High Breakdown Inference in the Mixed Linear Model

COPT, Samuel, VICTORIA-FESER, Maria-Pia

Abstract
Mixed linear models are used to analyze data in many settings. These models have a multivariate normal formulation in most cases. The maximum likelihood estimator (MLE) or the residual MLE (REML) is usually chosen to estimate the parameters. However, the latter are based on the strong assumption of exact multivariate normality. Welsh and Richardson have shown that these estimators are not robust to small deviations from the multivariate normality. This means that in practice a small proportion of data (even only one) can drive the value of the estimates on their own. Because the model is multivariate, we propose a high-breakdown robust estimator for very general mixed linear models that include, for example, covariates. This robust estimator belongs to the class of S-estimators, from which we can derive the asymptotic properties for inference. We also use it as a diagnostic tool to detect outlying subjects. We discuss the advantages of this estimator compared with other robust estimators proposed previously and illustrate its performance with simulation studies and analysis of three datasets. We also consider robust [...]
High-Breakdown Inference for Mixed Linear Models

Samuel COPT and Maria-Pia VICTORIA-FESER

Mixed linear models are used to analyze data in many settings. These models have a multivariate normal formulation in most cases. The maximum likelihood estimator (MLE) or the residual MLE (REML) is usually chosen to estimate the parameters. However, the latter are based on the strong assumption of exact multivariate normality. Welsh and Richardson have shown that these estimators are not robust to small deviations from the multivariate normality. This means that in practice a small proportion of data (even only one) can drive the value of the estimates on their own. Because the model is multivariate, we propose a high-breakdown robust estimator for very general mixed linear models that include, for example, covariates. This robust estimator belongs to the class of $S$-estimators, from which we can derive the asymptotic properties for inference. We also use it as a diagnostic tool to detect outlying subjects. We discuss the advantages of this estimator compared with other robust estimators proposed previously and illustrate its performance with simulation studies and analysis of three datasets. We also consider robust inference for multivariate hypotheses as an alternative to the classical $F$-test by using a robust score-type test statistic proposed by Heritier and Ronchetti, and study its properties through simulations and analysis of real data.

KEY WORDS: Constrained covariance; Robust $F$-test; $S$-estimator; Variance components.

1. INTRODUCTION

Mixed linear models are powerful models for the analysis of data in many settings. They include, for example, ANOVA models with repeated measures (so-called “within-subjects” designs or “randomized block” designs), hierarchical or multilevel models (random nested designs), longitudinal data (repeated measures), and others.

Mixed linear models are generally based on the normality assumption, and inference based on the likelihood function is the approach used to estimate and test the parameters of the model (see, e.g., Searle, Casella, and McCulloch 1992). However, likelihood-based inference under the assumption of normality is well known to be nonrobust to small model deviations (see Huggins and Staudte 1994; Stahel and Welsh 1997; Welsh and Richardson 1997). We also derive estimating equations that can be solved in an iterative manner. It guaranties a high breakdown point for the resulting estimator and robust inference for the model’s parameters can be developed in a straightforward manner. In particular, a robust score-type test (Heritier and Ronchetti 1994) can be used as a robust alternative to the $F$-test.

The article is organized as follow. In Section 2 we briefly present the multivariate normal formulation of the mixed linear model. In Section 3 we define the high-breakdown estimator and compare it analytically with other robust estimators proposed for the mixed linear model proposed so far in the literature. We also discuss some numerical issues. We cover multivariate hypothesis robust testing in Section 4 and present a simulation study in Section 5. In Section 6 we show, through the analysis of real data, that the classical and the robust estimators and testing procedures can give a different insight to the data analysis. Finally, in Section 7 we conclude.

2. MODEL FORMULATION

Mixed linear models can be expressed generally by the regression equation of the form

$$\mathbf{Y} = \mathbf{X}\mathbf{\alpha} + \sum_{j=1}^{r} \mathbf{Z}_j \mathbf{\beta}_j + \mathbf{\epsilon}, \quad (1)$$

where $\mathbf{Y}$ is the $N$-vector of all measurements; $\mathbf{X}$ is an $N \times q_0$ design matrix for the fixed effects; the $\mathbf{Z}_j$ are the $N \times q_j$ design matrices for the random effects $\mathbf{\beta}_j$; $\mathbf{\epsilon}$ is the $N$-vector of independent residual errors, with $\mathbf{\epsilon} \sim N(0, \sigma^2_\mathbf{I}_N)$; $\mathbf{\alpha}$ is a $q_0$-vector of unknown fixed effects; and $\mathbf{\beta}_j$ are the unobserved $q_j$-vectors of independent random effects, with $\mathbf{\beta}_j \sim N(0, \sigma^2_{\mathbf{\beta}_j})$. It follows that $\mathbb{E}[\mathbf{Y}] = \mathbf{X}\mathbf{\alpha}$ and $\text{var}(\mathbf{Y}) = \sum_{j=0}^{r} \sigma^2_j \mathbf{Z}_j \mathbf{Z}_j^T = \mathbf{V}$, with $\sigma^2_0 = \sigma^2_\mathbf{\epsilon}$ and $\mathbf{Z}_0 = \mathbf{I}_N$. We assume that all of the $q_0 + r + 1$ effects are identifiable and concentrate on models for which we can write

$$\mathbf{V} = \text{diag}(\Sigma). \quad (2)$$

For such models, we have an equivalent multivariate formulation for (1), which is

$$\mathbf{y}_i \sim N(\mu, \Sigma), \quad (3)$$

with $\mathbf{y}_i$ the $p$-vector of independent observations obtained by partitioning $\mathbf{Y}$ according to the covariance structure in (2) and

$$\mu = \mathbf{X}\mathbf{\alpha}. \quad (4)$$

© 0 American Statistical Association
Journal of the American Statistical Association
???? 0, Vol. 0, No. 00, Theory and Methods
DOI 10.1198/00
with \( x \) a \( p \times q_0 \) matrix obtained by partitioning \( X \) according to the covariance structure in (2). Furthermore, for most well-known models (see, e.g., Copt and Victoria-Feser 2003), we can write

\[
\Sigma = \sum_{j=0}^{r} \sigma_j^2 z_j z_j^T, \tag{5}
\]

where \( z_j \) is a \( p \times q_j \) random-effects design matrix. We also discuss the case with the presence in the model of covariates in which

\[
\mu = \mu_i = x_i \alpha. \tag{6}
\]

Under the normality assumption (3), the maximum likelihood estimator (MLE) or the residual MLE (REML) introduced by Patterson and Thompson (1971) can be used to estimate the parameters. For the latter, it not possible to recover a structure like in (3); thus we do not pursue this route here.

### 3. ROBUST ESTIMATION

#### 3.1 Constrained S-Estimator

Using the multivariate formulation of mixed linear models given in (3), we propose building a robust estimator of multivariate mean and constrained covariance. Robust estimators of multivariate mean and (free) covariance have been regularly proposed in the literature since the work of Maronna (1976) (see Maronna and Yohai 1998 for a review). In principle they can be built from weighted score functions, but, as Maronna (1976) showed, if the weights are not redescending (no weights of 0), then their breakdown point becomes smaller as the dimension increases. Because in multivariate models in general and in the mixed linear model in particular the dimension can be large, it is important to consider so-called “high-breakdown” estimators. These include the class of \( S \)-estimators (Rousseeuw and Yohai 1984). Davies (1987) investigated the existence, consistency, asymptotic normality, and breakdown point of these estimators. It is known (Lopuhaä 1989) that an \( S \)-estimator of multivariate mean and covariance can be found by an iterative procedure; however, the latter usually have multiple solutions, so that the starting point for their computation is important. Here we propose to adapt an \( S \)-estimator to the case where the covariance matrix is constrained; for the starting point, we adapt the orthogonalized Gnanadesikan–Kettering (OGK) robust estimator proposed by Maronna and Zamar (2002).

For a sample \( (y_i), i = 1, \ldots, n \), approximately generated from (3), an \( S \)-estimator of multivariate mean and covariance is defined as the solution for \( \mu \) and \( \Sigma \) that minimizes \( \text{det}(\Sigma) = |\Sigma| \) subject to

\[
n^{-1} \sum_{i=1}^{n} \rho(\sqrt{\sum_{j=1}^{p} (y_{ij} - \mu_j)^2 - \sum_{j=1}^{p} (y_{ij} - \mu_j)}) = b_0, \tag{7}
\]

where \( b_0 \) is a parameter chosen to determine the desired breakdown point and \( \rho \) is a function with the properties given by, for example, Rousseeuw and Yohai (1984). In particular, \( \rho \) is such that \( E_\Phi(\rho(\cdot)) / \rho(\cdot) = \epsilon^*, \) where \( \Phi \) is the standard normal model, \( c = \max_{x_1 \rho(x)} \), and \( \epsilon^* \) is the desired breakdown point (up to 1/2), so that it can be chosen for a fixed value of the breakdown point. When the mean vector is as in (4) and the covariance matrix is as in (5), we show in Section A.1, that an iterative system for finding the \( S \)-estimator for the mean and variance components is given by

\[
\alpha = (x^T \Sigma^{-1} x)^{-1} \sum_{i=1}^{n} u(d_i) x_i^T \Sigma^{-1} y_i \tag{8}
\]

and

\[
S_0 = \left[ \frac{1}{n} \sum_{i=1}^{n} u(d_i) d_i^2 \right]^{-1} Q^{-1} U, \tag{9}
\]

where \( d_i = \sqrt{(y_i - \mu)^T \Sigma^{-1} (y_i - \mu)} \) are the Mahalanobis distances, \( u(d_i) = \frac{\alpha^2}{R(d_i)/d_i} = \psi(d_i)/d_i, \) \( S_0 = (\sigma_2^2, \ldots, \sigma_p^2)^T, \) \( Q = [\text{tr} \Sigma^{-1} z_j z_j^T \Sigma^{-1} z_k z_k^T]_{j,k=1,\ldots,r} \) and \( U = \left[ \frac{1}{n} \sum_{i=1}^{n} u(d_i) \right] \times \left( (y_i - x_\alpha)^T \Sigma^{-1} z_j z_j^T \Sigma^{-1} (y_i - x_\alpha) \right)_{i=1,\ldots,r}. \) When \( \mu = \mu_i = x_i \alpha, \) (8) becomes (see Sec. A.1)

\[
\alpha = \left[ \sum_{i=1}^{n} u(d_i) x_i^T \Sigma^{-1} y_i \right]^{-1} \sum_{i=1}^{n} u(d_i) x_i^T \Sigma^{-1} y_i. \tag{10}
\]

Finally, note that the distances \( d_i \) can be used as a diagnostic tool to detect outlying observations (see Sec. 6).

A usual choice for the function \( \rho \) is Tukey biweight (Beaton and Tukey 1974). Rousseeuw (1985) instead proposed a translated Tukey biweight that can control the so-called asymptotic rejection probability (ARP). The ARP can be interpreted as the probability of an estimator giving a null (or nearly null) weight to an extreme observation (in large dimensions). The translated biweight function leads to the weight function

\[
u(d; c, M) = \begin{cases} 1, & 0 \leq d < M \\ \left(1 - \left(\frac{d - M}{c}\right)^2\right)^2, & M \leq d \leq M + c \\ 0, & d > M + c, \end{cases} \tag{11}
\]

where the constants \( c \) and \( M \) can be chosen to achieve the desired breakdown point and ARP (see Sec. 3.3). Using (11) in (8) and (9) defines the constrained translated biweight \( S \)-estimator (CTBS). It should be noted that Wellmann (2000) also used an \( S \)-estimator, but his framework was limited to very simple models.

#### 3.2 Properties

The following proposition states the asymptotic normality of the CTBS of \( \alpha \) and \( S_0. \)

**Proposition 1.** Let \( (y_i), i = 1, \ldots, n \), be a sequence of independent random vectors with the \( p \)-variate normal distribution \( F_{\mu, \Sigma} \) with \( \mu \) of the form (4) and \( \Sigma \) positive definite and constrained as in (5). Suppose that the inverse of the \( q_0 \times q_0 \)-dimensional matrix \( \Sigma^{1/2} \) exists and that the \( p^2 \times (r+1) \)-dimensional matrix \( D = [\text{vec}(z_1 z_1^T) \cdots \text{vec}(z_r z_r^T)] \) is such that the inverse of \( D^T D \) exists, where the operator \( \text{vec}(z_j z_j^T) \) stacks the columns of the matrix \( z_j z_j^T \) into a \( p^2 \times 1 \) vector. Also, let \( \hat{\theta} = (\hat{\alpha}, \hat{S}_0) \) be the CTBS of \( \theta = (\alpha, S_0) \), that is, the \( S \)-estimator defined as the solution of (8) and (9) that minimizes \( \text{det}(\Sigma) = |\Sigma| \).
\[ |\Sigma| \text{ subject to (7) with translated biweight } \rho \text{-function given by } \]
\[
\rho(d; c, M) = \begin{cases} 
\frac{d^2}{2}, & 0 \leq d < M \\
\rho_{M \leq d \leq M+c}(d; c, M), & M \leq d \leq M+c \\
\frac{M^2}{2} - c(5c + 16M), & d > M+c,
\end{cases}
\]  
(12)

with \( M + c < \infty \) and
\[
\rho_{M \leq d \leq M+c}(d; c, M)
\]
\[
= \frac{M^2}{2} - \frac{M^2(M^4 - 5M^2c^2 + 15c^4)}{30c^4}
\]
\[
+ d^2 \left( \frac{5M^4}{2c^4} - \frac{M^2}{c^2} \right) + d^3 \left( \frac{4M^3}{3c^2} - \frac{4M^3}{3c^2} \right)
\]
\[
+ d^4 \left( \frac{3M^2}{2c^4} - \frac{2c^2}{c^2} \right) - \frac{4Md^5}{5c^4} + \frac{d^6}{6c^4},
\]  
(13)

and leading to the weight function \( u(d) \) given in (11). Then \( \sqrt{n}(\theta - \theta) \) has a limiting normal distribution with mean \( 0 \) and covariance
\[
\begin{bmatrix}
V_\alpha & 0 \\
0 & V_{S_0}
\end{bmatrix},
\]  
(14)

with
\[
V_\alpha = \frac{e_1}{e_2}(x^T x)^{-1} x^T \Sigma x(x^T x)^{-T},
\]  
(15)

and
\[
V_{S_0} = (D^T D)^{-1} D^T V_\Sigma D(D^T D)^{-1},
\]  
(16)

with \( V_\Sigma \) as given in (A.3).

For the proof see Section A.2.

Note that in our formulation we suppose that the dimension of the response vector \( y_i \) is the same \( \forall i \), which makes our results valid only in these cases. If the dimension varies because of missing data, then our results can in principle be extended using the results of Cheng and Victoria-Feser (2002) and Copt and Victoria-Feser (2004) on multivariate S-estimators with missing data.

Welsh and Richardson (1997) defined a very general class of robust estimators for the mixed linear model (1) based on bounded influence estimating equations (BIEEs) through
\[
\sum x_i^T W_{0i} \Sigma^{-1/2} \psi_0(\Sigma^{-1/2} U_{0i}(y_i - x_i \alpha)) = 0
\]  
(17)

for the fixed effects and
\[
\frac{1}{2} \sum \psi_1(\Sigma^{-1/2} U_{1i}(y_i - x_i \alpha)) W_{1i} \Sigma^{-1/2} z_j^T W_{1i} \Sigma^{-1/2} = \text{tr}(K_{2j}\Sigma^{-1/2} z_j^T W_{1i} \Sigma^{-1/2} W_{1i} W_{1i})
\]
\[
\times \psi_2(\Sigma^{-1/2} U_{1i}(y_i - x_i \alpha)) = 0
\]  
(18)

for each variance component \( \sigma_j^2 \). The matrices \( K_{2j} \) ensure consistency at the normal model (see Welsh and Richardson 1997).

Depending on the choice for the weight matrices \( W_{0i}, W_{1i}, U_{0i}, U_{1i} \), and \( \psi_0, \psi_1, \psi_2 \), we can define the Mallows, Andrews, Hill and Ryan, and Schweppe estimators (see Hampel, Ronchetti, Rousseau, and Stahel 1986) and also the robust MLE I, MLE II, REMLE I, and REMLE II of Richardson and Welsh (1995) and Richardson (1997) (see also Welsh and Richardson 1997). Gill (2000) proposed a robust estimator for the analysis of longitudinal data that is actually defined through (16) and (17) with \( W_{0i} = W_{1i} = U_{0i} = U_{1i} = I \), \( \psi_0 = \psi_1 = \psi_2 = \psi \), and \( \varphi(t) \) the identity function. In the context of generalized mixed linear models, Yau and Kuk (2002) proposed a robust estimator that in the normal case is equivalent to the robust MLEs I and II of Richardson and Welsh (1995). Similarly, Sinha’s (2004) robust estimator for generalized mixed linear models can be considered a Mallows-type estimator and hence defined through BIEE in the normal case.

The CTBS can also be defined through BIEE, because (10) can be written as
\[
\sum u(d_i) x_i^T \Sigma^{-1}(y_i - x_i \alpha)
\]
\[
= \sum ((y_i - x_i \alpha)^T \Sigma^{-1}(y_i - x_i \alpha))^{1/2}
\]
\[
\times (y_i - x_i \alpha)^T \Sigma^{-1/2}(y_i - x_i \alpha)^{1/2} x_i^T \Sigma^{-1/2}(y_i - x_i \alpha)
\]
\[
= \sum \psi_0(y_i, x_i; \theta)
\]
\[
= 0,
\]  
(19)

with \( \theta = (x^T, S_{ij}^2)^T \) and (9) as
\[
\sum \{\psi^2(u(d_i))(y_i - x_i \alpha)^T \Sigma^{-1} z_j W_{1i} \Sigma^{-1/2} \Sigma^{-1} u(d_i)\}
\]
\[
= \sum \{\psi^2(u(d_i)) \text{tr}[\Sigma^{-1} z_j W_{1i} \Sigma^{-1/2} \Sigma^{-1} u(d_i)]\}
\]
\[
= \sum \psi_0(y_i, x_i; \theta)
\]
\[
= 0,
\]  
(20)

where \( \psi(d) = da(d) \). Therefore, by setting \( W_{0i} = W_{1i} = U_{0i} = U_{1i} = I, \forall i, \psi_0(v) = \psi_1(v) = \psi_2(v) = v \), \( \forall i, K_{2j} = u(d_i)d_i^2 \) in (16) and (17), one gets the estimating equations defining the CTBS. But the CTBS has the important difference that the weight function \( u \) (or indeed \( \psi \)) is based on the translated biweight \( \rho \)-function and hence is redescending, so that the breakdown point of the CTBS can be controlled. To our knowledge, all robust estimators belonging to the general class of BIEE estimators are based on a nonredescending \( \psi \)-function, such as Huber’s function \( \psi(v) = \min(1; c|v|) \).

From (18), we can deduce that \( \alpha \) is Fisher-consistent; that is, \( \frac{1}{n} \sum_{i=1}^{n} \int \psi_0(y, x_i; \theta) dF_{\mu, \Sigma}(y|x_i) = 0 \), where \( F_{\mu, \Sigma} \) is the normal distribution with mean \( \mu \) and covariance \( \Sigma \). Indeed, letting \( \Sigma = B B^T \) and \( v_i = B^{-1}(y|x_i - x_i \alpha) \), which
has a p-variate standard normal distribution $\Phi \forall i$, we have
\[ \frac{1}{\alpha} \sum_{i=1}^{p} \int (y - x_i\alpha)^T \Sigma^{-1} (y - x_i\alpha)^{-1/2} \psi \left( (y - x_i\alpha)^T \Sigma^{-1} \times (y - x_i\alpha) \right)^{1/2} \Sigma^{-1} \times \int u((v^T v)^{1/2}) v dF(v) = 0 \] by symmetry of the normal model.

The function $\psi_{\alpha}$ also satisfies the conditions for the resulting estimator to be asymptotically normal (see, e.g., Welsh 1996, p. 194). Moreover, supposing that the inverse of $\sum_{i=1}^{n} x_i^T x_i$ exists, we have that
\[ \text{var}(\alpha) = \frac{\alpha_1}{\alpha_2} \left( \sum_{i=1}^{n} x_i^T x_i \right)^{-1} \sum_{i=1}^{n} x_i^T \Sigma x_i \left( \sum_{i=1}^{n} x_i^T x_i \right)^{-1} (19) \]

When considering REML, it is not possible to recover a structure like in (3) so that an $S$-estimator cannot be defined.

However, one could consider the $\psi$-function corresponding to the CTBS in the estimating equations of for example the robust REML II of Richardson and Welsh (1995). The computational aspects might then become difficult not only due to the computation of the consistency corrections, but also because an iterative procedure as in (8) or (10) and (9) might be impossible to derive. Moreover, as we illustrate through a simulation study in Section 5, the CTBS seems to have very little, if any, bias in small samples.

An $S$-estimator with the translated biweight $\rho$-function has an asymptotic breakdown point of $e^* = \text{E}_{\rho}[x^2]/\text{max}_{d} \rho(d)$ and $M$ (Davies 1987). This is shown by, for example, computing (one version of) the finite breakdown point and letting $n \to \infty$ (see, e.g., Davies 1987). The fact that the covariance matrix $\Sigma$ is constrained does not modify the breakdown properties of the $S$-estimator. In the most constrained case when $\Sigma = \sigma^2 I$, for example, the result holds (see Rousseeuw and Yohai 1984).

### 3.3 Computational Aspects

To compute the CTBS, we must first set the parameters $M$ and $c$ for fixed values of the breakdown point $e^*$ and an ARP value $\alpha$ by means of $e^* \max_{d} \rho(d; c, M) = \text{E}_{\rho}[x^2]/\text{max}_{d} \rho(d; c, M)$ and $e^* = \sqrt{\text{E}[(x^2)^{-1}]}(1 - \pi)$ (see Rocke 1996). In our data analysis we chose the values of $e^* = .50$ and $\alpha = .01$. Also note that to compute the CTBS, we actually rescale the distances by a factor $k$, that is, $d_{i,j,k}$ with $k = d_{i,j}/\sqrt{\text{E}[(x^2)^{-1}]}(1/\alpha + 1)$, where $d_{i,j}$ denotes the $q$th ordered distance and $q = [(n + p + 1)/2]$ (see Rocke 1996).

To find a solution to the system (8) and (9), we need a starting point. We propose using the weighted sample means and covariances, with weights
\[ w_i = \begin{cases} 1 & \text{if } (y_i - \tilde{\mu}_{\text{OGK}})^T \tilde{\Sigma}_{\text{OGK}}^{-1} (y_i - \tilde{\mu}_{\text{OGK}}) \leq \lambda_2^2 \tau \\ 0 & \text{otherwise} \end{cases} \]
with $\tilde{\mu}_{\text{OGK}}$ and $\tilde{\Sigma}_{\text{OGK}}$ the adjusted OGK of Maronna and Zamar (2002) (see Copt and Victoria-Feser 2003) and $\tau = .9$. Then, given the structure of $\Sigma$, the means of the estimated covariances corresponding to the equal elements in $\Sigma$ are taken (see Copt and Victoria-Feser 2003 for a more detailed discussion). The same is done for $\alpha$. In the case of models with continuous covariates, we propose adopting a very pragmatic approach as proposed by Pinheiro et al. (2001). As suggested by a referee, instead of reweighted sample means and covariances based on the OGK, we could compute the (unrestricted) $S$-estimator with the algorithm of Ruppert (1992) and then adjust it to satisfy the constraints. But it is not clear which approach would lead to better estimates in the sense that they are faster to compute, and we leave the problem of choosing the starting point to future research.

Finally, the dimension $p$ of the model and the number of parameters (fixed and variance components) relative to the number of observations $n$ also play important roles. We study this point with a simulation in Section 5.

### 4. ROBUST TESTING

We are interested here in testing the null hypothesis that $q < 0$ linearly estimable functions of the vector of parameters $\alpha$ are 0. We treat the variance components as nuisance parameters. Let $\alpha^T = (\alpha^T_1, \alpha^T_2)$ denote the partition of the vector $\alpha$ into $q > q_0$ and $q$ components and let $A_{ij}, i, j = 1, 2, \alpha_0$ unspecified against $H_1: \alpha_0 \neq 0$, $\alpha_1$ unspecified. As an example, consider a so-called "one factor within subject" design, that is, $y_{ij} = \mu + \lambda_i + s_i + e_{ij}$ with $\sum_{i=1}^{n_1} \lambda_i = 0, s_i \sim N(0, \sigma^2_s)$ $\forall i = 1, \ldots, n, e_{ij} \sim N(0, \sigma^2_e)$ $\forall i = 1, \ldots, n, j = 1, \ldots, l$, so that in our notations
\[ \alpha = [\mu, \lambda_1, \ldots, \lambda_{l-1}]^T; \quad x = \left[ e_1 \right] \left[ \frac{1}{15} \right] \left[ e_{l-1} \right] \]
with $e_1$ an $l$-dimensional vector of 1’s; $(\sigma^2_s, \sigma^2_e) = (\sigma^2_s, \sigma^2_e)$; $z_0 z_0^T = I_l$ and $z_1 z_1^T = J_{l,1}$, where $J_{l,1}$ is a matrix of 1’s of dimension $l \times l$; and thus $q_0 = l (= p)$. A sensible hypothesis would be $H_0: \lambda_1 = \cdots = \lambda_{l-1} = 0, \mu$ unspecified against $H_1$: one of the $\lambda_i \neq 0, \mu$ unspecified so that $q = l - 1$.

Here we consider robust score-type test statistics (Heritier and Ronchetti 1994), which for a sample $(y_i, i = 1, \ldots, n)$ of an iid random vector generated from $F_\alpha$, is $R_\alpha = Z_\alpha^T C^{-1} Z_\alpha$, where $Z_\alpha = \frac{1}{n} \sum_{i=1}^{n} \psi(y_i, \alpha^0_{(2)})$ and $\alpha^0$ is the $M$-estimator in the reduced model; that is, the solution of the equation $\sum_{i=1}^{n} \psi(y_i, \alpha^0_{(1)}) = 0$ with $\alpha^0(2) = 0, C = M(22)^T M^{-1}(22), I_{n1} = \int \frac{\partial}{\partial \alpha} \psi(y, \alpha) dF_\alpha(y) = \int \psi(y, \alpha) dF_\alpha(y)$ and $(y, \alpha)$ is the score function of model $F_\alpha$. We also have that $H = M^{-1} K M^{-T}$, with $K = \int \psi(y, \alpha)^2 dF_\alpha(y)$ and $M = \int \psi(y, \alpha) dF_\alpha(y)$ and $M$ and $K$ both evaluated at $\alpha^0$. In practice, we actually use $M = \frac{1}{n} \sum_{i=1}^{n} \psi(y_i, \alpha^0_{(2)})$ and $K = \frac{1}{n} \sum_{i=1}^{n} \psi(y_i, \alpha^0_{(2)})^2$. For the CTBS estimator of $\alpha$, the $\psi$-function is $\psi_{\alpha}$ given in (18).

Heritier and Ronchetti (1994) showed that under $H_0$ and under mild conditions for $\psi$ (existence, consistency, and Fréchet differentiability), the statistic $R_\alpha^2$ is asymptotically $\chi^2$-distributed. Conditions (A.1)–(A.9) of Heritier and Ronchetti (1994) are satisfied when the $\psi$-function is $\psi_{\alpha}$, because $\tilde{\alpha}$ is Fisher-consistent and $\psi_{\alpha}$ is (twice) differentiable with respect to $\alpha$. Because the results are only asymptotic, we investigate in a simulation study the effect of sample size on the empirical level of the tests in Section 5.
5. SIMULATION STUDY

In this section we compare the performance of the CTBS in terms of robustness, efficiency, and small-sample behavior with the classical estimators and with the robust MLE I and II and robust REML I and II of Richardson and Welsh (1995). We also study the performance of the robust test for multivariate hypothesis and compare it with that of the classical F-test.

5.1 Comparing Estimators

Richardson and Welsh (1995) reported a simulation study that analyzed the performance of robust estimators in terms of bias and mean squared error (MSE). Here we perform the same simulation study and compare the same statistics to the ones they found.

Richardson and Welsh (1995) proposed a data-generating model \( y_{ij} = \alpha_1 + \alpha_2 x_{ij} + \beta_j + \epsilon_{ij} \) with \( \beta_j, j = 1, \ldots, 4 \), following a \( N(0, \sigma_{\beta}^2) \) and \( \epsilon_{ij}, i = 1, \ldots, 5, j = 1, \ldots, 4 \), following a \( N(0, \sigma_{\epsilon}^2) \). The \( x_{ij} \) were generated from a \( N(0,1) \) and then orthogonalized so that the \( 20 \times 2 \) matrix \( X = [1 \ x_{ij}] \) is such that \( X^T X = 20 I_2 \). The parameters values were chosen as \( \alpha_1 = \alpha_2 = \sigma_{\beta}^2 = \sigma_{\epsilon}^2 = 1 \). Data contamination was then allowed in either or both of the random-effect distributions, with the form \((1-\epsilon)N(0,1) + \epsilon N(0,11)\) with \( \epsilon = .10 \). The possible resulting data-generating models are denoted by \((0:0), (0:1), (1:0), \) and \((1:1)\), corresponding to the true model, the contaminated model on the second random effect \( \epsilon \), the contaminated model on the first random effect \( \beta \), and the contaminated model on both random effects. We generated 200 samples with each type of model and computed a robust bias statistic \( B = \text{median} (\hat{\theta}) - \theta \) for each parameter. (Note that \( \theta = 1 \) for all four parameters.) Figure 1 shows the results for the CTBS together with the boxplots of the estimates’ distributions. Compare the first line of our Figure 1 with the first line of figure 5 of Richardson and Welsh (1995) involving the MLE, the REML, and the robust MLE I, MLE II, REML I, and REML II. We can conclude the following. For the fixed effects \( (\alpha_1 \text{ and } \alpha_2) \), the performance of all estimators is very similar; all the biases are of the same magnitude and very small. For the variance components, the CTBS appears to be biased for \( \sigma_{\beta}^2 \) as measured by \( B \), but the bias is very small (i.e., the boxplots do not show an appreciable bias). For \( \sigma_{\epsilon}^2 \), the bias is nearly zero. On the other hand, the performance in terms of bias of the estimators of Richardson and Welsh (1995) appears to depend on the contamination settings. In practice, this means that the choice of a robust estimator should be made depending on the type of contamination expected. This is not the case with the CTBS, which is very stable across all contamination settings.

We also studied the performance of the CTBS under other contamination settings. More precisely, instead of normal mixtures for the approximate model, we considered mixtures with a point mass, that is, \((1-\epsilon)N(0,1) + \epsilon A_z\), with \( z \) an arbitrary value, as well as mixtures with lognormal distributions. Across all settings (contamination type, model, value of parameters, etc.), the CTBS remains very stable, whereas the MLE (or REML) can be seriously biased for some or all fixed effects and/or variance components. As an illustration, we consider the two-factors-within-subjects model

\[
\begin{align*}
\gamma_{ijk} &= \mu + \lambda_j + \gamma_k + (\lambda\gamma)_{jk} + s_i + (\lambda s)_{ij} + (\gamma s)_{ik} + \epsilon_{ijk},
\end{align*}
\]

\( i = 1, \ldots, n, j = 1, \ldots, l, k = 1, \ldots, g \), in which \( s_i \sim N(0, \sigma_{s}^2) \), \( (\lambda s)_{ij} \sim N(0, \sigma_{s\lambda}^2) \), \( (\gamma s)_{ik} \sim N(0, \sigma_{s\gamma}^2) \), and \( \epsilon_{ijk} \sim N(0, \sigma_{\epsilon}^2) \) \( \forall i, j, k \) are then random effects. (For the multivariate normal parameterization, see Sec. A.3). We chose \( \mu = 24, \lambda_j = \gamma_k = (\lambda\gamma)_{jk} = 0 \) \( \forall j, k = 1, \ldots, 3, \sigma_{s}^2 = 10, \sigma_{s\lambda}^2 = 6, \sigma_{s\gamma}^2 = 4 \), and \( \sigma_{\epsilon}^2 = 50 \). We contaminated the observations by generating 2\% (out of \( n = 100 \)) of the residual errors from a lognormal distribution with mean 3 and variance .25. The bias distribution of the

![Figure 1. Robust Bias Statistic and Bias Distribution for the CTBS for Different Contamination Models.](image-url)
estimators are presented in Figure 2. The MLE for the residual variance is completely biased, and estimates of the fixed effects are (slightly) biased. The CTBS, on the other hand, is very stable for all parameters. It should be stressed that Copt and Victoria-Feser (2003) present real data for which the two-factors-within-subject design we found that the CTBS numerically converges and remains very stable (unbiased) under all settings for a very small sample size and different sample sizes, and show that the CTBS efficiency is compared to the MLE, under different contamination setting and different sample sizes, and show that the CTBS remains very stable (unbiased) under all settings for a very small efficiency loss at the model.

Finally, the associate editor suggested studying the behavior of the CTBS in nearly overparameterized problems. In our experience we found that the CTBS numerically converges with samples sizes \( n \approx q_0 + r + 1 \). Copt and Victoria-Feser (2003) presented real data for which the two-factors-within-subjects model (20) with \( j, k = 1, \ldots, 3 \) is appropriate. The CTBS estimates are \((\mu, \lambda_1, \lambda_2, \gamma_1, \gamma_2, (\lambda \gamma)_{11}, (\lambda \gamma)_{12}, (\lambda \gamma)_{21}, (\lambda \gamma)_{22})^T = (430, 9, -6, -16, 11, 5, -2, 0, -2)\) and \((\sigma_1^2, \sigma_2^2, \sigma_{\gamma 1}^2, \sigma_{\gamma 2}^2) = (649, 158, 261, 733)\) (see also Sec. 6 for a similar analysis). Hence there are 13 parameters (9 means and 4 variance components). We generated samples of size \( n = 9 \) (i.e., 81 measurements) from the estimated model and estimated the parameters using the MLE, REML, and CTBS. With smaller sample sizes, the CTBS did not systematically converge. Figure 3 presents the bias distribution of the estimators for the variance components. For the fixed effects, all estimators present no apparent bias.) As expected, the REML shows no bias, whereas the MLE is biased for some variance components. However (which may come as a surprise), the CTBS has a behavior comparable to the one of the REML; that is, it shows no apparent bias. This result is actually not that surprising, because it has been observed for some time now that robust estimators or test statistics can have better small-sample behavior than their classical counterparts (see, e.g., Victoria-Feser 1997 in the context of robust model choice).

5.2 Comparing Testing Procedures

The aim of this simulation study is to compare the actual levels of the tests, which are fixed a priori, with the empirical levels given by the simulations. If the test behaves well, then only small differences between those two levels can be expected. We compared the performance of the classical and robust score tests as well as the classical F-test, which is known to be exact under the normal model. As a model, we considered a two-factors-within-subject design (20); for the samples, we considered uncontaminated and contaminated ones, with different sample sizes.

Supposing three levels for each factor, there are three sets of natural hypothesis, two sets for the main effects and one set for the interaction. For example, for the main effect of the first factor, we have

\[ H_0 : \lambda_1, \lambda_2 = 0 \text{ and } \mu, \gamma_1, \gamma_2, (\lambda \gamma)_{11}, (\lambda \gamma)_{12}, (\lambda \gamma)_{21}, (\lambda \gamma)_{22} \text{ unspecified}; \]

\[ H_1 : \lambda_1 \neq 0 \text{ or } \lambda_2 \neq 0 \text{ and } \mu \gamma_1, \gamma_2, (\lambda \gamma)_{11}, (\lambda \gamma)_{12}, (\lambda \gamma)_{21}, (\lambda \gamma)_{22} \text{ unspecified}. \]  

The results are presented in Table 1 for hypothesis (21), without data contamination for \( n = 20 \). With the larger sample size \( n = 100 \), all three tests behaved nicely in that the empirical levels corresponded to the actual ones (results not presented here), but with the smaller sample size, both score-type tests were a bit conservative for the low actual levels. We found the same results for the other hypotheses (second main effect and interaction).

To investigate the robustness properties of the tests on the levels, we contaminated the data by replacing a proportion, \( \epsilon \), randomly chosen by data generated from the model with parameters \( \mu + \lambda_1 = 48, \mu + \lambda_1 = \mu + \lambda_2 = 48, \) or \( \mu + \lambda_1 = \mu + \lambda_2 = 48, \) and \( \mu + \lambda_1 = \mu + \lambda_2 = 48, \) and \( \mu + \lambda_1 = \mu + \lambda_2 = 48, \) and \( \mu + \lambda_1 = \mu + \lambda_2 = 48, \) and \( \mu + \lambda_1 =\)
The study of semantic and associative priming in picture naming is well known in psychology (see, e.g., Alario, Segui, and Ferrand 2000; Holcomb and McPherson 1994). The data that we have come from an experiment in which subjects had to decide as quickly as possible whether a target (object’s drawing) appearing after a prime (action of a pantomime) was a real object or not (see Moy and Mounoud 2003). The delay between the pantomime and the appari-tion of the object was either short or long, the pantomime was of one of three types (related, neutral, or unrelated), and several different real objects were used.

### 6. DATA ANALYSIS

In this section we analyze a real dataset to evaluate the impact on the analysis of the robust method compared with the classical method. We compute the REML and the CTBS and corresponding standard errors for the fixed effects. As a diagnostic tool, we also provide a scatter plot of the Mahalanobis distances $d_i$ estimated using respectively the REML and the CTBS. Finally, we also test some natural multivariate hypotheses using the classical F-test and the robust score-type test. Other datasets were also analyzed by Copt and Victoria-Feser (2003).

The study of semantic and associative priming in picture naming is well known in psychology (see, e.g., Alario, Segui, and Ferrand 2000; Holcomb and McPherson 1994). The data that we have come from an experiment in which subjects had to decide as quickly as possible whether a target (object’s drawing) appearing after a prime (action of a pantomime) was a real object or not (see Moy and Mounoud 2003). The delay between the pantomime and the appari-tion of the object was either short or long, the pantomime was of one of three types (related, neutral, or unrelated), and several different real objects were used.

### Table 1. Proportion of Times (21) Is Rejected, With $n = 20$

<table>
<thead>
<tr>
<th>$\epsilon = 0%$</th>
<th>Actual levels</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical scores</td>
<td></td>
<td>.2%</td>
<td>3.9%</td>
<td>10.1%</td>
</tr>
<tr>
<td>Robust scores</td>
<td></td>
<td>.4%</td>
<td>4.1%</td>
<td>9.9%</td>
</tr>
<tr>
<td>F-test</td>
<td></td>
<td>1.1%</td>
<td>5.1%</td>
<td>9.5%</td>
</tr>
</tbody>
</table>

For each combination of objects, type of pantomime, and delay, five measures (i.e., time to decide if the object was real or not) were taken, of which the first measure (trial) and the errors (wrong object decision) were discarded and the means of the remaining were taken as the response variable. The underlying hypothesis is that the reaction time is shorter when there is a link between the priming (pantomime) and the object and an interaction with the delay. We consider here the subsample involving only the object “broom,” with 21 old subjects (age 70 and over).

The model used to analyze these data is given in (20) with $\lambda_j, j = 1, \ldots, 3$, the pantomime type (PT) and $\gamma_k, k = 1, 2$, the delay (DE). Table 3 gives the estimates for the REML and the CTBS and the standard errors for the contrasts (bold values are for significant contrasts at the 5% level). Figure 4 shows a scatterplot of the Mahalanobis distances computed with the REML and CTBS. The horizontal and vertical lines correspond to the quantile 97.5% of a $\chi^2_6$, that is, the asymptotic distribution of the Mahalanobis distances. One can see that both estimators detect one clear outlier (no. 19), and CTBS detects one influential observation (no. 12). We also tested the significance of each factor and each interaction (i.e., three hypotheses) using the F-test and the robust score-type test. The results, presented in Table 4, show that the outlier and the influential observation seem to exert quite substantial influence not only on the size of the effect estimates, but also on the conclusions of the main effects and interaction testing. With the F-test, only the main effects (PT and DE) are found to be significant, whereas with

### Table 3. Estimates and Standard Errors (SEs) for the REML and the CTBS for the Semantic Priming Data Fixed Effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CTBS</th>
<th>REML</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>603.35</td>
<td>633.43</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-52.54</td>
<td>-51.22</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>43.71</td>
<td>32.65</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>17.51</td>
<td>18.07</td>
</tr>
<tr>
<td>$\lambda_1\gamma_1$</td>
<td>-11.64</td>
<td>-16.80</td>
</tr>
<tr>
<td>$\lambda_2\gamma_1$</td>
<td>25.71</td>
<td>13.11</td>
</tr>
</tbody>
</table>

NOTE: Values in bold denote significant parameters at the 5% level.
the robust score-type test, the interaction between DE and PT is also found to be significant.

7. CONCLUSION

In this article we have proposed a high-breakdown estimator for very general mixed linear models. We have developed an S-estimator, namely the CTBS, and derived the estimating equations so that it can be implemented in a relatively straightforward manner. We have also proposed using a robust test statistic for multivariate hypotheses (i.e., a robust counterpart to the F-test). Through a real dataset, we have shown that a robust analysis including robust inference can provide other insights. It should be stressed that although the framework of this article is mixed linear models, the CTBS actually can be used for any model with a constrained covariance matrix. Indeed, the $z_j z_j^T$ in (5) can be considered as the derivative of $\Sigma$ with respect to the parameters of interest, so that any constrained covariance matrix can be written as in (5). Finally, the $R$ functions for computing robust estimators for models that can be expressed as in (3) with (6) and (5) are available from the authors on request.

APPENDIX: ???

A.1 Iterative System for the S-Estimator

The Lagrangian $L$ associated with the constraint (7) is

$$L = \log(||\Sigma||) + \lambda \left[ \frac{1}{n} \sum \rho(d_i) - b_0 \right].$$

Taking derivatives yields

$$\frac{\partial L}{\partial \mu} = \left( \frac{\partial \mu}{\partial \mu} \right)^T \frac{\partial L}{\partial \mu} = -\lambda \sum u(d_i)x^T \Sigma^{-1}(y_i - x\alpha) = 0,$$

$$\frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial \alpha} = \text{tr} [\Sigma^{-1} z_j z_j^T]$$

$$= -\frac{\lambda}{2n} \sum u(d_i)(y_i - x\alpha)^T \Sigma^{-1} z_j z_j^T \Sigma^{-1} (y_i - x\alpha) = 0.$$

From (A.1), we get (8). When $\mu_j = x_j \alpha$, we have simply $\left( \frac{\partial \mu_j}{\partial \mu} \right)^T \frac{\partial L}{\partial \mu} = -\lambda \sum u(d_i) x_j^T \Sigma^{-1}(y_i - x_j \alpha) = 0$, which reduces to (10).

For the covariance matrix, we take the sum of (A.2) in $j$ and get

$$\sum_{j=0}^r \text{tr} [\Sigma^{-1} z_j z_j^T]$$

$$-\frac{\lambda}{2n} \sum_{j=0}^n u(d_i)(y_i - x\alpha)^T \Sigma^{-1} z_j z_j^T \Sigma^{-1} (y_i - x\alpha) = 0,$$

$$= \frac{1}{n} \sum_{i=1}^n p u(d_i)(y_i - x\alpha)^T \Sigma^{-1} z_j z_j^T \Sigma^{-1} (y_i - x\alpha) = 0,$$

$$= 0.$$
It follows that $\rho$ has a continuous derivative, $\rho(0) = 0$, $\rho$ is strictly increasing on $(0, c_0)$, $c_0 = M + c$ and constant on $(c_0, \infty)$, $u(d, c, M) = \psi(d; c, M)/d$ and $\frac{d}{d}\psi(d; c, M)$ are bounded at $M + c$ and continuous. Because the model is the p-variate normal distribution, which is an elliptical distribution, then all conditions of corollary 5.1 of Lopuhaä (1989) are satisfied for $\mu$ and $\Sigma$ to have asymptotically independent normal distributions. The asymptotic covariance of $\sqrt{n} \hat{\mu}$ is $\frac{c_1}{n^{\infty}} \Sigma$, and that of $\sqrt{n} \hat{\Sigma}$ is

$$V_\Sigma = e_1 (I_{p,q} + K_{p,p}) \Sigma \otimes \Sigma + e_4 \text{vec}(\Sigma) \text{vec}(\Sigma)^T,$$

(A.3)

in which $\otimes$ denotes the Kronecker product, $K_{p,p}$ is a $p^2 \times p^2$-block matrix with the $(i,j)$-block being a $p \times p$ matrix with 1 at entry $(i, j)$ and 0 elsewhere, and $e_3$ and $e_4$ are as given by Lopuhaä [1989, eq. (5.5)]. $\alpha$ and $S_0$ are actually linear reparametrizations of $\mu$ and $\Sigma$, defined as $\mu = \alpha x$ and $\Sigma(S_0) = \sum_{j=0}^p \sigma_j^2 z_j^T z_j$. Using standard results on linear functions of random vectors, the asymptotic properties of $\hat{\alpha}$ follow immediately. For $S_0$, given that $\text{vec}(\Sigma(S_0)) = \sum_{j=0}^p \sigma_j^2 \text{vec}(z_j z_j^T)$ and $\frac{2}{\sqrt{8}} \text{vec}(\Sigma(S_0)) = D$, we have that $\sqrt{n}(S_0 - S_0)$ is asymptotic normal with asymptotic covariance given by (15).

### A.3 Two-Factors-Within-Subjects Design

The structural equation (20) leads to a $p = q \cdot l$ multivariate normal model $N(\alpha x, \sum_{j=0}^3 \sigma_j^2 z_j^T z_j)$, where

$$\alpha = (\mu, \lambda_1, \ldots, \lambda_{l-1}, \gamma_1, \ldots, \gamma_q - 1),$$

$$(\lambda \gamma^1, \lambda \gamma^2, \ldots, \lambda \gamma^{(l-1)(q-1)})^T$$

and

$$x = \begin{bmatrix} e_1 & e_4 \\ e_5 & \bigotimes I_{l-1} \end{bmatrix} \begin{bmatrix} z_1^T \\ -e_1^T \end{bmatrix} = \begin{bmatrix} I_{l-1} \otimes e_1 & e_4 \\ -e_5 & I_{l-1} \otimes e_1 \end{bmatrix} \begin{bmatrix} I_{q-1} \otimes e_1 & e_4 \\ -e_5 & I_{q-1} \otimes e_1 \end{bmatrix}.$$