

## GENERALIZED S-ESTIMATORS FOR LINEAR MIXED EFFECTS MODELS

Inna Chervoneva and Mark Vishnyakov

*Thomas Jefferson University*

*Abstract:* Linear mixed effects (LME) models are important statistical tools for analysis of clustered and correlated data. High breakdown estimators are currently the robust methods of choice for multivariate linear regression, but extensions of such estimators have been developed only for completely balanced LME models. In this work, we propose a generalized S-estimator for a general unbalanced LME model. Our GS-estimator reduces to the classic high breakdown S-estimator when the LME model reduces to a multivariate normal location and scale model or multivariate regression model. The asymptotic properties are established, and we show that the estimator may be viewed as a redescending M-estimator. A small simulation study is conducted to compare performance of the GS-estimates, monotone M-estimates, and restricted maximum likelihood (REML) estimates under various contamination patterns. The proposed estimator is used for analysis of age-related changes in hemoglobin levels of sickle cell disease patients.

*Key words and phrases:* Breakdown, clustered data, longitudinal data, redescending M-estimators, robust estimation.

### 1. Introduction

Linear mixed effects (LME) models are widely used for analysis of clustered and correlated data from the biomedical and agricultural sciences, psychology, sociology, economics, and more. Robust estimation of the LME models was originally proposed through robustified weighted log-likelihood (Huggins (1993); Huggins and Staudte (1994); Richardson and Welsh (1995)) or through the log-likelihood based on t-distribution (Stahel and Welsh (1997); Pinheiro, Liu, and Wu (2001)). Richardson (1997) and Welsh and Richardson (1997) proposed a general class of bounded influence estimating equations (BIEE) estimators (generalized M-estimators with monotone increasing  $\rho$ -function) for the LME models. Field, Pang, and Welsh (2010) developed robust bootstrap procedures for monotone M-estimators of covariance parameters in a general LME model. Samanta and Welsh (2013) developed fast and robust bootstrap for monotone M-estimators in LME models and investigated performance of various bootstrap estimates for highly unbalanced clustered data. Monotone M-estimators have good local robustness (bounded influence function), but do not necessary

have good breakdown properties (Maronna (1976); Maronna, Martin, and Yohai (2006)). Recent work in robust estimation of linear models has shifted toward high breakdown S-estimators (Bilodeau and Duchesne (2000)); Van Aelst and Willems (2005)) and redescending M-estimators (Yohai (1987); Tatsuoka and Tyler (2000)). For balanced LME models, Copt and Victoria-Feser (2006) proposed a constrained S-estimator defined similarly to the S-estimators of multivariate location and scale (Davies (1987); Lopuhaä (1989)). This estimator is limited to completely balanced data with the same number of observations and the same covariance matrix for each independent unit, and there is a need for extension to unbalanced LME models (Heritier et al. (2009)).

We consider unbalanced LME models for independent vector observations with variable dimensions and not necessarily identical covariance structure. The proposed generalized S-estimator (GS-estimator) minimizes the weighted sum of the log-transformed determinants of the covariance matrices of independent vector observations, subject to constraint insuring unbiasedness of the fixed effects estimates and a desired *breakdown parameter*. If an LME model reduces to a multivariate regression or location and scale model, then the GS-estimator reduces to the S-estimator and the *breakdown parameter* becomes a breakdown point. For balanced LME models, the GS-estimator is the constrained S-estimator of Copt and Victoria-Feser (2006).

In Section 2, we formally define the GS-estimators and derive the corresponding system of unbiased estimating equations. Asymptotic properties of the GS-estimators are presented in Section 3. Section 4 describes the computing algorithm for GS-estimators. The fixed effects inference based on GS-estimators is addressed in Section 5. We report on the results of the simulation studies in Section 6, and present a data example in Section 7. Section 8 concludes with a discussion. All proofs are given in the online supplement available at <http://www.stat.sinica.edu.tw/statistica>.

## 2. Generalized S-estimator

Consider the standard hierarchical LME model

$$\mathbf{y}_i = \mathbf{A}_i\beta + \mathbf{Z}_i\gamma_i + \varepsilon_i, \quad (2.1)$$

where  $m_i$ -dimensional vectors  $\mathbf{y}_i$ ,  $i = 1, \dots, M$ , are assumed to be independent,  $\beta$  is a  $q \times 1$  vector of fixed population parameters,  $\gamma_i$  is a  $g \times 1$  vector of random effects associated with  $\mathbf{y}_i$ ,  $\mathbf{A}_i$  and  $\mathbf{Z}_i$  are design matrices for the fixed and random effects, respectively, and  $\varepsilon_i$ ,  $i = 1, \dots, M$ , are independent vectors of error terms. It is assumed that  $\gamma_i \stackrel{i.i.d.}{\sim} \mathbf{N}_g(\mathbf{0}, \mathbf{G})$ ,  $\varepsilon_i \sim \mathbf{N}_{m_i}(\mathbf{0}, \mathbf{R}_i)$ , and  $\gamma_l$  is independent of any  $\varepsilon_i$  for any  $l$  and  $i$ . We denote by  $\theta = [\theta_1, \dots, \theta_l]'$  the vector of all unknown

covariance parameters in  $\mathbf{G}$  and  $\mathbf{R}_i$ ,  $i = 1, \dots, M$ , and by  $\xi$  the vector of all population parameters,  $\xi = [\beta', \theta']' \in \Theta \subset \mathbf{R}^{q+l}$ ,  $q = \dim(\beta)$ , and  $l = \dim(\theta)$ . We write  $\mathbf{y} = \mathbf{A}\beta + \mathbf{Z}\gamma + \varepsilon$ , where  $\mathbf{y} = [\mathbf{y}'_1, \dots, \mathbf{y}'_M]'$ ,  $\mathbf{A} = [\mathbf{A}'_1, \dots, \mathbf{A}'_M]'$ ,  $\mathbf{Z} = \text{Diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_M)$ ,  $\gamma = [\gamma'_1, \dots, \gamma'_M]'$  and  $\varepsilon = [\varepsilon'_1, \dots, \varepsilon'_M]'$ . Model (2.1) also has a multivariate formulation as

$$\mathbf{y}_i \sim \mathbf{N}_{m_i}(\mathbf{A}_i\beta, \mathbf{V}_i), \quad i = 1, \dots, M, \quad \text{where } \mathbf{V}_i = \mathbf{Z}_i\mathbf{G}\mathbf{Z}'_i + \mathbf{R}_i. \quad (2.2)$$

Let  $0 < \epsilon \leq 0.5$  be the breakdown parameter,  $m_{\max} = \max_i m_i$ , and  $m_{\min} = \min_i m_i$ . Consider a family of real-valued  $\rho$ -functions  $\{\rho_{m_i}^\epsilon(d)\}$ ,  $m_{\min} \leq m_i \leq m_{\max}$ ,  $d \in [0, \infty)$ , such that parameters of  $\rho_{m_i}^\epsilon$  are selected as solutions of the equation

$$\epsilon = \frac{E_{d^2 \sim \chi_{m_i}^2} [\rho_{m_i}^\epsilon(d)]}{\max_d \rho_{m_i}^\epsilon(d)}, \quad (2.3)$$

with  $\rho$ -functions that satisfy the conditions (Lopuhaä (1989))

**R1**  $\rho_{m_i}^\epsilon(d)$  is symmetric with continuous derivative and  $\rho_{m_i}^\epsilon(0) = 0$ ,

**R2** there exists  $c_0(m_i) > 0$  such that  $\rho_{m_i}^\epsilon(d)$  is strictly increasing on  $[0, c_0(m_i)]$  and constant on  $[c_0(m_i), \infty)$ ,

**R3**  $u_{m_i}^\epsilon(d) = (1/d) \frac{\partial}{\partial d} \rho_{m_i}^\epsilon(d)$  and  $\frac{\partial^2}{\partial d^2} \rho_{m_i}^\epsilon(d)$  are bounded and continuous.

Conditions **R1-R3** imply that  $\rho_{m_i}^\epsilon(d)$  are redescending functions. To obtain a generalized S (GS)-estimator of  $\xi$ , it is proposed to minimize

$$D = \sum_{i=1}^M v_i \ln [\det(\mathbf{V}_i)], \quad (2.4)$$

subject to

$$\sum_{i=1}^M \rho_{m_i}^\epsilon(d_i) - b_M = 0 \quad (2.5)$$

where  $d_i$  is the Mahalanobis distance for vector  $\mathbf{y}_i$ ,

$$d_i = [(\mathbf{y}_i - \mathbf{A}_i\beta)' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{A}_i\beta)]^{1/2}, \quad (2.6)$$

$$v_i = m_i^{-1} E_{d^2 \sim \chi_{m_i}^2} [d^2 u_{m_i}^\epsilon(d)], \quad (2.7)$$

$$b_M = \sum_{i=1}^M b_{m_i}, \quad b_{m_i} = E_{d^2 \sim \chi_{m_i}^2} [\rho_{m_i}^\epsilon(d)]. \quad (2.8)$$

The  $\rho$ -function applied to each vector  $\mathbf{y}_i$  is selected to correspond to dimension  $m_i$  of  $\mathbf{y}_i$  in the sense that  $E [\rho_{m_i}^\epsilon(d)]$  in (2.3) is computed under  $d^2 \sim \chi_{m_i}^2$ . Meanwhile, parameter(s) of  $\rho_{m_i}^\epsilon$  are solutions of (2.3) with a priori chosen breakdown

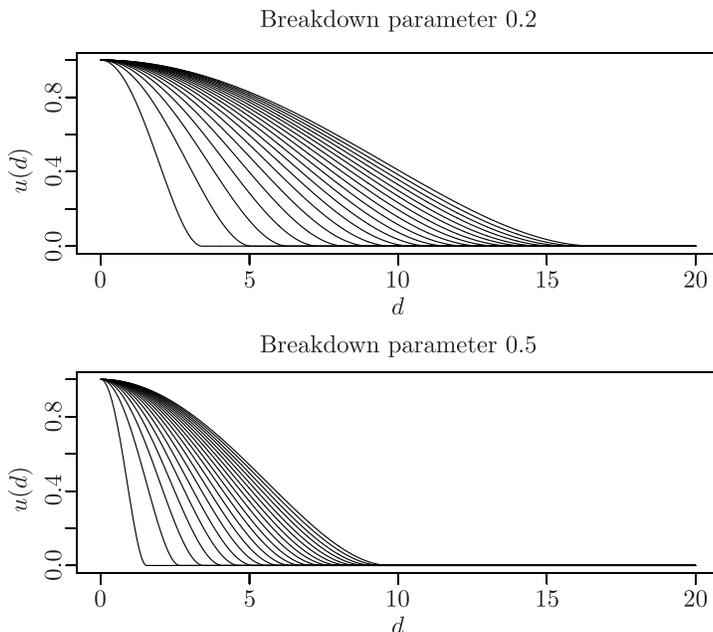


Figure 1. Tukey biweight weighting functions  $u(d)$  for dimensions  $m = 1$  to 20 and breakdown parameters 0.2 and 0.5. The curves shift right as dimension increases.

parameter  $\epsilon \leq 0.5$ , the same for all  $m_i$ . In numerical studies, we use the Tukey biweight  $\rho$ -functions (Beaton and Tukey (1974)).

$$\rho_B(d; c_{m_i}) = \begin{cases} \frac{1}{6}c_{m_i}^2 \left[ 1 - \left( 1 - \left( \frac{d}{c_{m_i}} \right)^2 \right)^3 \right] & \text{if } d \leq c_{m_i}, \\ \frac{1}{6}c_{m_i}^2 & \text{if } d \geq c_{m_i}. \end{cases} \quad (2.9)$$

Respectively, the weighting functions are

$$u_{m_i}(d) = \begin{cases} \left( 1 - \left( \frac{d}{c_{m_i}} \right)^2 \right)^2 & \text{if } d \leq c_{m_i}, \\ 0 & \text{if } d \geq c_{m_i}. \end{cases}$$

Constants  $c_{m_i}$ ,  $m_{\min} \leq m_i \leq m_{\max}$ , are chosen as solutions of (2.3). Figure 1 shows the weighting functions  $u_{m_i}(d)$  for  $m = 1, \dots, 20$  and  $\epsilon = 0.2$  or  $\epsilon = 0.5$ .

Constraint (2.5) is similar to the balanced case (up to variable dimension  $m_i$ ) and ensures that estimating equations for  $\beta$  are conditionally unbiased (Künsch, Stefanski, and Carroll (1989)) under the true model. Meanwhile, weights (2.7) are selected so that estimating equations for  $\theta$  are conditionally unbiased, as shown in Theorem 1.

We further suppress the subscript  $m_i$  in  $\rho_{m_i}(d_i)$  and  $u_{m_i}(d_i)$ . With the standard assumption that  $\mathbf{V}_i^{-1/2}$  exists for any  $i$ , we let  $\mathbf{r}_i = \mathbf{V}_i^{-1/2}(\mathbf{y}_i - \mathbf{A}_i\beta)$ , so that  $d_i = (\mathbf{r}_i'\mathbf{r}_i)^{1/2}$ .

**Theorem 1.** *The GS-estimator of  $\xi$  is a solution of the estimating equations*

$$\frac{1}{M} \sum_{i=1}^M u(d_i) \mathbf{A}_i' \mathbf{V}_i^{-1/2} \mathbf{r}_i = \mathbf{0}, \tag{2.10}$$

$$\frac{1}{M} \sum_{i=1}^M u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \mathbf{F}_{ji} \mathbf{V}_i^{-1/2} \mathbf{r}_i = 0, \quad j = 1, \dots, l, \tag{2.11}$$

$$\sum_{i=1}^M \rho(d_i) - b_M = 0, \tag{2.12}$$

that are unbiased under the assumed model, where

$$\mathbf{F}_{ji} = \left( \sum_{s=1}^l \frac{\partial \mathbf{V}_i}{\partial \theta_s} \right) \sum_{k=1}^M v_k \text{tr} \left( \mathbf{V}_k^{-1} \frac{\partial \mathbf{V}_k}{\partial \theta_j} \right) - \frac{\partial \mathbf{V}_i}{\partial \theta_j} \sum_{k=1}^M v_k \text{tr} \left( \mathbf{V}_k^{-1} \sum_{s=1}^l \frac{\partial \mathbf{V}_k}{\partial \theta_s} \right). \tag{2.13}$$

The proof is given at <http://www.stat.sinica.edu.tw/statistica>.

For the balanced case with  $m_i = m$  and  $v_i = v = m^{-1} E_{d^2 \sim \chi_m^2} [d^2 u(d)]$  for any  $i$ , (2.5) reduces to  $\sum_{i=1}^M \rho(d_i) = Mb_0$ . Minimizing (2.4) with equal weights  $v_i = v$  is equivalent to minimizing  $\det(\mathbf{V})$ . Hence, the proposed GS-estimator reduces to the constrained S-estimator in balanced LME models, and to the S-estimator in multivariate regression, which may be viewed as subclass of LME models. The breakdown parameter  $\epsilon$  is the same as the usual breakdown point if all fixed effects covariates are random (continuous) and all matrices  $\mathbf{V}_i$  are the same, which is not necessary in (2.1). The first problem is that a general LME model allows for any mixture of continuous and categorical fixed effects covariates, which implies that the standard definition of the breakdown point does not apply (Maronna and Yohai (2000)). The maximum possible breakdown point for multivariate regression models with only random covariates is 0.5 (Van Aelst and Willems (2005)), but for linear models with fixed designs corresponding to two or more groups, the maximum possible breakdown point of S-estimators may be lower than 0.5 (He and Fung (2000); Maronna and Yohai (2000)), and it is affected by between-subject fixed design covariates. A typical example is the dummy variables corresponding to group indicators. In particular, in case of one-way design for the fixed effects model, the results in Maronna and Yohai (2000) imply that  $\epsilon \leq (1/M) [(M_{\min} + 1)/2]$ , where  $M_{\min}$  is the number of independent

vectors in the smallest group. One can adapt the definition of the breakdown point for mixed designs proposed by Maronna and Yohai (2000) in the context of regression with continuous and categorical predictors to address the breakdown of the fixed effects parameter estimates in LME models. However, the breakdown point for covariance parameter estimates cannot be well defined in a LME model with different covariance matrices  $\mathbf{V}_i$  depending on different (but not necessarily disjoint) subsets of the covariance parameter vector  $\theta$ . In our data application, the covariance matrices for vector observations from two phenotypes depend on common error term variance and on phenotype-specific variance of the subject random effect.

The estimating equations in (2.10) look similar to the corresponding fixed effects part in bounded influence estimating equations (BIEE) of Welsh and Richardson (1997) with  $\mathbf{W}_{0i} = w_i \mathbf{I}_{m_i}$  ( $\mathbf{I}_{m_i}$  is  $m_i \times m_i$  identity matrix) and  $\mathbf{U}_{0i} = \mathbf{I}_{m_i}$ , but in (2.10) the  $\rho$ -functions and  $\psi$ -functions are applied to Mahalanobis distances corresponding to vectors  $\mathbf{r}_i$ , while in BIEE estimators, the  $\psi$ -functions are applied to each scalar component of vectors  $\mathbf{r}_i$ .

### 3. Asymptotic Properties

Let  $\Theta$  be the parameter space for  $\xi = [\beta', \theta']' \in \Theta \subset \mathbf{R}^{q+1}$ . The proof of Theorem 1 implies that the proposed GS-estimator satisfies the system of  $q + l$  unbiased estimating equations

$$\begin{aligned} \Psi_{\mathbf{M}}(\mathbf{y}, \xi) &= \left[ \Psi_M^\beta(\mathbf{y}, \xi)', \Psi_M^\theta(\mathbf{y}, \xi)' \right]' = \mathbf{0}, \\ \Psi_M^\beta(\mathbf{y}, \xi) &= \frac{1}{M} \sum_{i=1}^M u(d_i) \mathbf{A}'_i \mathbf{V}_i^{-1/2} \mathbf{r}_i, \end{aligned} \quad (3.1)$$

$$\Psi_M^\theta(\mathbf{y}, \xi) = \frac{1}{M} \sum_{j=1}^l e_j^l \otimes \sum_{i=1}^M \left[ u(d_i) \mathbf{r}'_i \mathbf{V}_i^{-1/2} \mathbf{F}_{ji} \mathbf{V}_i^{-1/2} \mathbf{r}_i + c_j (\rho(d_i) - b_{m_i}) \right]. \quad (3.2)$$

Here,  $e_j^l$  is the  $j^{\text{th}}$  canonical  $l \times 1$  vector,  $\otimes$  is the Kronecker product of two matrices,  $\mathbf{F}_{ji}$  is given (2.13), and  $b_{m_i} = E_{d^2 \sim \chi_{m_i}^2} [\rho_{m_i}^\epsilon(d)]$ . Denoting

$$g(\mathbf{y}, \xi) = \sum_{i=1}^M \rho_{m_i}^\epsilon(d_i) - b_M,$$

(3.2) can be written as

$$\Psi_M^\theta(\mathbf{y}, \xi) = \frac{1}{M} \left[ \left( \mathbf{1}'_{l \times 1} \frac{\partial g}{\partial \theta} \right) \frac{\partial D}{\partial \theta} - \left( \mathbf{1}'_{l \times 1} \frac{\partial D}{\partial \theta} \right) \frac{\partial g}{\partial \theta} + g \mathbf{c} \right],$$

where  $D$  is the objective function given in (2.4),  $\mathbf{1}_{l \times 1}$  is a  $l \times 1$  vector of ones, and  $\mathbf{c} = [c_1, \dots, c_l]$  is a non-zero  $l \times 1$  vector. The equations in (3.1) are the same  $q$  equations as in (2.10), and the  $l$  equations in (3.2) are linear combination of the  $l + 1$  equations (2.11) and (2.12). Adding  $g\mathbf{c} = \left(\sum_{i=1}^M \rho(d_i) - b_M\right) \mathbf{c}$  to (2.11) is necessary because equations in (2.11) are linearly dependent after excluding the Lagrange multiplier. Vector  $\mathbf{c}$  is any non-zero  $l \times 1$  vector such that the Jacobian of the estimating equations  $\Psi_M(\mathbf{y}, \xi_0)$  is non-singular, where  $\xi_0 = [\beta'_0, \theta'_0]'$  is the true value of  $\xi$  such that  $E[\Psi_M(\mathbf{y}, \xi_0)] = 0$ .

The following regularity conditions are used to establish consistency and asymptotic normal distribution of the GS-estimator.

**R4** There exists  $m_{\max}$  such that for any  $i$ ,  $m_i \leq m_{\max} < \infty$ .

**R5** For any  $i$ ,  $\mathbf{V}_i$  is twice continuously differentiable with respect to  $\theta$ .

**R6** For any  $\xi$  in some compact neighborhood of  $\xi_0$ , there exists

$$\Phi(\xi) = \lim_{M \rightarrow \infty} E \left[ \frac{\partial}{\partial \xi} \Psi_M(\mathbf{y}, \xi) \right] \text{ not singular at } \xi_0.$$

**R7** For any  $M$ , matrix  $\Omega_M(\xi_0) = (1/M) \sum_{i=1}^M E[\Psi_i(\mathbf{y}_i, \xi_0) \Psi_i(\mathbf{y}_i, \xi_0)']$  is positive definite, and there exists positive definite  $\Omega(\xi_0) = \lim_{M \rightarrow \infty} \{\Omega_M(\xi_0)\}$ .

**Theorem 2.** *If  $\xi_0 = [\beta'_0, \theta'_0]'$  is the true value of  $\xi$  such that  $E[\Psi_M(\mathbf{y}, \xi_0)] = 0$ , then under **R1–R5**, as  $M \rightarrow \infty$ , there exists a consistent sequence of GS-estimators such that, with probability 1,  $\Psi_M(\mathbf{y}, \hat{\xi}_M) = \mathbf{0}$  and  $\hat{\xi}_M \rightarrow \xi_0$ .*

The proof is given at <http://www.stat.sinica.edu.tw/statistica>.

**Theorem 3.** *Under the assumptions of Theorem 2 and **R6–R7**, for a consistent sequence of GS-estimators  $\hat{\xi}_M$ ,  $M^{1/2}(\hat{\xi}_M - \xi_0) \rightarrow \mathbf{N}(0, \Sigma(\xi_0))$ , where*

$$\Sigma(\xi_0) = [Diag(\Phi_{\beta_0}, \Phi_{\theta_0})]^{-1} Diag(\Omega_{\beta_0}, \Omega_{\theta_0}) [Diag(\Phi_{\beta_0}, \Phi_{\theta_0})']^{-1}$$

$$\Phi_{\beta_0} = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M E_{d_i^2 \sim \chi_{m_i}^2} [u(d_i) + m_i^{-1} u'(d_i) d_i] \mathbf{A}'_i \mathbf{V}_i^{-1} \mathbf{A}_i$$

$$\Phi_{\theta_0} = \lim_{M \rightarrow \infty} \frac{1}{M} \left\{ \left( \mathbf{1}'_{l \times 1} \frac{\partial D}{\partial \theta} \mathbf{I}_{l \times l} - \frac{\partial D}{\partial \theta} \mathbf{1}'_{l \times 1} \right) \mathbf{G} - \mathbf{c} \frac{\partial D}{\partial \theta'} \right\},$$

$$\mathbf{G} = E \left[ \frac{\partial^2 g}{\partial \theta \partial \theta'} \right] - \frac{\partial^2 D}{\partial \theta \partial \theta'}$$

$$= \sum_{j=1}^l \sum_{k=1}^l e_j^l \otimes (e_k^l)' \otimes \sum_{i=1}^M \left\{ v_i tr \left( \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_j} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \right) \right\}$$

$$\begin{aligned}
& + \frac{E[u'(d_i)d_i^3]}{m_i(m_i+2)} \operatorname{tr} \left( \left[ \left( \mathbf{V}_i^{-1/2} \frac{\partial \mathbf{V}_i}{\partial \theta_k} \mathbf{V}_i^{-1/2} \right)' \otimes \left( \mathbf{V}_i^{-1/2} \frac{\partial \mathbf{V}_i}{\partial \theta_j} \mathbf{V}_i^{-1/2} \right) \right] \mathbf{H}_{m_i} \right) \Bigg\}, \\
\frac{\partial D}{\partial \theta} &= \sum_{j=1}^l e_j^l \otimes \sum_{i=1}^M v_i \operatorname{tr} \left[ \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \theta_j} \right], \\
\Omega_{\beta_0} &= \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M m_i^{-1} E_{d_i^2 \sim \chi_{m_i}^2} [u^2(d_i)d_i^2] \mathbf{A}_i' \mathbf{V}_i^{-1} \mathbf{A}_i, \\
\Omega_{\theta_0} &= \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{j=1}^l \sum_{k=1}^l e_j^l \otimes (e_k^l)' \otimes \left\{ c_j c_k \sum_{i=1}^M E [(\rho(d_i) - b_{m_i})^2] \right. \\
& \quad + \sum_{i=1}^M \frac{E[u^2(d_i)d_i^4]}{m_i(m_i+2)} \operatorname{tr} \left( \left[ \left( \mathbf{V}_i^{-1/2} \mathbf{F}_{ki} \mathbf{V}_i^{-1/2} \right)' \otimes \left( \mathbf{V}_i^{-1/2} \mathbf{F}_{ji} \mathbf{V}_i^{-1/2} \right) \right] \mathbf{H}_{m_i} \right) \\
& \quad \left. + \sum_{i=1}^M m_i^{-1} E[u(d_i)d_i^2(\rho(d_i) - b_{m_i})] \operatorname{tr} [\mathbf{V}_i^{-1} (c_j \mathbf{F}_{ji} + c_k \mathbf{F}_{ki})] \right\},
\end{aligned}$$

$\mathbf{H}_{m_i} = [\mathbf{I} + \mathbf{K}_{m_i} + \operatorname{vec}(\mathbf{I}) \operatorname{vec}(\mathbf{I})']$ . Here  $\mathbf{K}_{m_i}$  is a  $(m_i^2 \times m_i^2)$ -block matrix with  $(i, j)$ -th block  $K(i, j)$  of size  $(m_i \times m_i)$  with 1 at  $(i, j)$  and 0 otherwise,  $\otimes$  is the Kronecker product, and  $\mathbf{F}_{ji}(\theta)$  is as in (2.13).

The proof is given at <http://www.stat.sinica.edu.tw/statistica>.

#### 4. Computing GS-estimates

To compute the GS-estimator, one has to minimize (2.4) subject to (2.5). Theorem 1 implies that the GS-estimator satisfies (2.10)–(2.12), but not all solutions of (2.10)–(2.12) are solutions of (2.4)–(2.5). Furthermore, (2.4) may have several local minima, and (2.10)–(2.12) may have multiple solutions. A common approach to computing regression S-estimators is to employ a multi-start algorithm (Maronna, Martin, and Yohai (2006)). Similarly, it is proposed to compute a large number of candidate solutions of (2.4)–(2.5) or (2.10)–(2.12) and select among them the one that yields the global minimum of (2.4).

The proposed computational algorithm is the following.

- I Select a number  $N_s$  of candidate solutions to be computed and a size  $S$  of subsamples to be used for computing starting values.
- II The following steps are repeated  $N_s$  times.
  - II.1 For each  $n \leq N_s$ , randomly select a subsample  $\{\mathbf{y}_{i_s}^{(n)}, s = 1, \dots, S\}$  of independent vectors  $\{\mathbf{y}_i, i = 1, \dots, M\}$ .

- II.2** For  $\{\mathbf{y}_{i_s}^{(n)}, s = 1, \dots, S\}$ , compute the robust REML estimates  $\widehat{\beta}_{RREML}^{(n)}$  and  $\widehat{\theta}_{RREML}^{(n)}$ .
- II.3** Using  $\widehat{\beta}_{RREML}^{(n)}$  and  $\widehat{\theta}_{RREML}^{(n)}$  as starting values, compute a candidate solution  $(\widetilde{\beta}^{(n)'}, \widetilde{\theta}^{(n)'})$  by solving (2.4)–(2.5) directly or by solving (2.10)–(2.12).
- III** Among  $N_s$  candidate solutions  $(\widetilde{\beta}^{(n)'}, \widetilde{\theta}^{(n)'})$ , select the one which yields the global minimum of (2.4) as the GS-estimator of  $[\beta', \theta']$ .

The algorithm implements "random direction" search for starting values as proposed by Ruppert (1992). The size  $S$  of the sub-samples  $\{\mathbf{y}_{i_s}, s = 1, \dots, S\}$  should be small enough to increase probability of subsamples without outliers in random effects, and large enough to allow all parameters in the model to be estimable. The number of subsamples  $N_s$  may be selected to have a desired probability of getting at least one (random effects) outlier-free subsample of size  $S$  (Section 5.7.2 in Maronna, Martin, and Yohai (2006)). The proportion of outliers  $p_O$  in error terms result in fewer uncontaminated vectors  $\mathbf{y}_i$  than  $M(1-p_O)$  under the independent contamination model (Alqallaf et al. (2009)). Therefore, for robustness to contamination in error terms, it is proposed to use robust REML for computing starting values from each subsample  $\{\mathbf{y}_{i_s}^{(n)}, s = 1, \dots, S\}$ .

In Step II.3, direct minimization of (2.4)–(2.5) subject to (2.5) may be implemented using a constrained nonlinear optimization routine such as PROC NLP in SAS (SAS Institute Inc., Cary, NC, USA). Alternatively, (2.10)–(2.12) can be solved, for example, using the R package 'nleqslv' (Hasselman (2009) or 'BB' (Varadhan and Gilbert (2009)). In either case, it is not necessary that all  $N_s$  starting values yield valid solutions of (2.4)–(2.5) or (2.10)–(2.12).

To solve (2.10)–(2.12) it is also possible to use fixed-point equations if each covariance matrix  $\mathbf{V}_i$  can be written as  $\mathbf{V}_i = \sum_{g=1}^l \theta_g \mathbf{L}_{ig}, i = 1, \dots, M$ , where  $\mathbf{L}_{ig}$  are known (mostly singular) matrices. Then, as shown in online supplementary material, the solution  $\widehat{\theta}_M$  of (2.11) is also a solution of the fixed-point equations

$$\theta = - (P_M)^{-1} \mathbf{Q}(\theta)^{-1} [\mathbf{U}_M + M^{-1}g(\mathbf{y}, \xi) \mathbf{c}], \tag{4.1}$$

where  $\mathbf{Q}(\theta)$  is a symmetric  $l \times l$  matrix with entries

$$\begin{aligned} [\mathbf{Q}(\theta)]_{jk} &= \sum_{i=1}^M v_i \text{tr} [\mathbf{V}_i^{-1} \mathbf{L}_{ij} \mathbf{V}_i^{-1} \mathbf{L}_{ik}], \\ P_M &= M^{-1} \sum_{i=1}^M \left[ u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \left( \sum_{s=1}^l \theta_s \mathbf{L}_{is} \right) \mathbf{V}_i^{-1/2} \mathbf{r}_i \right], \end{aligned}$$

and  $\mathbf{U}_M$  is an  $l \times 1$  vector with entries

$$[U_M]_j = M^{-1} S_{Ml} \sum_{i=1}^M \left[ u(d_i) \mathbf{r}_i' \mathbf{V}_i^{-1/2} \mathbf{L}_{ij} \mathbf{V}_i^{-1/2} \mathbf{r}_i \right],$$

$$S_{ML} = \sum_{k=1}^M v_k \text{tr} \left( \mathbf{V}_k^{-1} \sum_{s=1}^l \mathbf{L}_{is} \right).$$

It is straightforward that the solution  $\hat{\beta}_M$  of (2.10) is a solution of the fixed-point equations

$$\beta = \left[ \sum_{i=1}^M u(d_i) \mathbf{A}'_i \mathbf{V}_i(\theta)^{-1} \mathbf{A}_i \right]^{-1} \left\{ \sum_{i=1}^M u(d_i) \mathbf{A}'_i \mathbf{V}_i(\theta)^{-1} \mathbf{y}_i \right\}. \quad (4.2)$$

Thus, Step II.3 in the algorithm may be replaced by the following.

**II.3\*** For a pre-determined number  $N_I$ , perform up to  $N_I$  iterations using (4.2)–(4.1) with the starting values computed in Step II.2 while evaluating the absolute value of the relative change in  $D$ ,  $|\Delta D|/D$ , between consecutive iterations. Record estimates for  $\beta$  and  $\theta$  only if  $|\Delta D|/D < \delta_D$  and  $|g(\mathbf{y}, \beta, \theta)| < \delta_g$  for some sufficiently small  $\delta_D$  and  $\delta_g$ .

## 5. Inference for Fixed Effects

For the fixed effects GS-estimates, the approximate covariance matrix is

$$\hat{\Sigma}_M(\hat{\beta}_M) = \frac{1}{M} \left[ \hat{\Phi}_{\beta_0 M}(\hat{\xi}_M) \right]^{-1} \hat{\Omega}_{\beta_0 M}(\hat{\xi}_M) \left[ \hat{\Phi}_{\beta_0 M}(\hat{\xi}_M)' \right]^{-1}, \quad (5.1)$$

where

$$\begin{aligned} \hat{\Omega}_{\beta_0 M}(\hat{\xi}_M) &= \frac{1}{M} \sum_{i=1}^M e_{1m_i} \mathbf{A}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{A}_i, \quad e_{1m} = m^{-1} E_{d^2 \sim \chi_m^2} [u^2(d) d^2], \\ \hat{\Phi}_{\beta_0 M}(\hat{\xi}_M) &= \frac{1}{M} \sum_{i=1}^M e_{2m_i} \mathbf{A}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{A}_i, \quad e_{2m} = E_{d^2 \sim \chi_m^2} [u(d) + m^{-1} u'(d) d]. \end{aligned}$$

With  $u(d) \equiv 1$ , which corresponds to the maximum likelihood estimation, (5.1) reduces to  $(\sum_{i=1}^M \mathbf{A}'_i \mathbf{V}_i^{-1} \mathbf{A}_i)^{-1}$ , the asymptotic covariance matrix of the maximum likelihood estimator. The approximate covariance matrix of  $\hat{\theta}_M$  can be computed by substituting  $\hat{\beta}_M$  and  $\hat{\theta}_M$  into expressions for  $\Phi_{\theta_0}$  and  $\Omega_{\theta_0}$  in Theorem 1. Alternatively, one can use an empirical sandwich estimator replacing the expectations by their empirical analogues. Such an approach simplifies computations of  $\hat{\Omega}_M(\hat{\xi}_M)$ , but the empirical analogue of  $\hat{\Phi}_M(\hat{\xi}_M)$ , which involves the derivatives of the estimating equations, is more complex than  $\Phi_{\theta_0}$ . For the balanced case with  $m_i = m$  and  $\mathbf{V}_i = \mathbf{V}$  for any  $i$ , (5.1) reduces to  $\hat{\Sigma}_M(\hat{\beta}_M) = (e_{1m}/e_{2m}^2) (\sum_{i=1}^M \mathbf{A}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{A}_i)^{-1}$ . This is the same result as in Heritier et al. (2009) for MM-estimators, which have the same estimating equations for the fixed effects as (3.1). Ratio  $e_2^2/e_1$  is the asymptotic relative efficiency of the location S-estimator.

Matrices  $\widehat{\Sigma}_M(\widehat{\beta}_M)$  and  $\widehat{\Sigma}_M(\widehat{\theta}_M)$  can be used to construct approximate normal confidence intervals for  $\widehat{\beta}_M$  and  $\widehat{\theta}_M$ . For fixed effects inference, one can use the Wald chi-square test of a general linear hypothesis  $\mathbf{H}_0 : \mathbf{G}\beta = \mathbf{0}$  vs.  $\mathbf{G}\beta \neq \mathbf{0}$ , where  $\mathbf{G}$  is a matrix of full column rank. Test statistic

$$\mathbf{T}^2 = \widehat{\beta}'_M \mathbf{G}' \left[ \mathbf{G} \left\{ \widehat{\Sigma}_M(\beta_0) \right\} \mathbf{G}' \right]^{-1} \mathbf{G} \widehat{\beta}_M$$

has an asymptotic  $\chi^2$  distribution with degrees of freedom equal to  $rank(\mathbf{G})$ .

## 6. Simulation Study

A simulation study was conducted to evaluate finite sample performance of the proposed GS-estimator, and to compare it to the standard REML and the robust REML II estimator of Richardson and Welsh (1995). In Richardson and Welsh (1995), the robust REML II estimator performed the best among other considered BIEE estimators. The design of our simulation study is similar to theirs, except that we generated unbalanced data and considered additional contaminations with “good” and “bad” leverage points.

Uncontaminated data were simulated from the one-way random effects model

$$y_{ij} = \alpha_1 + \alpha_2 x_{ij} + \gamma_i + \varepsilon_{ij}, \quad (6.1)$$

where  $i = 1, \dots, M$ ,  $j = 1, \dots, m_i$ ,  $m_i \in \{2, 3, 4, 5, 6\}$ ,  $x_{ij} \sim N(0, 1)$ ,  $\gamma_j \sim N(0, \sigma_1^2)$ , and  $\varepsilon_{ij} \sim N(0, \sigma_2^2)$ . Two, four, or six replicates of 5 groups ( $M = 10, 20$ , or  $30$ ) of sizes  $m_i = 2, 3, 4, 5, 6$  were generated, so that the total sample size (40, 80, and 120) was always a multiple of  $\sum_{j=2}^6 j = 20$ . The true parameter values were  $\alpha_1 = \alpha_2 = \sigma_1^2 = \sigma_2^2 = 1$ .

Following Richardson and Welsh (1995), contaminations were introduced by replacing one or both normal distributions  $\gamma_j \sim N(0, 1)$  and  $\varepsilon_{ij} \sim N(0, 1)$  with  $0.9N(0, 1) + 0.1N(0, 11)$ . Contamination scenarios are denoted by (0,0), (0.1,0), (0,0.1), (0.1,0.1), where the first number in the pair stands for contamination proportion in the distribution of random effects  $\gamma_j$  and the second number stands for contamination proportion in the distribution of error terms  $\varepsilon_{ij}$ . Scenario (0.1,0) is the classic Tukey-Huber contamination model, while (0,0.1) is the independent contamination model (Alqallaf et al. (2009)). Additionally, leverage points were introduced by replacing  $x_{ij} \sim N(0, 1)$  with  $x_{ij} \sim 0.9N(0, 1) + 0.1N(0, 11)$ . For generating “good” leverage points (scenario (0.1 glp X)), the  $x_{ij}$  were contaminated *before*  $y_{ij}$  were generated according to (6.1). For generating “bad” leverage points (scenario (0.1 blp X)), the  $x_{ij}$  were contaminated *after* generating  $y_{ij}$  according to (6.1).

Two hundred data sets were simulated for each scenario and analyzed using the standard REML, Robust REML II, and GS-estimators with  $\epsilon = 0.2$  and  $\epsilon =$

0.5. For Robust REML II, the Huber  $\psi$ -function,  $\psi_{Hc}(x; c) = \min\{c, \max\{x, -c\}\}$  was used with  $c = 1.34$  for location and  $c = 2.0$  for scale estimation as in Richardson and Welsh (1995). The GS-estimators were computed using the algorithm of Section 4 with  $S = 4$  for  $M = 10$ ,  $S = 6$  for  $M = 20$ ,  $S = 8$  for  $M = 30$ ,  $N_s = 50$ , and utilizing PROC NLP in SAS (SAS Institute Inc., Cary, NC, USA) for computing candidate solutions in Step II.3. For  $M = 10$  and 0.1 proportion of outliers in random effects, using  $S = 4$  implies probability  $\binom{9}{4}/\binom{10}{4} = 0.60$  of a random effects outlier-free subsample of size  $S$  (sampling without replacement). Then, the probability of at least one random effects outlier-free subsample among  $N_s = 50$  subsamples is  $\delta = 1 - (1 - 0.60)^{50} = 1 - 1.3 \times 10^{-20}$ . The same probability of at least one random effects outlier-free subsample among  $N_s = 50$  of sizes  $S = 6$  from  $M = 20$  and  $S = 8$  from  $M = 30$  are  $1 - 7.0 \times 10^{-15}$  and  $1 - 4.4 \times 10^{-11}$ , respectively.

Figure 2 shows the boxplots of error distributions of parameter estimates, for which differences in performance of considered estimators were observed. For scenario (0.1 glp X), we observe substantial estimation bias ( $\sim 30\%$  on average) for the slope  $\alpha_2$  estimates using REML and Robust REML II, but generally unbiased GS-estimates. For scenario (0.1 blp X), there is a downward bias in estimating  $\alpha_2$  using all estimators, but for GS-estimates with  $\epsilon = 0.5$ , the empirical bias is  $\sim 40\%$  lower ( $\sim 0.3$  vs  $\sim 0.5$ ). For scenarios with contaminations in error terms ((0,0.1) and (0.1,0.1)), the GS-estimators and Robust REML II yield similar errors for estimating the error terms variance component  $\sigma_2^2$ , while these errors are substantially higher for the corresponding REML estimates. For scenarios (0.1,0) and (0.1,0.1), GS-estimators, especially with  $\epsilon = 0.5$ , yield smaller errors in estimates of random effects variance  $\sigma_1^2$  as compared to both REML and Robust REML II. The boxplots of error distributions for all parameter estimates when  $N = 40$  and  $N = 80$  are shown in Supplementary Figures 1 and 2, respectively. Supplementary Figures 3-4 show the QQ-plots of the GS-estimates with  $\epsilon = 0.2$  and  $\epsilon = 0.5$  for  $N = 80$ . They suggest that for both fixed effects and covariance parameter estimates, the normal distribution approximation is generally appropriate for  $N \geq 80$ . The results for  $N = 120$  are similar to results for  $N = 80$  and not shown.

The root mean squared errors (RMSE) are displayed in Figure 3 for parameters and scenarios with differences in performance of considered estimators and in Supplementary Figure 5 for all parameters and scenarios. For fixed effects parameter estimates, RMSE is similar for all estimation methods and all contamination patterns, except for the ones with leverage points. For slope parameter  $\alpha_2$ , in (0.1 glp X) with  $N = 40$ , the RMSE of the GS-estimators is about half as large as compared to the RMSE of REML and Robust REML II. For (0.1 blp X), the GS-estimator with  $\epsilon = 0.5$  yields the smallest RMSE for all considered sample sizes (Figure 3). For covariance parameter estimates, the RMSE is

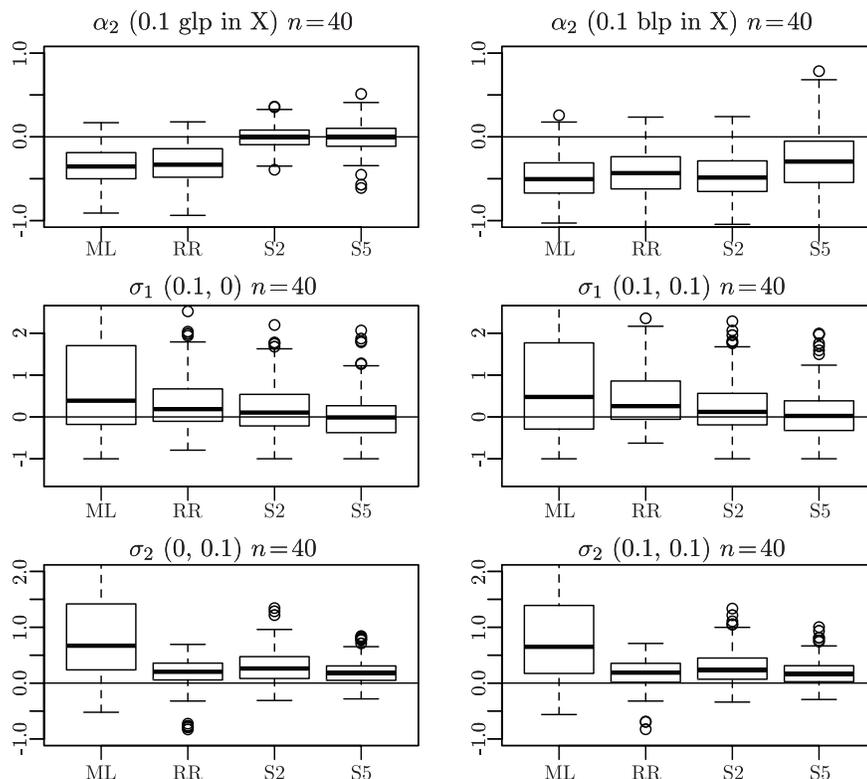


Figure 2. Error distribution of parameter estimates for simulated data sets with  $n=40$  observations from ten groups with 10% contamination in error terms  $(0,0.1)$ , with 10% contamination in random effect  $(0.1, 0)$ , with 10% contamination in both error terms and random effect  $(0.1,0.1)$ , and with 10% of “good”  $(0.1 \text{ glp in } X)$  and 10% of “bad”  $(0.1 \text{ blp in } X)$  leverage points. Estimation methods used are REML (ML), Robust REML II (RR), GS-estimator with breakdown parameter 0.2 (S2), and GS-estimator with breakdown parameter 0.5 (S5).

substantially smaller for GS as compared to REML and Robust REML II for variance components corresponding to contaminated sources of variability. For scenarios shown in Figure 3, the Robust REML II estimates of  $\sigma_1^2$  and  $\sigma_2^2$  resulted in a small proportion of very large errors (4-6% of data sets for  $N = 40$ , 2% for  $N = 80$ , and 1% for  $N = 120$ ). Because of these large errors, the corresponding RMSE of Robust REML II are higher than 300% and outside of the plotting areas in subplots  $\sigma_1^2(0.1,0)$ ,  $\sigma_2^2(0,0.1)$ ,  $\sigma_1^2(0.1,0.1)$ , and  $\sigma_2^2(0.1,0.1)$ .

The empirical coverage of asymptotic 95% confidence intervals is shown in Supplementary Figure 6 for all scenarios, and in Figure 4 for selected scenarios where some differences between  $\epsilon = 0.2$  and  $\epsilon = 0.5$  were observed. For

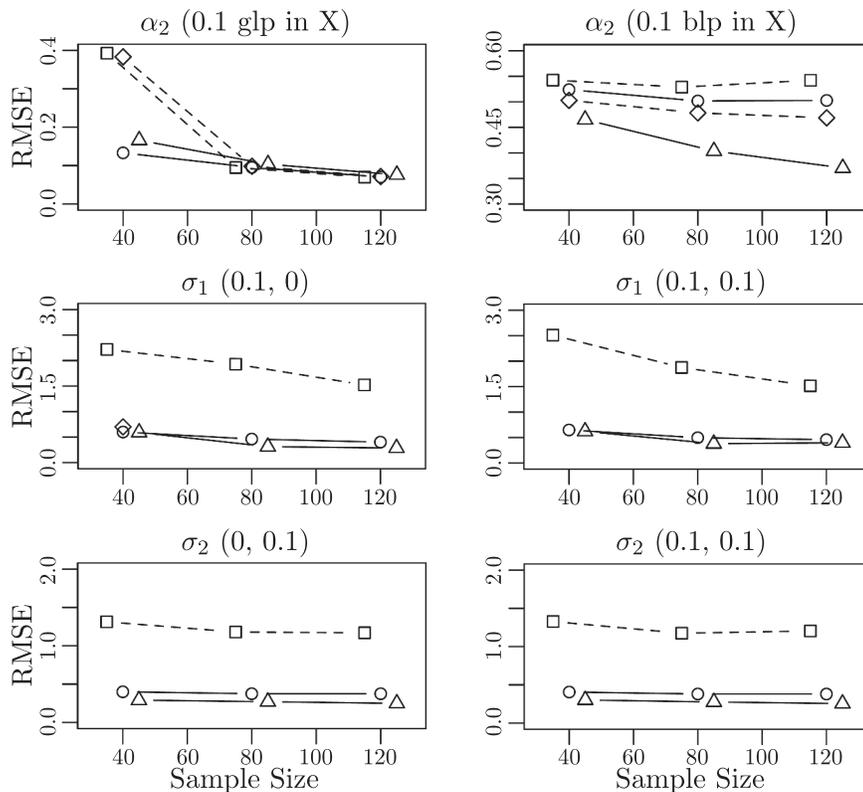


Figure 3. Root mean squared errors (RMSE) as a function of sample size (40, 80 or 120) for parameter estimates in simulated data sets with 10% contamination in error terms (0,0.1), with 10% contamination in random effect (0.1, 0), with 10% contamination in both error terms and random effect (0.1,0.1), and with 10% of “good” (0.1 glp in X) and 10% of “bad” (0.1 blp in X) leverage points. Estimation methods used are REML (squares, dashed lines), Robust REML II (diamonds, dashed lines), GS-estimator with breakdown parameter 0.2 (circles, solid lines), and GS-estimator with breakdown parameter 0.5 (triangles, solid lines). Missing points for Robust REML II correspond to the mean squared errors far outside the plotting range.

$N \geq 80$ , the nominal 95% coverage was generally maintained (empirical coverage within the approximate acceptance region  $[0.92, 0.98]$  for the null hypothesis  $H_0 : p = 0.95$ ) for fixed effects parameter estimates in all scenarios, except for estimates of  $\alpha_2$  in scenario (0.1 blp X). For covariance parameter estimates, the empirical coverage is either close to nominal or conservative for GS-estimates with  $\epsilon = 0.5$  and  $N \geq 80$  in all contaminated scenarios. Meanwhile, the coverage is significantly lower than nominal for  $\sigma_2^2$  with  $\epsilon = 0.2$  in scenarios with contamination in error terms ((0:0.1) and (0.1:0.1)).

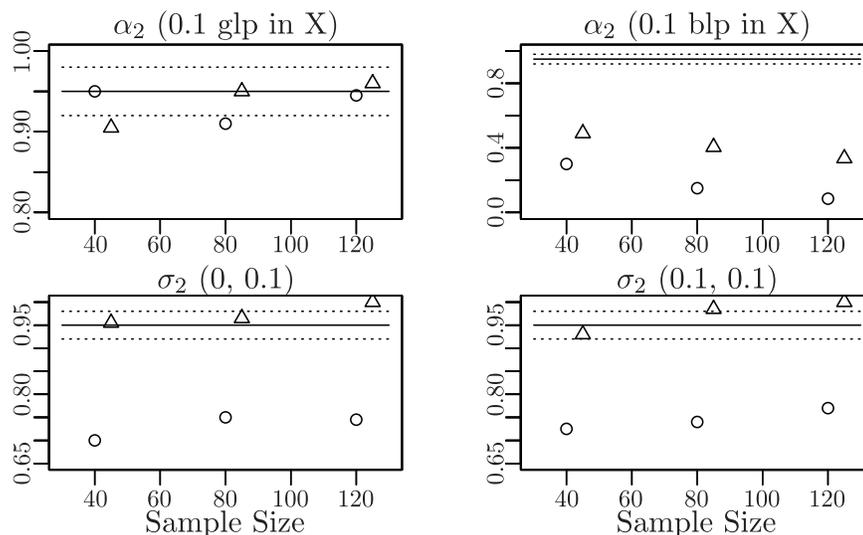


Figure 4. Coverage of asymptotic 95% confidence intervals for sample sizes 40, 80 and 120 in simulated data sets with 10% contamination in error terms (0,0.1), with 10% contamination in random effects (0.1,0), with 10% contamination in both error terms and random effects (0.1,0.1), and with 10% of “good” (0.1 glp in X) and 10% of “bad” (0.1 blp in X) leverage points. Results for GS-estimator with breakdown parameter 0.2 plotted with circles and for GS-estimator with breakdown parameter 0.5 plotted with triangles. Solid lines indicates the nominal 95% coverage, and the dotted lines limit the approximate acceptance region  $[0.92, 0.98]$  for the null hypothesis  $H_0: p = 0.95$ .

In conclusion, the GS-estimators provided less biased results with smaller RMSE as compared to REML and Robust REML when contaminations were present in random effects or predictors. For the contaminations considered, the coverage of asymptotic confidence intervals based on GS-estimators with  $\epsilon = 0.5$  was consistent with the nominal level, unless the contamination represented “bad” leverage points.

## 7. Data Example

Hemoglobin (Hb) is the standard measure to diagnose sickle cell anemia and differentiate between the phenotypes HbSS and HbSC. Our data include 227 Hb measures from 134 patients with sickle cell disease (88 with HbSS and 46 with HbSC phenotype) of ages from 22 month to 23 years old. The number of repeated measures per patient varies from 1 to 4 (75 patients (56%) with 1 observation, 34 patients (25%) with 2 observations from , 17 patients (13%) with 3 observations, and 17 patients (6%) with 4 observations). We are interested in evaluating age

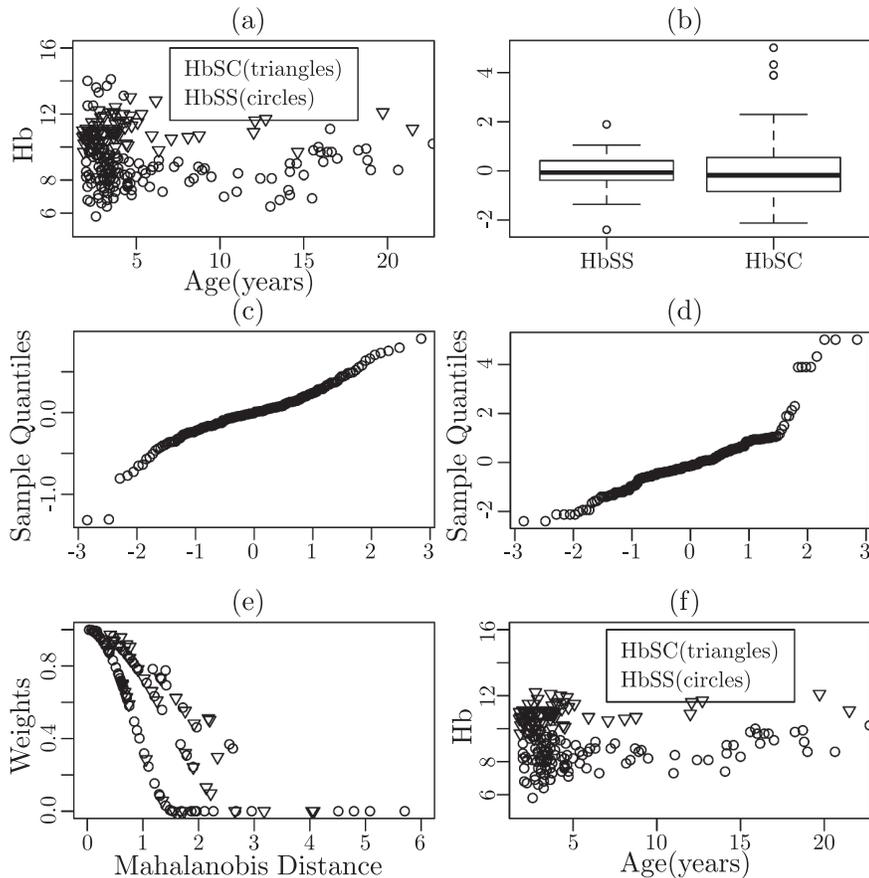


Figure 5. Scatterplot of the sickle cell patients Hemoglobin (Hb) vs. Age (a); boxplots of the best linear predictors of random effects by genotype (b); diagnostics for the standard LME model: (c) residual quantiles, (d) quantiles of the best linear predictors of random effects; (e) weights for GS-estimator with breakdown parameter 0.5 vs. Mahalanobis distances at the final GS-estimated model; and (f) Hemoglobin (Hb) vs. Age only in sickle cell patients with GS-estimator weights greater than 0.01.

related Hb trends in each phenotype since it is still not well established whether Hb tends to increase or decrease with age for more severe HbSS phenotype. Figure 5(a) shows the Hb measurements by phenotype as a function of age. It indicates a small group of potential high end outliers among HbSS patients and leverage points that correspond to sparse data for older ages. Log transformation of age was considered, but it did not alleviate potential problems with outliers and leverage points.

Initially, the data were analyzed using the standard LME model in SAS

Table 1. Parameter estimates (Estimate) and standard errors (SE) for the linear mixed effects model fitted to the sickle cell data.

Group	Parameter	REML			GS		
		Estimate	SE	p-value	Estimate	SE	p-value
SC	$\beta_1$	10.68	0.178	<0.001	10.52	0.203	<0.001
SS	$\beta_2$	8.70	0.247	<0.001	7.95	0.275	<0.001
SC	$\beta_3$	0.037	0.027	0.169	0.071	0.039	0.066
SS	$\beta_4$	-0.002	0.027	0.945	0.068	0.033	0.038
SC	$\sigma_{SC}^2$	0.523	0.136	<0.001	0.319	0.129	0.014
SS	$\sigma_{SS}^2$	1.907	0.312	<0.001	0.988	0.231	<0.001
SC/SS	$\tau^2$	0.190	0.028	<0.001	0.171	0.008	<0.001

PROC MIXED. The variability of the best linear predictors (BLUPs) of the random effects appeared different in the two phenotypes (Figure 5(b)). The final LME models incorporates random effect variance components different by phenotype, but common variance of the error terms. Even accounting for different phenotype random effects variance components, normal quantile plots of residuals (Figure 5(c)), and especially of the BLUPs of the random effects (Figure 5(d)), suggest deviations from the normal distribution assumption.

Denote by  $y_{ij}$  the  $j^{\text{th}}$  Hb measure in the  $i^{\text{th}}$  patient of age  $x_{ij}$ . The model used to model dependence of Hb on age and phenotype was

$$y_{ij} = (\beta_1 + \beta_2 x_{ij}) I_{\{i \in \text{HbSC}\}} + (\beta_3 + \beta_4 x_{ij}) I_{\{i \in \text{HbSS}\}} + \gamma_i^{\text{SC}} I_{\{i \in \text{HbSC}\}} + \gamma_i^{\text{SS}} I_{\{i \in \text{HbSS}\}} + \varepsilon_{ij}, \quad (7.1)$$

where  $I_{\{i \in \text{HbSS}\}} = 1$  if patient  $i$  has HbSS phenotype and  $I_{\{i \in \text{HbSS}\}} = 0$  otherwise,  $I_{\{i \in \text{HbSC}\}} = 1$  if patient  $i$  has HbSC phenotype and  $I_{\{i \in \text{HbSC}\}} = 0$  otherwise,  $i = 1, \dots, 134$ ,  $j = 1, \dots, m_i$ ,  $1 \leq m_i \leq 4$ . It is assumed that  $\gamma_i^{\text{SS}} \stackrel{i.i.d.}{\sim} N(0, \sigma_{SS}^2)$  when patient  $i$  has HbSS phenotype,  $\gamma_i^{\text{SC}} \stackrel{i.i.d.}{\sim} N(0, \sigma_{SC}^2)$  if patient  $i$  has HbSC phenotype, and the  $\varepsilon_{ij} \stackrel{i.i.d.}{\sim} N(0, \tau^2)$  are independent for any  $i$  and  $j$ .

Table 1 presents the results of estimating (7.1) using the REML and the GS-estimator with Tukey biweight  $\rho$ -functions and breakdown parameter  $\epsilon = 0.5$ . The latter yielded the least biased results across all scenarios in the simulation study. The intercept estimates for HbSC and HbSS phenotypes are very similar for both estimation methods, and the intercepts are significantly different ( $p < 0.001$ ) between phenotypes using either the t-test with the standard LME model, or the Wald test described in Section 5 with GS-estimators. This result is consistent with all sickle cell disease literature. In contrast, the slope estimates are different for REML and robust GS estimation methods. The REML estimates  $\hat{\beta}_3 = 0.037$  and  $\hat{\beta}_4 = -0.002$ , and their difference, are not significantly

different from zero. Meanwhile, the GS-estimator with  $\epsilon = 0.5$  yields very similar  $\hat{\beta}_3 = 0.071$  and  $\hat{\beta}_4 = 0.068$ , with  $\hat{\beta}_4$  significantly ( $p = 0.038$ ) and  $\hat{\beta}_3$  borderline significantly ( $p = 0.066$ ) different from zero. Figures 5(e) and 5(f) demonstrate that the GS-estimator with  $\epsilon = 0.5$  successfully downweights the subgroup of potential outliers in HbSS phenotype with uncharacteristically high levels of Hb.

Use of the GS-estimator with  $\epsilon = 0.5$  suggests changing the conclusion about the age-related trend in HbSS phenotype from nonsignificant negative, using REML, to significant positive slope. The robust analysis results also indicate that age-dependent trends are essentially the same for the two phenotypes of sickle cell disease.

## 8. Discussion

We have developed a generalized S-estimator that provides foundation for further developments in robust inference for general unbalanced LME models, including construction of efficient MM-estimators and robust tests, similar to the work of Copt and Heritier (2007) for fully balanced LME models. The approximate covariance matrix of the fixed effects GS parameter estimates is relatively simple to compute. It is different from the covariance matrix for the maximum likelihood estimates only in terms of incorporating additional weights that depend on the dimension of independent vectors and breakdown parameter through the corresponding  $\rho$ -function. The asymptotic covariance matrix of covariance parameter estimates has rather complicated expression, but it is not required when the interest is in the fixed effects inference. It is expected that arguments similar to the ones used in Field, Pang, and Welsh (2010) can yield asymptotic validity of suitably selected bootstrap procedure(s), which would eliminate the need to evaluate the asymptotic covariance matrix of the covariance parameter estimates. Such developments will be a focus of further investigations.

Danilov, Yohai, and Zamar (2012) recently introduced a generalized S-estimator (GSE) for multivariate location and scatter in the presence of missing data. Both their estimator and ours reduce to the S-estimator for multivariate location and scale under the multivariate normal distribution model, but involve quite different model assumptions and objective functions for minimization.

In numerical studies, we use the Tukey biweight  $\rho$ -function, but alternative re-descending  $\rho$ -functions (e.g., proposed by Rocke (1996)) can be used instead. Our simulation results suggest that the advantage of the GS-estimator over Robust REML is most notable when there is a contamination in distribution of random effects or predictor variable. This result is consistent with known differences between S-estimators and M-estimators in simpler models with contamination in design variables.

## Acknowledgement

The authors thank Dr. Marie Stuart from Jefferson Medical College, Thomas Jefferson University and Dr. Carlton Dampier from the Emory University School of Medicine for permission to use sickle cell disease data. This work was supported by the grant from the US National Institutes of Health (NIH/NIAMSD AR054596). We thank an anonymous referee and an associate editor for constructive suggestions that resulted in substantial improvements.

## References

- Alqallaf, F., Van Aelst, S., Zamar, R., and Yohai, V. J. (2009). Propagation of outliers in multivariate data. *Ann. Statist.* **37**, 311-331.
- Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16**, 147-185.
- Bilodeau, M. and Duchesne, P. (2000). Robust estimation of the SUR model. *Canad. J. Statist.* **28**, 277-288.
- Copt, S. and Heritier, S. (2007). Robust alternatives to the F-Test in mixed linear models based on MM-estimates. *Biometrics* **63**, 1045-1052.
- Copt, S. and Victoria-Feser, M.-P. (2006). High-breakdown inference for mixed linear models. *J. Amer. Statist. Assoc.* **101**, 292-300.
- Danilov, M., Yohai, V. J. and Zamar, R. H. (2012). Robust estimation of multivariate location and scatter in the presence of missing data. *J. Amer. Statist. Assoc.* **107**, 1178-1186.
- Davies, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15**, 1269-1292.
- Field, C. A., Pang, Z. and Welsh, A. H. (2010). Bootstrapping robust estimates for clustered data. *J. Amer. Statist. Assoc.* **105**, 1606-1616.
- He, X. and Fung, W. K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *J. Multivariate Anal.* **72**, 151-162.
- Heritier, S., Cantoni, E., Copt, S. and Victoria-Feser, M.-F. (2009). *Robust Methods in Biostatistics*. John Wiley, Cornwall, UK.
- Huggins, R. M. (1993). A robust approach to the analysis of repeated measures. *Biometrics* **49**, 715-720.
- Huggins, R. M. and Staudte, R. G. (1994). Variance components models for dependent cell populations. *J. Amer. Statist. Assoc.* **89**, 19-29.
- Hasselman, B. (2009). nleqslv: Solve Systems of Non Linear Equations. R package version 2.0, URL <http://CRAN.R-project.org/package=nleqslv>.
- Künsch, H. R., Stefanski, L. A., and Carroll, R. J. (1989). Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *J. Amer. Statist. Assoc.* **84**, 460-466.
- Lopuhaä, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17**, 1662-1683.
- Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4**, 51-67.

- Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust Statistics*. John Wiley, Cornwall, UK.
- Maronna, R. A. and Yohai, V. J. (2000). Robust regression with both continuous and categorical predictors. *J. Statist. Plann. Inference* **89**, 197-214.
- Pinheiro, J. C., Liu, C., and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *J. Comput. Graph. Statist.* **10**, 249-276.
- Richardson, A. M. (1997). Bounded influence estimation in the mixed linear model. *J. Amer. Statist. Assoc.* **92**, 154-161.
- Richardson, A. M. and Welsh, A. H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics* **51**, 1429-1439.
- Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *Ann. Statist.* **24**, 1327-1345.
- Ruppert, D. (1992). Computing S-estimators for regression and multivariate location/dispersion. *J. Comput. Graph. Statist.* **1**, 253-270.
- Samanta, M. and Welsh, A. H. (2013). Bootstrapping for highly unbalanced clustered data. *Comput. Statist. Data Anal.* **59**, 70-81.
- Stahel, W. A. and Welsh, A. H. (1997). Approaches to robust estimation in the simplest variance components model. *J. Statist. Plann. Inference* **57**, 295-319.
- Tatsuoka, K. S. and Tyler, D. E. (2000). The uniqueness of S and M-functionals under non-elliptical distributions. *Ann. Statist.* **28**, 1219-1243.
- Van Aelst, S. and Willems, G. (2005). Multivariate regression S-estimators for robust estimation and inference. *Statist. Sinica* **15**, 981-1001.
- Varadhan, R., Gilbert, P. (2009). BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *J. Statist. Software* **32**, 1-26.
- Welsh, A. H. and Richardson, A. M. (1997). Approaches to the robust estimation of mixed models. In *Handbook of Statistics* **15** (Edited by G. S. Maddala and C. R. Rao), 343-384. Elsevier, Amsterdam.
- Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.* **15**, 642-656.

Division of Biostatistics, Thomas Jefferson University, Philadelphia, PA 19107, USA.

E-mail: Inna.Chervoneva@jefferson.edu

Division of Biostatistics, Thomas Jefferson University, Philadelphia, PA 19107, USA.

E-mail: Mark.Vishnyakov@jefferson.edu

(Received September 2012; accepted August 2013)