\text{SST}(\boldsymbol{\beta}_1)] &= \text{tr} \left\{ \begin{pmatrix} \mathbf{X}_1^T \mathbf{X}_1 & \mathbf{X}_1^T \mathbf{X}_2 \\ \mathbf{X}_2^T \mathbf{X}_1 & \mathbf{X}_2^T \mathbf{X}_2 \end{pmatrix} E(\boldsymbol{\beta} \boldsymbol{\beta}^T) \right\} + n \sigma_e^2.
\end{aligned} \tag{15}$$

The reduction in sum of squares due to introducing \mathbf{X}_2 in the model with only \mathbf{X}_1 is

$$\text{SSR}(\boldsymbol{\beta}_2 | \boldsymbol{\beta}_1) = \text{SSR}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - \text{SSR}(\boldsymbol{\beta}_1). \tag{16}$$

The expectation of this reduction is given by

$$\begin{aligned}
E[\text{SSR}(\boldsymbol{\beta}_2 | \boldsymbol{\beta}_1)] &= \text{tr} \{ \mathbf{X}_2^T [\mathbf{I}_n - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T] \mathbf{X}_2 E(\boldsymbol{\beta} \boldsymbol{\beta}^T) \} \\
&\quad + [\text{rank}(\mathbf{X}) - \text{rank}(\mathbf{X}_1)] \sigma_e^2.
\end{aligned} \tag{17}$$

Now consider model (1) 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 equates the sum of squares $\text{SSR}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ in (14) and $\text{SSR}(\boldsymbol{\beta}_2 | \boldsymbol{\beta}_1)$ in (16) to their expectations in (12) and (17) respectively, 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 [21], chapter 5). Let $\hat{\mathbf{e}}$ and $\hat{\boldsymbol{\varepsilon}}$ be the vectors of residuals obtained by fitting the two models (10) and (13) respectively, considering $\boldsymbol{\beta}_2$ as fixed. 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

$$\hat{\sigma}_{e,H3}^2 = \frac{\sum_{d=1}^D \sum_{j=1}^{n_d} \hat{e}_{dj}^2}{n - p - D}, \quad \hat{\sigma}_{u,H3}^2 = \frac{\sum_{d=1}^D \sum_{j=1}^{n_d} \hat{\varepsilon}_{dj}^2 - \hat{\sigma}_e^2(n - p)}{\text{tr}\{\mathbf{Z}^T[\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]\mathbf{Z}\}}, \quad (18)$$

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 corresponding in model (13).

4 Diagnostic methods

Limited work has been done on diagnostic methods for linear mixed models. Christensen et al. [2] considered the case deletion diagnostics and Galpin and Zewotir [6] provided a definition of residuals, leverages and outliers when some variance components are known.

Fitted values of the response variable are

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}},$$

and residuals are then

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{R}\mathbf{y},$$

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

Studentized residuals (internal studentization):

$$t_{dj} = \frac{\hat{e}_{dj}}{\sqrt{\text{var}(\hat{e}_{dj})}} = \frac{\hat{e}_{dj}}{\hat{\sigma}_e\sqrt{r_{dj}}}$$

where r_{dj} is the dj -th diagonal element of matrix \mathbf{R} and e_{dj} is the dj -th element of vector $\hat{\mathbf{e}} = \mathbf{R}\mathbf{y}$.

Studentized residuals (external studentization): Let $\hat{\sigma}_{e(dj)}$ denote the estimate of σ_e when the dj -th observation is deleted. If $\hat{\sigma}_{e(dj)}^2$ is used in place of $\hat{\sigma}_e^2$ we obtain the dj -th externally Studentized residual, given by

$$t_{dj}^* = \frac{\hat{e}_{dj}}{\hat{\sigma}_{e(dj)}\sqrt{r_{dj}}}.$$

The estimator t_{dj}^* satisfies that $t_{dj}^{*2} \sim \frac{n-1}{n-p-1}F(1, n-p-1)$ where $F(1, n-p-1)$ is an F -distribution with degrees of freedom 1 and $(n-p-1)$.

Note that element the r_{dj} used to standardized residuals depends on the variance components σ_e^2 and σ_u^2 , which are unknown. When there are outliers, these might affect the estimators of variance components, and these estimators will change the distribution of standardized residuals.

To illustrate this, we have simulated data from model (1), with $D = 15$ groups and total sample size $n = 2500$. The theoretical values of the variance components are $\sigma_e^2 = 0.5$ and $\sigma_u^2 = 0.5$. In order to increase the estimator of the error variance σ_e^2 , we introduced atypical data on \mathbf{y} as mean shifts, by increasing the values of the some of the response values by k times the theoretical standard deviation with $k = 5$. Index plots of internally studentized residuals, using the true variance components and the estimated ones, appear in the left and right panels of Figure 1 respectively. This example illustrates how the estimation of variance components affect the studentized residuals. On the right plot obtained with estimated variances, all residuals appear in the interval $(-2.5, 2.5)$; as a consequence, using the standard rule applied to these residuals, outlying observations will not be detected.

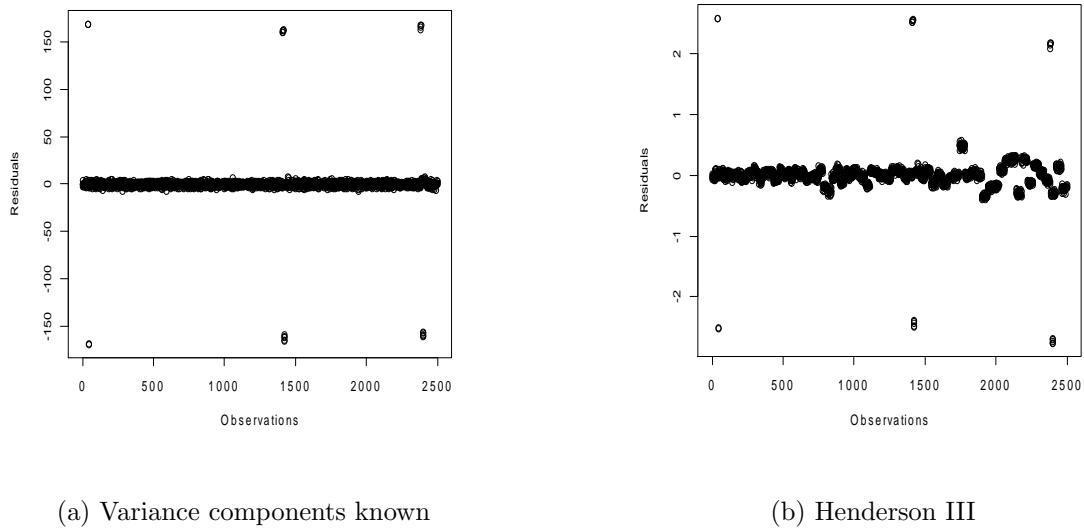


Figure 1: Internally studentized residuals (a) using the true variance components and (b) when they are estimated using H3 method.

Leverage effect in the nested-error model

Assuming that θ is known, the vector of predicted values is

$$\tilde{\mathbf{y}} = (\mathbf{I} - \mathbf{R})\mathbf{y} \quad (19)$$

This relation evokes the definition of the Hat matrix, as

$$\mathbf{H}_{\tilde{\mathbf{y}}} = \mathbf{I} - \mathbf{R}.$$

The diagonal elements $(1 - r_{dj})$ of this matrix are measures of the leverage effect of the observations and are called *leverages*. Galpin and Zewotir [6] proposed the use of the r_{dj} s to identify influential observations. If r_{dj} approaches zero, this indicates that the corresponding observation has a large leverage effect.

Due to the grouped data structure in linear mixed models with one random factor, it seems more relevant to study the leverage effect of groups instead of that of isolated observations. The leverage effect of group d is defined here as

$$h_d = \bar{\mathbf{x}}_d^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \bar{\mathbf{x}}_d, \quad , \quad d = 1, \dots, D \quad (20)$$

where $\bar{\mathbf{x}}_d = n_d^{-1} \sum_{j=1}^{n_d} \mathbf{x}_{dj}$. In practice, \mathbf{V} could be estimated using the robust variance components estimators described in the next section.

5 Robust Henderson method III

Consider the linear regression model with random effects given in (1). The estimators of variance components obtained by Henderson method III (H3 estimators) are given by

$$\hat{\sigma}_{e,H3}^2 = \frac{\sum_{d=1}^D \sum_{j=1}^{n_d} \hat{e}_{dj}^2}{n - (p + D)}, \quad \hat{\sigma}_{u,H3}^2 = \frac{\sum_{d=1}^D \sum_{j=1}^{n_d} \hat{\varepsilon}_{dj}^2 - \hat{\sigma}_e^2(n - p)}{\text{tr} \{ \mathbf{Z}^T [\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{Z} \}}, \quad (21)$$

where \hat{e}_{dj} is the residual corresponding to observation $(\mathbf{x}_{dj}^T, y_{dj})$ in the full model (10) with group effects assumed to be fixed and $\hat{\varepsilon}_{dj}$ is the corresponding residual in the reduced model (13).

Remark 1 *Henderson III estimators are scale equivariant, that is,*

$$\hat{\sigma}_{e,H3}(cy) = |c| \sigma_{e,H3}(y) \quad \text{and} \quad \hat{\sigma}_{u,H3}(cy) = |c| \sigma_{u,H3}(y).$$

The estimator $\hat{\sigma}_{e,H3}^2$ can be expressed as

$$\hat{\sigma}_{e,H3}^2 = \hat{\sigma}_{e,H3}^2(\mathbf{y}) = \frac{SSE(\boldsymbol{\beta}^*)}{n - \text{rank}(\mathbf{X}^*)} = \frac{\mathbf{y}^T(\mathbf{I}_n - \mathbf{H}^*)\mathbf{y}}{n - (p + D)}$$

where $\mathbf{H}^* = \mathbf{X}^*(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}$, $\mathbf{X}^* = (\mathbf{X}|\mathbf{Z})$ and $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^T, \mathbf{u}^T)^T$.

Then,

$$\begin{aligned} \hat{\sigma}_{e,H3}(c\mathbf{y}) &= \sqrt{\frac{(c\mathbf{y})^T(\mathbf{I}_n - \mathbf{H}^*)(c\mathbf{y})}{n - (p + D)}} \\ &= \sqrt{\frac{c^2\mathbf{y}^T(\mathbf{I}_n - \mathbf{H}^*)\mathbf{y}}{n - (p + D)}} \\ &= |c| \sqrt{\frac{\mathbf{y}^T(\mathbf{I}_n - \mathbf{H}^*)\mathbf{y}}{n - (p + D)}} \\ &= |c| \hat{\sigma}_{e,H3}(\mathbf{y}) \end{aligned}$$

Therefore, the estimator $\hat{\sigma}_{e,H3}$ is scale invariant. Now we check that $\hat{\sigma}_{u,H3}$ is also scale equivariant.

The estimator $\hat{\sigma}_{u,H3}^2$ is given by

$$\hat{\sigma}_{u,H3}^2 = \hat{\sigma}_{u,H3}^2(y) = \frac{SSE(\boldsymbol{\beta}) - \hat{\sigma}_{e,H3}^2(n - p)}{\text{tr}[\mathbf{Z}^T(\mathbf{I}_n - \mathbf{H})\mathbf{Z}]} = \frac{\mathbf{y}(\mathbf{I}_n - \mathbf{H})\mathbf{y} - \left[\frac{\mathbf{y}^T(\mathbf{I}_n - \mathbf{H}^*)\mathbf{y}}{n - (p + D)} \right] (n - p)}{\text{tr}[\mathbf{Z}^T(\mathbf{I}_n - \mathbf{H})\mathbf{Z}]}$$

denoting $m = \text{tr}[\mathbf{Z}^T(\mathbf{I}_n - \mathbf{H})\mathbf{Z}]$

$$\hat{\sigma}_{u,H3}^2 = \frac{1}{m} \left\{ \mathbf{y}(\mathbf{I}_n - \mathbf{H})\mathbf{y} - \frac{n - p}{n - (p + D)} \mathbf{y}^T(\mathbf{I}_n - \mathbf{H}^*)\mathbf{y} \right\}$$

thus,

$$\hat{\sigma}_{u,H3}(\mathbf{y}) = \sqrt{\frac{1}{m} \left\{ \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y} - \frac{n - p}{n - (p + D)} \mathbf{y}^T(\mathbf{I}_n - \mathbf{H}^*)\mathbf{y} \right\}}$$

Then,

$$\begin{aligned} \hat{\sigma}_{u,H3}(c\mathbf{y}) &= \sqrt{\frac{1}{m} \left\{ (c\mathbf{y})^T(\mathbf{I}_n - \mathbf{H})(c\mathbf{y}) - \frac{n - p}{n - (p + D)} (c\mathbf{y})^T(\mathbf{I}_n - \mathbf{H}^*)(c\mathbf{y}) \right\}} \\ &= \sqrt{\frac{c^2}{m} \left\{ \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y} - \frac{n - p}{n - (p + D)} \mathbf{y}^T(\mathbf{I}_n - \mathbf{H}^*)\mathbf{y} \right\}} \\ &= |c| \sqrt{\frac{1}{m} \left\{ \mathbf{y}^T(\mathbf{I}_n - \mathbf{H})\mathbf{y} - \frac{n - p}{n - (p + D)} \mathbf{y}^T(\mathbf{I}_n - \mathbf{H}^*)\mathbf{y} \right\}} \\ &= |c| \hat{\sigma}_{u,H3}(\mathbf{y}) \end{aligned}$$

Therefore, the estimator $\hat{\sigma}_{u,H3}$ is scale invariant.

Let us express Henderson III estimators in terms of the means of squared residuals

$$\hat{\sigma}_{e,H3}^2 = \frac{n \left[\sum_{d=1}^D \sum_{j=1}^{n_d} \hat{e}_{dj}^2 / n \right]}{n - (p + D)}, \quad \hat{\sigma}_{u,H3}^2 = \frac{n \left[\sum_{d=1}^D \sum_{j=1}^{n_d} \hat{\varepsilon}_{dj}^2 / n \right] - \hat{\sigma}_e^2 (n - p)}{\text{tr} \{ \mathbf{Z}^T [\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{Z} \}}, \quad (22)$$

We propose to robustify these estimators using, first, robust methods to fit the two models (10) and (13) and, after that, replacing in (22) the means of squared residuals by other robust functions.

Model (13) 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 [30]. Model (10) is a model with fixed group effects, which can be robustly fitted using an adaptation of the principal sensibility components method of Peña and Yohai [30] to the grouped data structure. An alternative approach is the M-S estimation of Maronna and Yohai [31].

These fitting methods will provide better residuals \hat{e}_{dj} and $\hat{\varepsilon}_{dj}$, which are in turn used to find robust estimators of the variance components. Below we describe different estimators based on robust functions of these new residuals.

MADH3 estimators: In the two estimators given in (22), we substitute the means of squared residuals by the square of the normalized medians of absolute deviations (MAD), given by

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

where $\hat{\xi}_{dj}$ is the residual of observation $(\mathbf{x}_{dj}^T, y_{dj})$ under the corresponding fitted model, either (10) or (13).

TH3 estimators: Trimming consists of giving zero weight to a percentage of extreme cases. In this case, in the two equations given in (22) we trim residuals that are outside the interval (b_1, b_2) with

$$b_1 = q_1 - k(q_3 - q_1) \quad \text{and} \quad b_2 = q_3 + k(q_3 - q_1). \quad (23)$$

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 residuals appearing in (22) according to an appropriate smoothing function. Here we consider Tukey's biweight function, given by

$$\varphi(x) = x[1 - (x/k)^2]^2, \quad \text{if } |x| \leq k. \quad (24)$$

In this case, the robust Henderson III estimators are given by

$$\hat{\sigma}_{e,RH3}^2 = \frac{\sigma_{e,MAD}^2 \sum_{d=1}^D \sum_{j=1}^{n_d} \varphi^2(\hat{e}_{dj}/\sigma_{e,MAD})}{n - (p + D)}, \quad (25)$$

$$\hat{\sigma}_{u,RH3}^2 = \frac{\sigma_{u,MAD}^2 \sum_{d=1}^D \sum_{j=1}^{n_d} \varphi^2(\hat{e}_{dj}/\sigma_{u,MAD}) - \hat{\sigma}_{e,RH3}^2(n - p)}{\text{tr}\{\mathbf{Z}^T(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{Z}\}}. \quad (26)$$

Remark 2 The function $h(x) = \sigma_x \varphi(x/\sigma_x)$ is scale invariant, where σ_x is a scale such that $\sigma_{cx} = c\sigma_x, c > 0$. If we consider $\sigma_x = MAD(x)$, let us verify that

$$MAD(cx) = cMAD(x), \quad c > 0.$$

By definition $MAD(x) = 1.4826 \text{ median}(|x - \text{median}(x)|)$

$$\begin{aligned} MAD(cx) &= 1.4826 \text{ median}(|(cx) - \text{median}(cx)|) \\ &= 1.4826 \text{ median}(|c|(x - \text{median}(x))) \\ &= |c|[1.4826 \text{ median}(|x - \text{median}(x)|)] \\ &= |c|MAD(x). \end{aligned}$$

Since $\sigma_{cx} = c\sigma_x$, we have that

$$h(cx) = c\sigma_x \psi\left(\frac{cx}{c\sigma_x}\right) = c\sigma_x \psi\left(\frac{x}{\sigma_x}\right) = h(x).$$

Remark 3 RH3 estimators of σ_e^2 and σ_u^2 are scale invariant.

Consider the estimator $\hat{\sigma}_{e,RH3}^2$

$$\hat{\sigma}_{e,RH3}^2 = \frac{\sigma_{e,MAD}^2 \sum_{d=1}^D \sum_{j=1}^{n_d} \varphi^2(\hat{e}_{dj}/\sigma_{e,MAD})}{n - (p + D)} = \sqrt{\frac{\sum_{d=1}^D \sum_{j=1}^{n_d} h^2(\hat{e}_{dj})}{n - (p + D)}},$$

where $h(\cdot)$ is scale invariant. Therefore, $\hat{\sigma}_{e,RH3}$ is scale invariant.

Let $m = \text{tr}\{\mathbf{Z}^T(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{Z}\}$. The estimator $\hat{\sigma}_{u,RH3}^2$ is given by

$$\begin{aligned}\hat{\sigma}_{u,RH3}^2 &= \frac{1}{m} \left\{ \sigma_{\varepsilon,MAD}^2 \sum_{d=1}^D \sum_{j=1}^{n_d} \varphi^2 \left(\frac{\hat{\varepsilon}_{dj}}{\sigma_{\varepsilon,MAD}} \right) - \hat{\sigma}_{e,RH3}^2(n-p) \right\} \\ &= \frac{1}{m} \left\{ \sum_{d=1}^D \sum_{j=1}^{n_d} h^2(\hat{\varepsilon}_{dj}) - \frac{\sum_{d=1}^D \sum_{j=1}^{n_d} h^2(\hat{\varepsilon}_{dj})}{n-(p+D)}(n-p) \right\}.\end{aligned}$$

Similarly, since $h(\cdot)$ is scale invariant, $\hat{\sigma}_{u,RH3}$ is scale invariant.

5.1 Simulation experiment

This section describes a Monte Carlo simulation study that compares the robust estimators of the variance components with the traditional non-robust ones. For this, we generated data coming from $D = 10$ groups. The group sample sizes n_d , $d = 1, \dots, D$ 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 $L = 500$ Monte Carlo replicates. In each iteration, we generated group effects as $u_d \stackrel{iid}{\sim} N(0, \sigma_u^2)$ with $\sigma_u^2 = 0.25$. Similarly, we generated errors as $e_{dj} \stackrel{iid}{\sim} N(0, \sigma_e^2)$ with $\sigma_e^2 = 0.25$. Then we generated the model responses y_{dj} , $j = 1, \dots, n_d$, $d = 1, \dots, D$, 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 $\mathcal{D}_c \subseteq \{1, 2, \dots, D\}$ 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

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

Method	Estimators		Bias		MSE	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
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

increases the within group variability σ_e^2 .

Then, we calculated the traditional estimators H3, ML and REML, and the proposed robust estimators, MADH3, TH3 and RH3. After the $L = 500$ replicates, we computed the empirical bias and mean squared error (MSE) of the estimators.

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 estimators resist quite well.

5.2 Discussion

This part introduces three robust versions of H3 estimators called MADH3, TH3 and RH3 estimators. These estimators are obtained by first, fitting in a robust way the two submodels (10) and (13) and, then, replacing the means of squared residuals in H3 estimators by other robust functions of the residuals coming from those robust fittings.

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

Method	Estimators		Bias		MSE	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
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	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
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	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
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	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_e^2$
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

In simulations we have analyzed the robustness of our proposed estimators under two different contamination scenarios: when the between groups variability is increased by including a mean shift in some of the groups, and when the within group variability is increased by introducing given percentages of outliers within the groups. The new robust estimator RH3 achieves great efficiency under both types of contamination and at the same time preserves good efficiency when there is not contamination.

6 Robust estimation of regression coefficients

This section deals with robust estimation of regression coefficients using the estimators of variance components introduced above. These estimators are then used to derive robust predictors of the means in small areas.

6.1 Small area estimators

Small area estimation is usually done under the setup of finite population. Thus, we have a population U of size N that is assumed to be partitioned into D subpopulations U_1, \dots, U_D of sizes N_1, \dots, N_D called small areas. Particular quantities of interest are the means of the small areas,

$$\bar{Y}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} y_{dj}, \quad d = 1, \dots, D$$

A sample s_d of size n_d is drawn from U_d , $d = 1, \dots, D$. We assume that the model holds for all population units, that is, for units in the sample and out of the sample. Under this setup, the target area means are random. Therefore, is it common to say predicting \bar{Y}_d rather than estimating \bar{Y}_d . The mean of small area d can be split into two terms, one for the sample elements and the other for the out-of-sample elements, obtaining a linear combination of the sample mean \bar{y}_{s_d} and the out-of-sample mean $\bar{y}_{s_d^c}$.

$$\bar{Y}_d = \frac{1}{N_d} \left[\sum_{j \in s_d} y_{dj} + \sum_{j \in s_d^c} y_{dj} \right] = \frac{n_d}{N_d} \bar{y}_{s_d} + \left(\frac{n_d}{N_d} \bar{y}_{s_d^c} \right), \quad d = 1, \dots, D$$

When studying outliers in finite population inference, the existing literature is developed exclusively under one of the following assumptions:

Assumption 1. Non representative outliers: We assume that atypical observations appear only in the sample but not in the non-sample part of the population. Then, it seems natural to project the working model into the entire non-sampled part of the population. Chambers [27] call these type of outliers non-representative outliers. In this case, the appropriate methods for estimating model parameters are called *Robust Projective*, meaning that they project sample non-outlier behavior on to the non-sampled part of the population.

Assumption 2. Representative outliers We assume that atypical observations appear in the sample and non-sample part of the population. In this case, robust projective methods will provide biased estimators of the small area means; therefore, it is necessary to correct for this bias using an appropriate correction factor.

Next section introduces two robust projective methods given in the literature, Fellner's approach and Sinha and Rao's procedure.

6.2 Previous robust procedures

6.2.1 Fellner's approach

Fellner [4] derived robust estimators of variance components and regression coefficients $\boldsymbol{\beta}$, together with a robust predictor of \mathbf{u} , which could in turn be used to derive a robust EBLUP. The joint probability density function of \mathbf{y} is given by

$$f(\boldsymbol{\beta}, \theta | \mathbf{y}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}. \quad (27)$$

Similarly, the joint density function of $\mathbf{u} = (u_1, \dots, u_D)^T$ is

$$g(\mathbf{u}; \sigma_u^2) = (2\pi\sigma_u^2)^{-D/2} \exp\{-\mathbf{u}^T \mathbf{u} / 2\}.$$

Assuming θ known, the BLUE of $\boldsymbol{\beta}$ and the BLUP of \mathbf{u} can be obtained simultaneously by maximizing the joint loglikelihood of \mathbf{y} and \mathbf{u} , $\ln f(\boldsymbol{\beta}, \theta | \mathbf{y}, \mathbf{u}) = \ln f(\theta | \mathbf{y}) + \ln g(\mathbf{u})$, with respect to $\boldsymbol{\beta}$ and \mathbf{u} . The resulting system of normal equations is given by

$$\begin{bmatrix} \mathbf{X}^T \mathbf{X} / \sigma_e^2 & \mathbf{X}^T \mathbf{Z} / \sigma_e^2 \\ \mathbf{Z}^T \mathbf{X} / \sigma_e^2 & \mathbf{I} / \sigma_u^2 + \mathbf{Z}^T \mathbf{Z} / \sigma_e^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{y} / \sigma_e^2 \\ \mathbf{Z}^T \mathbf{y} / \sigma_e^2 + (\mathbf{I} / \sigma_u^2) \mathbf{0}_D \end{bmatrix}$$

Fellner's method is based in the idea of replacing in these equations, observations y_{di} and random effects u_d that are far from their predicted values $\hat{y}_{di} = \mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}} + \hat{u}_d$ and \hat{u}_d by what

he called pseudo-observations. More explicitly, Fellner's method solves the system

$$\begin{bmatrix} \mathbf{X}^T \mathbf{X} / \sigma_e^2 & \mathbf{X}^T \mathbf{Z} / \sigma_e^2 \\ \mathbf{Z}^T \mathbf{X} / \sigma_e^2 & I / \sigma_u^2 + \mathbf{Z}^T \mathbf{Z} / \sigma_e^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{y}^* / \sigma_e^2 \\ \mathbf{Z}^T \mathbf{y}^* / \sigma_e^2 + (\mathbf{I} / \sigma_u^2) \mathbf{0}_D^* \end{bmatrix}, \quad (28)$$

where $\mathbf{y}^* = (y_{di}^*, i = 1, \dots, n_d, d = 1, \dots, D)$ with $y_{di}^* = \mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}} + \hat{u}_d + \sigma_e \psi(\hat{e}_{dj} / \sigma_e)$ and $\mathbf{0}_D^* = (\hat{u}_d - \sigma_u \phi(\hat{u}_d / \sigma_u); d = 1, \dots, D)$ and ψ is an odd, monotonic and bounded function such as Huber's psi function.

Equations (28) assume that variance components are known, but Fellner [4] also gave REML equations for variance components which, solved jointly with (28), yield also a robust estimator of $\boldsymbol{\beta}$ together with a robust predictor of \mathbf{u} . For this, he proposes to robustify REML equations in the form

$$\begin{aligned} \hat{\sigma}_u^2 &= \{h(D - v^*)\}^{-1} \hat{\sigma}_u \sum_{d=1}^D \psi^2(\hat{u}_d / \hat{\sigma}_u), \\ \hat{\sigma}_e^2 &= \{h(n - p - D + v^*)\}^{-1} \hat{\sigma}_e \sum_{d=1}^D \psi^2(\hat{e}_{dj} / \hat{\sigma}_e), \end{aligned}$$

where h is an appropriately chosen constant to adjust for the bias in $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ at the normal distribution. This leads to $h = E\{\psi^2(X)\}$, where $X \sim N(0, 1)$.

6.2.2 REBLUP estimators

Sinha and Rao [28] proposed a two-step procedure for constructing robust estimators of model parameters. The steps of the procedure are the following:

- **Step 1.** The estimators $\hat{\boldsymbol{\beta}}^{SR}$ and $\hat{\theta}^{SR}$ are obtained simultaneously based on robustified ML equations.
- **Step 2.** The predictor $\hat{\mathbf{u}}^{SR}$ is obtained using the estimators of Step 1.

In Step 1, the ML equations for $\boldsymbol{\beta}$ and θ are defined by

$$\begin{aligned} \mathbf{X}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) &= \mathbf{0}, \\ (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_\ell} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) - \text{tr} \left\{ \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_\ell} \right\} &= \mathbf{0}, \quad \ell = 1, 2, \end{aligned}$$

where θ_ℓ is the ℓ -th element of $\theta = (\sigma_u^2, \sigma_e^2)^T$.

If some fitted values $\hat{y}_{dj} = \mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}}$ are unusually different from the corresponding observed values y_{dj} , then we have the indication of apparent outliers in the data. To handle outliers in the response values, they proposed robustified ML equations in the form

$$\begin{aligned} \mathbf{X}^T \mathbf{V}^{-1} U^{\frac{1}{2}} \Psi(\mathbf{r}) &= \mathbf{0}, \\ \Psi(\mathbf{r})^T U^{\frac{1}{2}} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_\ell} \mathbf{V}^{-1} U^{\frac{1}{2}} \Psi(\mathbf{r}) - \text{tr} \left\{ K \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_\ell} \right\} &= \mathbf{0}, \quad \ell = 1, 2, \end{aligned}$$

where

$$\mathbf{r} = U^{-\frac{1}{2}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad U = \text{diag}(\mathbf{V}), \quad K = E\{\psi_b^2(X)\} \mathbf{I}_n \text{ with } X \sim N(0, 1), \quad \Psi(u) = (\psi_b(u_1), \psi_b(u_2), \dots)^T \text{ with } \psi_b(u) = u \cdot \min(1, \frac{b}{|u|}) \text{ and } b = 1.345.$$

The complete algorithm for robust estimation of $\boldsymbol{\beta}$ and θ is:

1. Choose starting values $\boldsymbol{\beta}^{(0)}$ and $\theta^{(0)}$. Set $m = 0$.
2. (a) Calculate $\boldsymbol{\beta}^{(m+1)}$. (b) Calculate $\theta^{(m+1)}$. (c) Set $m = m + 1$.
3. Repeat until convergence is achieved. Denote the estimates at convergence as $\hat{\boldsymbol{\beta}}^{SR}$ and $\hat{\theta}^{SR}$.

In Step 2, the predictor $\hat{\mathbf{u}}^{SR}$ is obtained using the estimators of $\boldsymbol{\beta}$ and θ obtained in Step 1 and solving the following robustified equation

$$\hat{\sigma}_e \mathbf{Z}^T \Psi \{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})/\hat{\sigma}_e\} - \hat{\sigma}_u \Psi(\mathbf{u}/\hat{\sigma}_u) = 0$$

Sinha and Rao [28] proposed to solve this equation using the Newton-Raphson method. Finally, the Robust EBLUPs (REBLUPs) of the small area means are given by

$$\hat{Y}_d^{SR} = \frac{1}{N_d} \left(\sum_{j \in s_d} y_{dj} + \sum_{j \in s_d^c} \hat{y}_{dj}^{SR} \right), \quad d = 1, \dots, D$$

where $\hat{y}_{dj}^{SR} = \mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}}^{SR} + \hat{u}_d^{SR}$.

Some comments

The Newton-Raphson procedure is a commonly used iterative method for the solution of nonlinear equations. To solve the equation $h(t) = 0$, at each iteration the function h is linearized in the sense that it is replaced by its Taylor expansion of order one about

the current approximation. Let us denote by t_m the m -th approximation. Then the next value is the solution of

$$h(t^m) + h'(t^m)(t^{m+1} - t^m) = 0$$

that is,

$$t^{m+1} = t^m - \frac{h(t^m)}{h'(t^m)}$$

If the procedure converges, the convergence is very fast; but it is not guaranteed to converge. If h' is not bounded away from zero, the denominator may become very small, making the sequence t^m unstable unless the initial value t^0 is very near to the solution (Maronna et al., [29]).

6.3 Procedure using RH3

We propose a two-step procedure that provides robust estimators of model parameters based on the robust estimators of variance components given in (5).

- **Step 1.** Obtain the estimator $\hat{\theta}^{RH3}$ using the robustified version of Henderson Method III given in (25) and (26).
- **Step 2.** Obtain the estimator $\hat{\beta}^{RH3}$ and the predictor $\hat{\mathbf{u}}^{RH3}$ similarly as in Sinha and Rao [28], solving the robustified normal equations (28).

Then, the new robust EBLUPs, called here RH3-EBLUPs of the small area means are given by

$$\hat{Y}_d^{RH3} = \frac{1}{N_d} \left(\sum_{j \in s_d} y_{dj} + \sum_{j \in s_d^c} \hat{y}_{dj}^{RH3} \right), \quad d = 1, \dots, D$$

where $\hat{y}_{dj}^{RH3} = \mathbf{x}_{dj}^T \hat{\beta}^{RH3} + \hat{u}_d^{RH3}$.

6.4 Simulation experiment

In this simulation study we generated data coming from $D = 30$ groups. Concerning the group sample sizes, half of them were taken of size $n_d = 10$ and the other half of size $n_d = 20$, with a total sample size of $n = 450$. 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.

More concretely, the values of the four auxiliary variables were generated respectively as $X_1 \sim N(3.31, 0.68)$, $X_2 \sim N(1.74, 1.23)$, $X_3 \sim N(1.70, 1.65)$ and $X_4 \sim N(2.41, 2.61)$.

The number of Monte Carlo samples was $L = 200$. In each replicate, group effects were generated as $u_d \stackrel{iid}{\sim} N(0, \sigma_u^2)$ with $\sigma_u^2 = 1$. Similarly, individual errors were generated as $e_{dj} \stackrel{iid}{\sim} N(0, \sigma_e^2)$ with $\sigma_e^2 = 1$. Finally, model responses y_{dj} , $j = 1, \dots, n_d$, $d = 1, \dots, D$, were generated from model (1). Using each Monte Carlo sample, the two models (10) and (13) were fitted robustly using respectively the M-S estimator of Maronna and Yohai [31] and the PSC method of Peña and Yohai [30]. We assume that outliers are representative and use the correction factor proposed by Joingo et al. [?]. Firstly, data are generated without contamination. After that, contamination is introduced according to the following scenarios:

- **Type 0.** No contamination
- **Type 1.** Outlying areas: For each selected outlying domain, we substitute all their sample observations y_{dj} by the constant $C_1 = \bar{Y}_d + c \cdot \sqrt{\frac{\sum_{j=1}^{N_d} (y_{dj} - \bar{Y}_d)^2}{N_d}}$, where $c = 4$ and $\bar{Y}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} y_{dj}$.
- **Type 2.** Outlying individuals within areas: We replace some observations within selected domains by C_1 and some others by $C_2 = \bar{Y}_d - c \cdot \sqrt{\frac{\sum_{j=1}^{N_d} (y_{dj} - \bar{Y}_d)^2}{N_d}}$.

To compare several predictors of the prediction of the small area means, we use the following measures averaged over areas

Average Absolute Relative Bias (\overline{ARB}):

$$\overline{ARB} = \frac{1}{D} \sum_{d=1}^D \left| \frac{1}{L} \sum_{t=1}^L \left(\frac{\hat{Y}_d - \bar{Y}_d}{\bar{Y}_d} \right) \right|$$

Average Relative Root MSE (\overline{RRMSE}):

$$\overline{RRMSE} = \frac{1}{D} \sum_{d=1}^D \frac{\overline{MSE}(\hat{Y}_d)^{\frac{1}{2}}}{\bar{Y}_d}$$

Method	Bias		MSE	
	σ_u^2	σ_e^2	σ_u^2	σ_e^2
ML	-0,044	0,070	0,160	0,125
RML	-0,125	0,141	0,247	0,195
RH3	-0,174	0,075	0,279	0,142

Table 6: Scenario Type 0: No contamination

Parameter	ML		REML		RH3	
	Bias	MSE	Bias	MSE	Bias	MSE
β_0	-0,037	0,264	-0,033	0,312	-0,034	0,321
β_1	0,316	0,014	0,314	0,015	0,312	0,014
β_2	0,001	0,012	0,001	0,013	0,003	0,013
β_3	-0,007	0,004	-0,006	0,005	-0,008	0,005

Table 7: Scenario Type 0: No contamination

6.5 Discussion

In this part we compare two ways to estimate regression coefficients in the linear with random effects. Then, these estimators were used to derive robust predictors of the means in small areas. Our simulation studies show that the new robust procedure RH3 gets the best results in the case of outlying areas at the same time good efficiency when there is not contamination.

Method	\overline{ARB}	\overline{RRMSE}
EBLUP	0,3667	0,3825
REBLUP	0,4015	0,5056
RH3-EBLUP	0,3843	0,4884

Table 8: Scenario Type 0: No contamination

Method	Bias		MSE	
	σ_u^2	σ_e^2	σ_u^2	σ_e^2
ML	2,346	-0,022	6,248	0,119
RML	0,838	0,335	1,430	0,362
RH3	0,437	-0,167	0,586	0,227

Table 9: Scenario Type 1: One outlying domain.

Parameter	ML		REML		RH3	
	Bias	MSE	Bias	MSE	Bias	MSE
β_0	0,250	0,319	0,092	0,308	0,087	0,306
β_1	0,318	0,015	0,324	0,016	0,324	0,016
β_2	-0,013	0,012	-0,005	0,013	-0,006	0,013
β_3	-0,003	0,004	-0,007	0,005	-0,008	0,005

Table 10: Scenario Type 1: One outlying domain.

Method	\overline{ARB}	\overline{RRMSE}
EBLUP	0,4161	0,5301
REBLUP	0,4192	0,5251
RH3-EBLUP	0,4193	0,5248

Table 11: Scenario Type 1: One outlying domain.

Method	Bias		MSE	
	σ_u^2	σ_e^2	σ_u^2	σ_e^2
ML	5,027	-0,267	26,706	0,186
RML	3,205	0,478	15,848	0,541
RH3	2,386	-0,319	6,076	0,277

Table 12: Scenario Type 1: Two outlying domains.

Parameter	ML		REML		RH3	
	Bias	MSE	Bias	MSE	Bias	MSE
β_0	0,637	0,688	0,336	0,453	0,307	0,459
β_1	0,304	0,015	0,317	0,018	0,318	0,018
β_2	-0,016	0,013	-0,009	0,014	-0,008	0,015
β_3	-0,009	0,004	-0,012	0,005	-0,010	0,005

Table 13: Scenario Type 1: Two outlying domains.

Method	\overline{ARB}	\overline{RRMSE}
EBLUP	0,4162	0,6296
REBLUP	0,4316	0,5652
RH3-EBLUP	0,4338	0,5502

Table 14: Scenario Type 1: Two outlying domains.

Method	Bias		MSE	
	σ_u^2	σ_e^2	σ_u^2	σ_e^2
ML	-0,095	0,959	0,173	1,046
RML	-0,159	0,363	0,296	0,349
RH3	-0,216	0,364	0,293	0,322

Table 15: Scenario Type 2: 10% outlying observations within groups.

Parameter	ML		REML		RH3	
	Bias	MSE	Bias	MSE	Bias	MSE
β_0	-0,014	0,342	-0,028	0,337	-0,018	0,323
β_1	0,316	0,016	0,315	0,016	0,314	0,015
β_2	0,001	0,009	0,002	0,014	0,004	0,009
β_3	-0,006	0,005	-0,006	0,005	-0,008	0,005

Table 16: Scenario Type 2: 10% outlying observations within groups.

Method	\overline{ARB}	\overline{RRMSE}
EBLUP	0,4417	0,5286
REBLUP	0,3963	0,5002
RH3-EBLUP	0,3849	0,4881

Table 17: Scenario Type 2: 10% outlying observations within groups.

Method	Bias		MSE	
	σ_u^2	σ_e^2	σ_u^2	σ_e^2
ML	-0,180	1,912	0,184	3,783
RML	-0,214	0,604	0,293	0,567
RH3	-0,232	0,575	0,286	0,554

Table 18: Scenario Type 2: 20% outlying observations within groups.

Parameter	ML		REML		RH3	
	Bias	MSE	Bias	MSE	Bias	MSE
β_0	0,005	0,367	-0,028	0,352	-0,018	0,353
β_1	0,306	0,020	0,316	0,018	0,314	0,017
β_2	-0,006	0,015	-0,002	0,015	-0,001	0,015
β_3	-0,007	0,006	-0,008	0,005	-0,009	0,005

Table 19: Scenario Type 2: 20% outlying observations within groups.

Method	\overline{ARB}	\overline{RRMSE}
EBLUP	0,4265	0,5440
REBLUP	0,3895	0,4920
RH3-EBLUP	0,3825	0,4845

Table 20: Scenario Type 2: 20% outlying observations within groups.

References

- [1] Banerjee M, Frees EW (1997) Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association* **92**:999–1005.
- [2] Christensen R, Pearson LM, Johnson W (1992) Case-deletion diagnostics for mixed models. *Technometrics* **34**:38–45.
- [3] Demidenko E (2004) Mixed models. Theory and applications. Wiley.
- [4] Fellner WH (1986) Robust estimation of variance components. *Technometrics* **28**:51–60.
- [5] Galpin JS, Zewotir T (2005) Influence diagnostics for linear mixed models. *Journal of Data Science* **3**:153–177.
- [6] Galpin JS, Zewotir T (2007) A unified approach on residuals, leverages and outliers in the linear mixed models. *Test* **16**:58–75.
- [7] Henderson CR (1953) Estimation of Variance and Covariance Components. *Biometrics* **9**:226–252.
- [8] Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**:423–447.
- [9] Huggins, RM (1993). A robust approach to the analysis of repeated measures. *Biometrics* **49**:715–720.
- [10] Jiang J (2007) Linear and generalized linear models and their applications. Springer Series in Statistics.
- [11] Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* **38**:963–974.
- [12] Maronna R, Martin D, Yohai V (2006) Robust statistics. Theory and methods. Wiley.
- [13] Maronna, RA and Yohai, VJ (2000) Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference* **89**:197–214.
- [14] McCulloch Ch, Searle S (2001) Generalized, linear and mixed models. Wiley.

- [15] Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**:545–554.
- [16] Peña, D and Yohai, VJ (1999) A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association* **94**:434–445.
- [17] Richardson, AM (1995). Some problems in estimation in mixed linear models. PhD thesis. Australian National University, Canberra.
- [18] Richardson AM (1997) Bounded influence estimation in the mixed linear model. *Journal of the American Statistical Association* **92**:154–161.
- [19] Richardson AM, Welsh AH (1995) Robust restricted maximum likelihood in mixed linear models. *Biometrics* **51**:1429–1439.
- [20] Royall RM (1976) The linear least squared prediction approach to two-stage sampling. *Journal of the American Statistical Association* **71**:657–664.
- [21] Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley series in probability and mathematical statistics.
- [22] Vangeneugden T, Laenen A, Geys H, Renard D, Molenberghs G (2004) Applying linear mixed models to estimate reliability in clinical trial data with repeated measurements. *Controlled Clinical Trials* **25**:13–30.
- [23] Verbeke G, Molenberghs G (2009) Linear mixed models for longitudinal data. Springer.
- [24] Wellenius GA, Yeh GY, Coull BA, Suh HH, Phillips RS, Mittlemann MA (2007) Effects of ambient air pollution on functional status in patients with chronic congestive heart failure: a repeated-measures study. *Environmental Health* 6–26.
- [25] Jennrich, RI, Schluchter, MD (1986) Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 80-5-820
- [26] Lindstrom MJ, Bates DM (1988) Newton-Rhapson and EM Algorithm for linear mixed-effects model for repeated-measures. *Journal of the American Statistical Association* 83-1014-1022
- [27] Chambers RL (1986) Outlier robust finite population estimation. *Journal of the American Statistical Association* 81-1063-1069

- [28] Sinha SK, Rao JNK (2009) Robust small area estimation. *The Canadian Journal of Statistics* 37-381-399
- [29] Maronna R, Martin D, Yohai V (2006) Robust statistics. Theory and methods. *Wiley*
- [30] Peña D, Yohai VJ (1999) A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association. Theory and Methods* 94-434-445
- [31] Maronna, RA and Yohai, VJ (2000) Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference* 89-197-214
- [32] Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31-423-447