Robust *MM*-Estimation and Inference in Mixed Linear Models

Samuel Copt and Stephane Heritier

No 2006.01

Cahiers du département d'économétrie
Faculté des sciences économiques et sociales
Université de Genève

Janvier 2006

Département d'économétrie
Université de Genève, 40 Boulevard du Pont-d'Arve, CH -1211 Genève 4
http://www.unige.ch/ses/metri/

ROBUST *MM*-ESTIMATION AND INFERENCE IN MIXED LINEAR MODELS

Samuel Copt and Stephane Heritier

NHMRC Clinical Trials Centre, University of Sydney

Abstract: Mixed linear models are used to analyse data in many settings. These models generally rely on the normality assumption and are often fitted by means of the maximum likelihood estimator (MLE) or the restricted maximum likelihood estimator ($REML$). However, the sensitivity of these estimation techniques and related tests to this underlying assumption has been identified as a weakness that can even lead to wrong interpretations. Recently Copt and Victoria-Feser (2005) proposed a high breakdown estimator, namely an S-estimator, for general mixed linear models. It has the advantage of being easy to compute - even for highly structured variance matrices - and allow the computation of a robust score test. However this proposal cannot be used to define a likelihood ratio type test which is certainly the most direct route to robustify an F-test. As the latter is usually a key tool to test hypothesis in mixed linear models, we propose two new robust estimators that allow the desired extension. They also lead to resistant Wald-type tests useful for testing contrasts and covariate effects. We study their properties theoretically and by means of simulations. An analysis of a real data set illustrates the advantage of the new approach in the presence of outlying observations.

Key words and phrases: Mixed models, Robustness, MM-estimator, Breakdown point, Likelihood ratio test, Wald test.

1 Introduction

Mixed linear models are very popular models when there are multiple sources of error and are widely used in many scientific fields. Estimation of parameters

The first author acknowledges the support of the Swiss National Science Foundation (grant no PP001–106465)

in these models is usually a preliminary step to inference and the primary goal of many experimental designs is more often than once hypothesis testing. An example of such a design is given in Moy and Mounoud (2003). The data come from an experiment in which 23 old subjects (between 60 and 65) had to decide as quickly as possible if a target (object's drawing), which appeared after a prime (action of a pantomime), was a real object or not. The delay between the pantomime and the apparition of the object was either short or long and the pantomimes were of three types (related, unrelated and neutral). For each combination of pantomime and delay five measures (time to decide wether the object is real or not) were taken, with the first and last one discarded and the mean of the remaining ones taken as the response variable. The underlying hypothesis is that the reaction time is shorter when there is a link between the priming and the object and researchers suspect an interaction with the delay. A two-way ANOVA model with repeated measures can be fitted to these data, namely

$$y_{ijk} = \mu + \lambda_j + \gamma_k + (\lambda\gamma)_{jk} + s_i + (\lambda s)_{ij} + (\gamma s)_{ik} + \varepsilon_{ijk}, \tag{1}$$

with $i = 1, .., 21,\ \ j = 1, ..., 2$ and $k = 1, ..., 3$. μ is the grand mean, λ_j, γ_k are the fixed effects for the delay and the priming respectively and $(\lambda\gamma)_{jk}$ is the interaction factor between the two fixed effects. s_i is the effect due to subject i, which we assume to be a random variable $N(0, \sigma_s^2)$; $(\lambda s)_{ij}$ and $(\gamma s)_{ik}$ are interaction random variables with distribution $N(0, \sigma_{\lambda s}^2)$ and $N(0, \sigma_{\gamma s}^2)$ respectively; and ε_{ijk} is the error term coming from $N(0, \sigma_\varepsilon^2)$. We assume that all the variables on the right-hand side are independent.

A second example is described in Pinheiro, Liu, and Wu (2001) and was originally reported by Potthoff and Roy (1964). The data come from an orthodontic study on 16 boys and 11 girls between the ages of 8 and 14. The response variable is the distance (in millimeters) between the pituitary and the pterygomaxillary fissure, which was measured at 8, 10, 12 and 14 years for each boy and girl. Pinheiro et al. (2001) suggest that a potential model for these data is

$$y_{ijt} = \beta_0 + \beta_1 t + (\beta_{0g} + \beta_{1g} t)\, J_i(j) + \gamma_{0i} + \gamma_{1i} t + \varepsilon_{ijt},$$

with y_{ijt} the response for the $i^t h$ subject $(i = 1, \ldots, 27)$ of sex j ($j = 1$ for boys and $j = 2$ for girls) at age $t = 8.10, 12, 14$,

$$J_i(j) = \begin{cases} 0 & j = 1 \\ 1 & j = 2, \end{cases}$$

a dummy variable for sex. $\beta_0, \beta_1, \beta_{0g}, \beta_{1g}^T$ are the fixed effects and $\gamma_{0i}, \gamma_{1i}, \varepsilon_{ijt}$ the random effects with zero mean and respective variances of $\sigma^2_{\gamma_0}, \sigma^2_{\gamma_1}, \sigma^2_{\varepsilon}$. This model is actually a random slope and intercept model.

Both of the above models belong to the class of mixed linear model which is of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \sum_{j=1}^{r} \mathbf{Z}_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}, \tag{2}$$

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

$$\mathbf{V} = \mathrm{diag}(\boldsymbol{\Sigma}). \tag{3}$$

The usual procedures to derive estimates and tests for the various parameters in (2), namely the maximum likelihood estimator (MLE) or the restricted maximum likelihood estimation ($REML$) of Patterson and Thompson (1971), rely heavily on the normality assumption - see for instance Searle, Casella, and McCulloch (1992) for a review. Small departures from normality can have disastrous effects on estimators (bias) and tests (increased type I error), see e.g. Welsh and Richardson (1997) and Copt and Victoria-Feser (2005). Several alternative estimation techniques that are far less sensitive to model misspecifications have been proposed in the last decade; see Huggins and Staudte (1994), Stahel and Welsh (1997), Richardson and Welsh (1995), Richardson (1997) and Welsh and Richardson (1997). They are mainly based on a weighted version of the likelihood function. Although some of the above proposals can theoretically deal with leverage points, they are technically very difficult to compute and for that reason virtually unused. A high breakdown estimator has recently been suggested by Copt and Victoria-Feser (2005) with the major advantage to be intuitive and computationally simple. Far less has been said on how to robustify classical tests for mixed models. Researchers are 'de facto' led to use the asymptotic standard errors and 95% confidence intervals with the notable exception of the robust score test proposed by Copt and Victoria-Feser (2005). As F-tests and contrast tests are more commonly used in practise we would like to generalise this approach and propose robust counterparts that can cope with outliers in the covariates, have reasonable breakdown properties, and retain the simplicity of the previous proposal.

The paper is organised as follows. In section 2 we reformulate the mixed linear model in a convenient way and briefly review the high breakdown estimator (i.e. an S-estimator) proposed by Copt and Victoria-Feser (2005) in that setting. Two new estimators are introduced in Section 3 to pave the way for further testing developments. They can be seen as M or MM-type estimators based on the initial highly robust estimates of the scale parameters σ_j's. Unlike the initial S-estimator, these new proposals allow the construction of robust alternatives to the F-test presented and studied in Section 4. They will also be used to define a robust Wald test typically useful for testing contrasts. A simulation study illustrating the behavior of the new procedures is then carried out in Section 5. An analysis of a real data set is presented in Section 6 showing the benefit of our approach over the standard procedures. Finally, Section 7 concludes.

2 High breakdown estimation approach in mixed linear models

We review briefly the approach proposed by Copt and Victoria-Feser (2005) to extend the definition of multivariate S-estimators to mixed linear models. The key idea is to reformulate a mixed linear model in term of multivariate normal distribution with a structured covariance matrix. Specifically, model (2) can be rewritten as:

$$\mathbf{y}_i|\mathbf{x}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \tag{4}$$

with $\mathbf{y}_i$ the p-vector of independent observations obtained by partitioning $\mathbf{Y}$ according to the covariance structure in (3) and

$$\boldsymbol{\mu}_i = \mathbf{x}_i\boldsymbol{\alpha}, \tag{5}$$

with $\mathbf{x}_i$ a $p \times q_0$ matrix obtained by partitioning $\mathbf{X}$ according to the covariance structure in (3). When there is no covariate $\mathbf{x}_i = \mathbf{x}\ \forall i$. Following (Copt and Victoria-Feser 2005), we can write

$$\boldsymbol{\Sigma} = \sum_{j=0}^{r} \sigma_j^2 \mathbf{z}_j \mathbf{z}_j^T, \tag{6}$$

where $\mathbf{z}_j$ is a $p \times q_j$ random effects design matrix. Formula (6) clearly specifies the structure of the covariance matrix of $\mathbf{y}_i$ arising from the random part of the model. In the rest of paper we will assume that the $\mathbf{z}_j$'s are fixed or at least well controlled.

A high breakdown point estimator, namely an S-estimator can then be easily adapted to the model. When the mean vector is as in (5) and the covariance matrix Σ *is structured* as in (6), Copt and Victoria-Feser (2005) introduced an S-estimator for the mean and variance components as the solution for $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ which minimizes $\det(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|$ subject to

$$n^{-1}\sum_{i=1}^{n} \rho\left(\sqrt{(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha})}\right) = b_0, \tag{7}$$

where ρ is a function having the properties given in Rousseeuw and Yohai (1984) and b_0 a parameter typically chosen to achieve a pre-specified breakdown point. Let $d_i = d_i(\boldsymbol{\alpha}) = \sqrt{(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha})}$ be the Mahalanobis distance for observation i, $\mathbf{S}_0$ be the vector of random effects parameters, i.e. $\mathbf{S}_0 = (\sigma_0^2, ..., \sigma_r^2)^T$ and $\boldsymbol{\theta} = \left(\boldsymbol{\alpha}^T, \mathbf{S}_0^T\right)^T$ the overall parameter. Then straightforward calculations show that an S-estimator for the fixed effects $\boldsymbol{\alpha}$ is solution of

$$\sum u(d_i)\mathbf{x}_i^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha}) = \sum \Psi_{\boldsymbol{\alpha}}(\mathbf{y}_i, \mathbf{x}_i; \theta) = 0, \tag{8}$$

and for the random effects

$$
\begin{aligned}
& \sum \left\{ pu(d_i)(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha})^T \boldsymbol{\Sigma}^{-1}\mathbf{z}_j\mathbf{z}_j^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha}) - u(d_i)d_i^2 \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{z}_j\mathbf{z}_j^T\right]\right\} \\
= & \sum \Psi_{\sigma_j^2}\left(\mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta}\right) = 0, \qquad (9)
\end{aligned}
$$

where $u(d_i) = \frac{\partial}{\partial d_i}\rho(d_i)/d_i = \psi(d_i)/d_i$ and $\psi(d_i) = d_i u(d_i)$.

Both equations (8) and (9) are M-type equations for the overall parameter $\boldsymbol{\theta}$, i.e. satisfy $\sum \Psi(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) = 0$ with $\Psi = (\Psi_{\boldsymbol{\alpha}}^T, \Psi_{\sigma_1^2}, ..., \Psi_{\sigma_r^2})^T$.

They may have multiple roots but Copt and Victoria-Feser (2005) showed that this difficulty can be easily overcome to find the S-solution. This can be achieved by solving an iterative system derived from (8) and (9) using a good high breakdown estimator as a starting point, e.g. the OGK estimator by Maronna and Zamar (2002). They also recommended to use the translated Tukey's biweight of Rocke (1996) as the ρ function but other choices can also be made. We will follow that path for simplicity. The estimator is called $CTBS$ for constrained translated biweight S-estimator. Using similar arguments as Davies (1987) and Lopuhaä (1989), Copt and Victoria-Feser (2005) showed Fisher consistency and asymptotic normality for the resulting estimator.

3 Other robust alternative estimators

Our purpose it to define a robust alternative to an F-test for mixed models. As F-tests are asymptotically equivalent to likelihood ratio tests (LRT) it seems natural to look for a robust version possibly based on the $CTBS$ estimator. Attempts in this direction rely on M-estimators and have been suggested first by Ronchetti (1982) in linear regression and then Heritier and Ronchetti (1994) in a more general framework. Unfortunately, despite that the $CTBS$ estimator (or any S-estimator) is asymptotically equivalent to an M-estimator the same approach collapses in the mixed effects model. This is mainly due to the constraint (7) as we shall see in Section 4. We propose two new procedures to overcome this problem. For both of them, the idea is to dissociate the estimation of the fixed effects from that of the random component. In other words we propose to obtain first a highly-robust consistent estimate for the covariance matrix via the $CTBS$

estimator, say $\hat{\boldsymbol{\Sigma}}_S$, then use a different robust procedure for the regression part $\boldsymbol{\alpha}$ holding the variance parameter fixed and equal to the previous estimate. This modification may look minimal but is actually essential to define robust LRT tests as we will explain in Section 4.

3.1 3.1. Huber estimator

Let us assume that $\boldsymbol{\Sigma}$ is known and that the only remaining parameter to estimate is the vector of fixed effects $\boldsymbol{\alpha}$. Following Lopuhaä (1992), the Huber estimator can naturally be defined as the solution for $\boldsymbol{\alpha}$ of the minimisation problem: $\min \sum \rho\big(d_i(\boldsymbol{\alpha})\big)$ where ρ is the Huber objective function (quadratic in the middle and linear in the tails). Equivalently it can be obtained by solving the first order equation with

$$\Psi_H\left(\mathbf{y}, \mathbf{x}; \boldsymbol{\alpha}\right) = u_c(d)(\mathbf{x}^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{x}\boldsymbol{\alpha})), \tag{10}$$

and $u_c(d)$ is the Huber-weight defined as usual as $u_c(d) = \min(1, c/|d|)$ with c the tuning constant controlling the desired efficiency. As $\boldsymbol{\Sigma}$ is unknown we will typically replace it in (10) by a preliminary HBP estimate $\hat{\boldsymbol{\Sigma}}_S$ - see Lopuhaä (1992) - yielding

$$\sum u_c(d_i)\mathbf{x}_i^T\hat{\boldsymbol{\Sigma}}_S^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha}) = 0. \tag{11}$$

This estimator can thus be seen as a two-stage estimator. In a first stage, a consistent high breakdown estimator of $\boldsymbol{\Sigma}$ is obtained via the $CTBS$ estimator. Then in a second stage, we estimate the fixed effect parameter by using an Huber M-estimator with $\boldsymbol{\Sigma}$ being held constant and equal to $\hat{\boldsymbol{\Sigma}}_S$ for all practical purposes. As long as only the response vector $\mathbf{y}$ is concerned and the design matrix $\mathbf{x}$ is fixed, techniques analogues to Lopuhaä (1992) show that the M-estimate of $\boldsymbol{\alpha}$ defined through (11) with a Huber objective function ρ will inherit at least the breakdown point of $\hat{\boldsymbol{\Sigma}}_S$. Things are radically different when the breakdown point is considered with respect to both $\boldsymbol{x}$ and $\boldsymbol{y}$ as the breakdown point is simply zero. The reason is that the influence function (IF) which measures the worst asymptotic bias caused to an estimator - see Hampel, Ronchetti, Rousseeuw, and Stahel (1986) - is unbounded in that case. Even one single observation, namely a leverage point, can ruin the estimator. To see this simply remark that the IF of an M-estimator is proportional to its defining ψ function in general and in the present case the function Ψ_H is unbounded in $\boldsymbol{x}$. This estimator has therefore the traditional drawback of all Huber estimates: it can be severely biased in the presence of contamination in the factor space or (bad) leverage points. This difficulty will be overcome with our next proposal. Note though that the Huber estimator retains its full potential for all designs involving only factors, categorical variables or well controlled covariates (e.g. ANOVA), situations that are frequently encountered in practice.

3.2 MM estimator

The class of MM-estimators was first introduced by Yohai (1987) in the linear regression setting. Such estimates are interesting as they combine high efficiency and high breakdown point in a simple and intuitive way. Typically one starts first with a highly-robust regression estimator, typically an S-estimator. Then one can use the scale based upon this preliminary fit along with a better tuned ρ function to obtain a more efficient M-estimator of the regression parameter. In practice the initial regression estimator is based on a loss function ρ_0, the final estimator on ρ_1 and both functions are related to each other via $\rho_0(u) = \rho(u/c_0)$ and $\rho_1(u) = \rho(u/c_1)$ with $0 < c_0 < c_1$ - see remark 2.3 in Yohai (1987). Tuning constants need to be adjusted to achieve a specific breakdown point and efficiency at the model. A multivariate version of this method was later suggested by Lopuhaä (1992). We will simply extend this approach to mixed linear models. Let us assume we have two functions ρ_0 and ρ_1 satisfying the conditions (A1) of Yohai (1987) and the remark above. An MM-estimator of $\boldsymbol{\alpha}$ is then defined as any solution of an M-type equation where

$$\Psi_{MM}(\mathbf{y}, \mathbf{x}; \boldsymbol{\alpha}) = u_{MM}(d)(\mathbf{x}^T \boldsymbol{\Sigma}_S^{-1}(\mathbf{y} - \mathbf{x}\boldsymbol{\alpha})). \tag{12}$$

This looks similar to the previous proposal. The difference with the Huber estimator lies in the definition of the weight function $u_{MM}(d)$ now based on a redescending score. Technically we have $u_{MM}(d) = \frac{\partial}{\partial d}\rho_1(d)/d = \psi_1(d)/d$ where ψ_1 *is redescending* as it is the derivative of a bounded loss function ρ_1, e.g. chosen in the Tukey's biweight family:

$$\rho_B(d; c) = \begin{cases} 3(\frac{d}{c})^2 - 3(\frac{d}{c})^4 + (\frac{d}{c})^6 & |d| \leq c \\ 1 & |d| \geq c. \end{cases} \tag{13}$$

In practice $\mathbf{\Sigma}$ is replaced by its high breakdown estimate and, as before, the MM-estimator is a solution of

$$\sum u_{MM}(d_i)\mathbf{x}_i^T \hat{\mathbf{\Sigma}}_S^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha}) = 0. \tag{14}$$

The exact definition of the functions ρ's and choice of tuning constants will be given in Section 4. The proper solution to (14) with a high breakdown point can be obtained via the iteratively reweighed algorithm of Rocke (1996), using as a starting point the initial regression S-estimate and the variance matrix being held fixed. As in Section 3.1 for the Huber estimator the MM-estimator has a breakdown point with respect to $\boldsymbol{y}$ that is at least equal to the one of $\widehat{\mathbf{\Sigma}}_S$. The advantage is that it now performs better when leverage points are present as illustrated in Section 5. Although its global robustness properties are not fully derived here, it is clear that the breakdown point of the MM-estimator with respect to $(\mathbf{y}, \mathbf{x})$ is positive. Recently multivariate regression S-estimators have been proposed by Van Aelst and Willems (2005) and their breakdown property extensively studied. It is likely that their approach can be extended to MM-estimators in mixed effects models. The formal derivation of the breakdown point of MM-estimators in this setting is however beyond the scope of this paper.

3.3 Asymptotic distribution

Proposition 1. *Let $(\mathbf{y}_i|\mathbf{x}_i), i = 1, \ldots, n$ be a sequence of independent random vectors conditionally distributed as a p-variate normal distribution $F_{\boldsymbol{\alpha},\mathbf{\Sigma}}$ with mean $\boldsymbol{\mu} = \mathbf{x}\boldsymbol{\alpha}$ and variance $\mathbf{\Sigma}$ positive definite and structured as in (6). Let K be the distribution of a covariate matrix $\mathbf{x}$ and suppose that the $q_0 \times q_0$ dimensional matrix $\mathbf{\Gamma} = E_K\left[\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right]$ exists and is invertible. For ψ chosen as either (10) or (12) we denote $\widehat{\boldsymbol{\alpha}}$ a solution of the corresponding equation. Then $\sqrt{n}\,(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ has a limiting normal distribution with zero mean and covariance matrix $\mathbf{H} = \mathbf{M}^{-1}\mathbf{Q}\mathbf{M}^{-T}$ where $\mathbf{M} = e_1\mathbf{\Gamma}$ and $\mathbf{Q} = e_2\mathbf{\Gamma}$. As both matrices are $\mathbf{M}$ and $\mathbf{Q}$ are proportional a simpler representation for $\mathbf{H}$ can be given*

$$\mathbf{H} = \frac{e_1}{e_2^2}\mathbf{\Gamma}^{-1} = \frac{e_1}{e_2^2}E_K\left[\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right]^{-1}, \tag{15}$$

where

$$
\begin{aligned}
e_1 &= \frac{1}{p} E_\Phi \left[d^2 u(d)^2 \right] && (16) \\
e_2 &= E_\Phi \left[u(d) + \frac{1}{p} d \frac{\partial}{\partial d} u(d) \right], && (17)
\end{aligned}
$$

u is the weight function associated to the chosen ψ function and Φ is the standard p-variate normal distribution.

Proof. See Appendix 4.7 □

Formula (15) shows that the asymptotic covariance matrix of our two proposals is proportional to that of the MLE. The scalar e_1/e_2^2 depends only on tuning constants and the definition of ρ as shown in (16) and (17). It is typically calibrated to achieve a specific efficiency at the model; see next section for details. In practice the asymptotic variance (15) is estimated by its sample counterpart to avoid any specific distributional assumption on $\mathbf{x}$. More details will be given in the appendix. As indicated in Lopuhaä (1992) any preliminary affine-equivariant covariance estimator with a high breakdown point that tends to $\mathbf{\Sigma}$ with probability one could be used instead of $\widehat{\mathbf{\Sigma}}_S$. The proposition would still hold in that case.

4 Robust tests

Testing in mixed linear models is probably the central issue. In a crossover trial one would like to test for example if there is a significant difference between two or several treatments or whether specific contrasts are significant. In the first example of Sec. 1, based on a two-way ANOVA model, the main issue for the researcher is clearly to see whether the reaction time is shorter when a link between the priming and the object exists or whether the delay impact on the reaction time, possibly interacting with the type of pantomime. This can be easily formulated using the notation introduced in Sections 1-2. Basically we are interested in testing the null hypothesis that q $(< p)$ linearly estimable functions of the vector of parameters $\boldsymbol{\alpha}$ are zero, the variance components being treated as nuisance parameters. Denote by $\boldsymbol{\alpha}^T = (\boldsymbol{\alpha}_{(1)}^T, \boldsymbol{\alpha}_{(2)}^T)$ the partition of the vector $\boldsymbol{\alpha}$ into $p-q$ and q components and by $A_{(ij)}, i, j = 1, 2$ the corresponding partition of $p \times p$ matrices. Through a linear transformation of the parameters, the hypotheses to be tested can be reformulated as

$$\begin{aligned} H_0 &: \boldsymbol{\alpha} = \boldsymbol{\alpha}_0 \text{ where } \boldsymbol{\alpha}_{0(2)} = 0, \, \boldsymbol{\alpha}_{0(1)} \text{ unspecified} \\ H_1 &: \boldsymbol{\alpha}_{0(2)} \neq 0, \, \boldsymbol{\alpha}_{(1)} \text{ unspecified.} \end{aligned}$$

The need for robust testing in this setting is obvious as classical F-tests and contrast tests have reportedly been found unreliable under sometimes mild deviations; see for instance Welsh and Richardson (1997) and Copt and Victoria-Feser (2005). In the robustness paradigm, robust tests must have i) a stable type I error under small, arbitrary departures from the null hypothesis (robustness of validity), ii) a good power under small arbitrary departures from the specified alternative (robustness of efficiency). In principle such tests and the related theory results exist in a very general framework. For instance Heritier and Ronchetti (1994) proposed a robust version of the Wald, score and LRT tests for general parametric models. This follows earlier work in linear regression by Ronchetti (1982) and others in the linear model. We will not discuss the case of the score type test as it has already been implemented in Copt and Victoria-Feser (2005) but we will focus on the two other alternatives.

4.1 Difficulty inherent to robust LRT statistics

Unfortunately the approach based on a likelihood-ratio type test statistic which is probably the most natural route to robustify the F-test presents an intrinsic difficulty. Before we explain the problem let us review the basic idea. In linear regression the classical LRT is based on the difference of sum of squares basically computed at the full and reduced models with the respective maximum likelihood estimators plugged in. It is therefore natural to base a robust LRT on a difference in a dispersion (or loss) function properly chosen where this time robust M-estimators have been substituted for the regression parameter $\boldsymbol{\alpha}$ in the full and reduced models. This idea can be more generally extended to any log-likelihood based test as shown by Heritier and Ronchetti (1994). However such an approach requires a very stringent condition to hold for the theory to be valid. Indeed the partial derivative of the dispersion function $\rho(u;\boldsymbol{\alpha})$ with respect to the parameter (usually denoted $\psi(u;\boldsymbol{\alpha})$) must be bounded to guarantee robustness of the resulting testing procedure. As this condition is difficult to fulfill outside the linear model, robust LRT procedures have been barely used with the notable exception of the work on robust deviances by Cantoni and Ronchetti (2001) in generalised linear models. In the mixed models framework the situation is similar. The approach by Copt and Victoria-Feser (2005) does not lead to a proper LRT statistic either as the latter simply vanishes in that case. To see this, just notice that the S-estimator proposed in this work is asymptotically equivalent to an M-estimator but requires the constraint (7) to hold under both the full and reduced models. The left-hand term is precisely *what should be chosen* as the test statistic. Since it is constrained to be b_0 for the estimator to exist, the resulting LRT statistic is zero. In the following section we will show that the estimators we have introduced in Section 3 do not present this drawback and allow for the desired extension.

4.2 Robust likelihood-Ratio and Wald type tests

Let us reintroduce $d_i(\boldsymbol{\alpha}) = \sqrt{(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha})^T \hat{\boldsymbol{\Sigma}}_{\mathbf{S}}^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\alpha})}$ be the Mahalanobis distance for observation i. The robust LRT test is defined by a test statistic of the form

$$S_n^2 = \frac{2}{n}\sum_{i=1}^{n}\left[\rho\big(d_i(\hat{\boldsymbol{\alpha}})\big) - \rho\big(d_i(\dot{\boldsymbol{\alpha}})\big)\right], \tag{18}$$

where $\hat{\boldsymbol{\alpha}}$ and $\dot{\boldsymbol{\alpha}}$ are the robust estimators in respectively the full and reduced model and ρ the corresponding loss function. More specifically the LRT test statistic associated to the Huber estimator is defined through (18) where $\hat{\boldsymbol{\alpha}}$ and $\dot{\boldsymbol{\alpha}}$ are the solution of (11) in respectively the full and reduced model, ρ is the Huber dispersion function:

$$\rho_H(d) = \begin{cases} \frac{1}{2}d^2 & |d| \le c \\ -\frac{1}{2}c^2 + c\,|d| & |d| \ge c \end{cases}, \tag{19}$$

and c is a tuning constant controlling efficiency at the model. The LRT test associated to the MM-estimator can be defined in a very similar way but use a redescending dispersion function, typically Tukey's biweight (13). The corresponding statistic is then given by (18) where $\hat{\boldsymbol{\alpha}}$ and $\dot{\boldsymbol{\alpha}}$ are solution of (14) in respectively the full and reduced model, ρ is simply

$$\rho_{MM}(d) = \rho_B(d; c_1), \tag{20}$$

and c_1 is chosen to achieve a specific efficiency as above. The initial starting point and the variance matrix estimate $\hat{\boldsymbol{\Sigma}}_S$ are S-estimators based on $\rho_0(u) = \rho_B(u, c_0)$ with $c_0 < c_1$. Of course it is also possible to translate the biweight as in Copt and Victoria-Feser (2005) or use other dispersion functions as in Rocke (1996) but we stick to this common choice for simplicity.
A robust Wald-type test statistic is naturally defined by

$$W_n^2 = \hat{\boldsymbol{\alpha}}_{(2)}^T \hat{\mathbf{H}}_{(22)}^{-1} \hat{\boldsymbol{\alpha}}_{(2)},$$

where $\hat{\boldsymbol{\alpha}}_{(2)}$ is the robust estimator of $\boldsymbol{\alpha}_{(2)}$ in the full model and $\hat{\mathbf{H}}_{(22)}$ the corresponding variance estimate. Subscripts indicating the type of estimator used are omitted as before. This definition can be easily extended to testing contrasts or more generally null hypotheses of the type $\mathbf{L}\boldsymbol{\alpha} = 0$ where $\mathbf{L}$ is a specific fixed matrix.
The constants c, c_1 introduced above are generally tuned to achieve a predetermined efficiency at the model. Another option is available for the Huber tests. It stems from the fact that (19) has for argument d, the Mahalanobis distance. As d^2 has a chi-square distribution with p degrees of freedom χ^2_p, c can be chosen as the square-root of a specific quantile of this distribution. In the simulation below we followed that option and set $c = \sqrt{(\chi^2_p)^{-1}(.90)}$. Regarding the MM-tests, both c_0 and c_1 with $c_0 < c_1$ have to be set accordingly for the tests to be properly defined. The constant c_0 is normally chosen to ensure a high asymptotic breakdown point for the initial estimate $\hat{\boldsymbol{\Sigma}}_S$, 50% in our case.

Table 1 about here

The other constant c_1 is computed to achieve a predetermined efficiency of the estimator at the model, 95% in this paper. Both constants depend on the dimension of the parameter and can be obtained by Monte-Carlo simulations. Table 1 summarises these values for different p when the function ρ is defined as above in (20). When p gets large enough an asymptotic approximation given in Rocke (1996) p. 1330 can be used, i.e. $c_1 = \sqrt{p}/M$ where M is defined by $\rho_B(M) = .5\rho_B(M)$ with ρ_B as in (13) with $c = 1$. This formula already gives reasonable results when $p > 10$. Note that the values of c_0 and c_1 depend on the choice of the dispersion function and must therefore be recomputed if another ρ function is to be used.

4.3 Asymptotic distribution and robustness properties

The general theory developed in Heritier and Ronchetti (1994) applies to the robust LRT tests defined above. We have the following proposition.

Proposition 2. *If the score function defining the estimator is like in (11) or in (12), then under H_0 the corresponding statistic nS_n^2 is asymptotically distributed as the weighted sum of q independent chi-square variables with one degree of freedom. The weights are simply the q positive eigenvalues of the matrix $\mathbf{Q}[\mathbf{M}^{-1} - (\mathbf{M}^*)^+]$ and $(\mathbf{M}^*)^+$ is a $p \times p$ matrix where blocks (12), (21), (22) are zero and block (11) is $\mathbf{M}^{-1}_{(11)}$. The matrices $\mathbf{M}$ and $\mathbf{Q}$ refer to the corresponding estimator and are defined as in Section 3.3.*

Proof. See Appendix 4.7 □

The asymptotic distribution obtained for the Huber or MM-estimator defined in Section 3 translates easily to the Wald test. If the score function defining the estimator is like in (11) or in (12), then under H_0 the statistic nW_n^2 is asymptotically χ^2_q.

Robustness properties of the tests can theoretically be studied using the techniques of Heritier and Ronchetti (1994) for the Wald test in general parametric models and Cantoni and Ronchetti (2001) for the LRT test in generalised linear models. The idea is to define a "neighborhood" of the null hypothesis that shrinks around it at a rate of $1/\sqrt{n}$ and study the asymptotic level of the test under any (contaminated) distribution in this neighborhood. One can show that the LRT or Wald-type tests of H_0 have a stable asymptotic level if the influence function of the underlying M-estimator is bounded. More exactly robustness of validity is guaranteed if the second part of the influence function of the underlying M-estimator, i.e. the component related to the parameter to be tested $\alpha_{(2)}$, is bounded. Similar derivations can also be carried out for the power and show that the same condition holds to ensure robustness of efficiency; see Ronchetti and Trojani (2001).

The same approach can be applied here to the tests based on the Huber-estimator assuming that the variance parameters are known as in Hampel, Ronchetti, Rousseeuw, and Stahel (1986), chapter 7. Because Ψ_H is a bounded function of the response $\mathbf{y}$ so is the influence function of the corresponding estimator. This in turn guarantees the stability of the level of the Huber LRT and Wald test provided that problems occur in the response only. In the presence of leverage points both procedures collapses as the underlying Huber estimator itself breaks down in that case - this will be further illustrated in Section 5. For the tests based on the MM-estimator the same argument could be used. A slight difficulty arises as such an estimator is initially based on a minimisation problem and is therefore only *asymptotically equivalent* to an M-estimator. It turns out that its influence function which is proportional to Ψ_{MM} is bounded in $\mathbf{y}$ but not in $\mathbf{x}$. However as the MM-estimator has a positive breakdown point this global robustness property is carried over to the testing procedures as illustrated in the simulation study. In other words, the level of either the LRT or Wald type test based on the MM-estimator is stable even in the presence of leverage points (outliers in the covariates).

4.4 Simulation

In this section we study the behavior of the robust Wald and likelihood ratio tests defined in Section 3 through a simulation study. In principle a few potential contenders can be considered: the robust and classical Wald tests, the robust and classical likelihood ratio tests computed with the M-estimator or the MM-estimator and the F-test. We want to study the performance of the different procedures under various model misspecifications. We specifically focus on the level of the tests by comparing the theoretical type I error, which is fixed a priori, and the experimental ones given by the simulations. If the test behaves well, one can expect small differences between those two levels. Two different designs will be included in the simulations, one with fixed carriers and one with random covariates, enabling contamination in both the response and the design matrix. The first design we consider is the one way ANOVA model with repeated measures given by the equation

$$y_{ij} = \mu + \lambda_j + s_i + \varepsilon_{ij}.$$

Values for the model's parameters are $\mu = 85$, $\lambda_j = 0 \ \forall j = 1, \ldots 4$, $\sigma_s = 10$, $\sigma_\varepsilon = 4$ and $n = 100$. The values for the parameter are the ones usually encountered when measuring diastolic blood pressure. For example, one of the reasons behind such an experiment could be the need to compare different treatments (say 4) and to see whether there are differences between them. Recall that, for the parameters to be identifiable we need $\Sigma_{j=1}^{l} \lambda_j = 0$. In this case $\boldsymbol{\alpha} = [\mu, \lambda_1, \lambda_2, \lambda_3]^T$ and the hypothesis 'there is no measuring difference' is stated as

$$\begin{aligned} H_0 &: \lambda_1, \lambda_2, \lambda_3 = 0, \ \mu \text{ unspecified} \\ H_1 &: \lambda_1, \lambda_2, \lambda_3 \neq 0, \ \mu \text{ unspecified.} \end{aligned}$$

To create a small model deviation $(1 - \epsilon)\%$ of the data are generated from the multivariate normal distribution with parameters $\boldsymbol{\mu} = [85 \ 85 \ 85 \ 85]$ and $\boldsymbol{\Sigma} = 10^2 \mathbf{J}_4 + 8^2 \mathbf{I}_4$ with $\mathbf{J}_4$ being a 4×4 matrix of ones. $\epsilon\%$ of the data are generated from a multivariate normal distribution with the same covariance matrix, but with a shifted mean $\boldsymbol{\mu} = [90 \ 85 \ 85 \ 80]$ or $\boldsymbol{\alpha} = [85, 5, 0, 0]^T$. This type of model deviation produces so-called shift outliers (Woodruff and Rocke 1994) which are supposed to be the most difficult to be detected. We generated 10000 samples under the null hypothesis and recorded the proportion of times the null hypothesis was rejected for different amounts of contamination. The nominal level is chosen to be 5%. The results are summarized in Table 2.

Table 2 about here

We can see that all the tests perform similarly when the distribution of the response is indeed normal. The observed levels are all close to each other and to the nominal level of 5%. When contamination is introduced all robust tests remain stable irrespectively of the percentage of contamination. Meanwhile the classical LRT and F tests exhibit a larger type I error than expected with for example an observed level of more than 19% for $\epsilon = 5\%$ with the classical F-test. Note that only the LRT tests and F-test are presented here. The results for the Wald tests are omitted but are similar to those of the LRT tests in their classical and robust derivations.

The second design includes continuous covariates. To motivate such a design imagine an experiment where an adjusted analysis has to be performed with the inclusion of a laboratory parameter (e.g. triglycerides, or white blood cell count) as a covariate in the model. Such a variable can be viewed as a predictor or potential confounder than can take on naturally large values making the identification of gross errors more difficult by routine checks. This design will thus allow us to study the behavior of the tests when the design matrix is contaminated, i.e. with the presence of leverage points.
Suppose that we observe $n = 100$ subjects at different points of time $t = 1, 2, 3, 4$. A simple model including a continuous covariate could be expressed by the following equation.

$$y_{it} = \mu + \lambda e_i + \gamma x_{it} + s_i + \varepsilon_{it},$$

where e_i could be a dummy variable, x_{it} the continuous covariate for the laboratory measurement measured on the ith subject at time t, s_i a random effect for this patient and ε_{it} the error term. The parameters are $\mu = 2$, $\lambda = 0.5$ and $\gamma = 0$. The covariate x is normally distributed $N(0, 1)$ and $\sigma_s^2 = 1.5$, $\sigma_\varepsilon^2 = 1$ respectively. We are interested in testing the effect of the covariate. To create leverage points the response was generated with a covariate sampled from a standard normal distribution $N(0, 1)$. Then, for a proportion ϵ of the measurements, the x_{it}'s values were substituted by random numbers drawn from $N(5, 1)$, for $\epsilon = .01, .02$ and .03 adopting a similar strategy to the one used by Van Aelst and Willems (2005). For each of these situations we generated 10000 samples under the null hypothesis $H_0 : \gamma = 0$ and reported how many times H_0 was rejected for the selected amounts of contamination. The theoretical level was again set to 5%. Results are summarised in Table 3.

Table 3 about here

As for such a hypothesis we commonly use a Wald test we will only present results for this specific test. Similar results would be obtained with a LRT test. When there is no contamination in the data, all tests have a level closed to the nominal type I error $\alpha = 5\%$. In the presence of leverage points with a percentage of contamination as small as 1% the classical tests (Wald and F) are seriously biased. The position of the leverage points is of course critical for this to happen. Here we describe a situation with a low percentage of highly influential points that completely ruin the classical analysis. Intermediate and possibly more realistic situations can be thought of but this design somehow illustrates a "worst-case" scenario. Differences with the first simulation can be noticed as the robust Wald test based on the Huber estimator is now also biased. One single extreme observation in the factor space can drive the level of the test beyond any acceptable value. This does not come as a surprise as by definition this type of test is only built to deal with distributional problems in the response. On the contrary, the robust Wald test based on MM estimator remains stable. When the percentage of contamination increases the pattern remains the same, the classical and Huber tests displaying an even worse behavior. We now generate a design with a non-null slope $\gamma_0 \neq 0$ e.g. $\gamma_0 = .5$ to illustrate a common situation where the covariate is indeed needed. Assume that we are interested in testing $H_0 : \gamma = \gamma_0$ or equivalently in computing a 95% confidence interval for the slope. In such a case similar results would be observed (results not shown). Confidence intervals based on either both the classical or Huber approach would exhibit a poor coverage and would even be completely misleading. In that case only the MM-estimator can provide a proper confidence interval with the right nominal coverage.

4.5 Data Analysis

We analyse a real data set based on the first design given in the introduction. Our goal is to compare the classical and robust inference on that particular sample. The model used to analyse these data is given in (1) with $\lambda_j, j = 1, , 3$ the pantomime type (PT) and $\gamma_k, k = 1, 2$ the delay (DE). As an explanatory diagnostic tool, we provide a scatter plot of the robust Mahalanobis distances d_i obtained from an initial fit with the robust $CTBS$ estimator. This graphic is displayed in Figure 1 and reveals a few potential outliers. Note that the results obtained with the Huber or MM estimator (omitted here) are virtually the same as the $CTBS$. The horizontal line correspond to the quantile 97.5% of a χ^2_6, i.e. the asymptotic distribution of the Mahalanobis distance. An observation with a Mahalanobis distance which exceeds this cutoff value will be seen as outlier. In our example, the robust estimator detects one clear outlier (#12) which lies far away from the bulk of the data. It also detects two additional outlying observations (#19) and(#20).

We are primarily interested in testing whether the reaction time is different when there is a link between the priming and the object. We also would like to know whether the delay impacts on the reaction time, possibly with an interaction with the type of pantomime.

Figure 1 about here

We tested the significance of each factor and each interaction (i.e. 3 hypotheses) using the F-test, the classical and robust likelihood ratio tests. The results are presented in Table 4. The influence of outliers present in the data set seems to be quite substantial on the conclusions of the main effects and interaction testing. With the F-test and the classical LRT, only the effect of the pantomime type is found significant, whereas with the two robust LRT, the delay is also found significant. No evidence of an interaction effect was found by the different analyses. The robust Wald test (results not presented here) leads to the same conclusions.

Table 4 about here

4.6 Concluding remarks

In this paper we have proposed a Huber- and an MM-estimator for mixed linear models. These estimators are robust, easily computable and extend the previous work by Copt and Victoria-Feser (2005) based on S-estimators. They also allow the computation of a direct robust alternative to the F-test, namely a likelihood-ratio type test, something that was not possible with the previous proposal. Robust Wald tests based on these estimators have also been suggested as a more stable alternative to contrasts tests or test of covariate effects. We have derived the asymptotic properties of these testing procedures and studied their robustness properties. Through a real data set, we have shown that a robust analysis can provide further insight on the data. The proposed procedures have nevertheless the following limitations. Like the S-estimator these estimators and related tests have been developed for balanced designs, i.e data where the same number of measurements are recorded per observation. Future research is needed to release this condition especially in light of many applications in biostatistics where data with an unequal number of readings per subject naturally arise. The theoretical results presented here are asymptotic by nature: their validity in smaller samples needs to be examined. Alternative techniques like the fast bootstrap proposed by Salibian-Barrera and Zamar (2002) can probably be extended to these estimators and testing procedures. Another possibility would be to use more refined robust tests based on saddle-point approximations as in Robinson, Ronchetti, and Young (2003). The adaptation of this promising approach is left as future work.

4.7 Appendix

Proof of proposition 1

The proof is essentially the same as the proof of Theorem 3.2 in Lopuhaä (1992). The difference is due to the presence of covariates and the structure of the covariance matrix $\mathbf{\Sigma}$. We therefore give only a sketch of the derivations involved and focus on the Huber estimator. Similar arguments could be used for the MM-estimator.

We assume first that there is no structure on $\mathbf{\Sigma}$ but the mean of $\mathbf{y}$ is $\mu = \mathbf{x}\boldsymbol{\alpha}$ as in (5). Let $\boldsymbol{\alpha}(F)$ be the functional associated to $\widehat{\boldsymbol{\alpha}}$ the Huber estimator of $\boldsymbol{\alpha}$ and denote $F_{\boldsymbol{\alpha},\mathbf{\Sigma}}$ the model distribution, i.e. a distribution with density $f_{\boldsymbol{\alpha},\mathbf{\Sigma}}(\mathbf{y},\mathbf{x})k(x)$. The density $f_{\boldsymbol{\alpha},\mathbf{\Sigma}}$ is the p-variate normal density with mean $\boldsymbol{\mu} = \mathbf{x}\boldsymbol{\alpha}$ and variance $\mathbf{\Sigma}$ and k is the density corresponding to K. We also denote s the score function for that model, i.e. the derivative of the log-density with respect to the regression parameter $\boldsymbol{\alpha}$. At any distribution F the functional $\boldsymbol{\alpha}(.)$ is defined as the vector $\boldsymbol{\alpha}(F)$ that minimises the function $R_F(\alpha) = \int \rho_H \left(\sqrt{(\mathbf{y} - \mathbf{x}\boldsymbol{\alpha})^T \boldsymbol{\Sigma}(F)^{-1} (\mathbf{y} - \mathbf{x}\boldsymbol{\alpha})} \right) dF(\mathbf{y}, \mathbf{x})$ where $\boldsymbol{\Sigma}(.)$ is the functional corresponding to $\widehat{\boldsymbol{\Sigma}}_S$. This definition is similar to (2.4) in Lopuhaä (1992). $R_F(\alpha)$ has a derivative with respect to $\boldsymbol{\alpha}$ proportional to

$$\int u_c \left(\sqrt{(\mathbf{y} - \mathbf{x}\boldsymbol{\alpha})^T \boldsymbol{\Sigma}(F)^{-1} (\mathbf{y} - \mathbf{x}\boldsymbol{\alpha})} \right) \mathbf{x}^T \boldsymbol{\Sigma}(F)^{-1} (\mathbf{y} - \mathbf{x}\boldsymbol{\alpha}) dF(\mathbf{y}, \mathbf{x})$$

where as before $u_c(d) = \psi_c(d)/d = \min(1, c/|d|)$. Hence the functional $(\boldsymbol{\alpha}(F), \boldsymbol{\Sigma}(F))$ is a zero of $G(\boldsymbol{\theta}) = E_F\left[g(.;\theta)\right]$ where for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\Sigma})$,

$$g(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) = \Psi_H(\mathbf{y}, \mathbf{x}; \boldsymbol{\alpha}),$$

as previously defined in Section 3.1 with no specific structure on $\boldsymbol{\Sigma}$. This is again similar to equation (3.7) in Lopuhaä (1992). Then the same developments used in the proof of Theorem 3.2 can be used. Note that all the conditions mentioned there are satisfied. The function ρ is ρ_H which obviously satisfies condition (R) of Lopuhaä (1992) and u_c is of bounded variation. The functional $\boldsymbol{\alpha}(F)$ is uniquely defined and is a point of symmetry when F is the model distribution $F_{\boldsymbol{\alpha}, \boldsymbol{\Sigma}}$. $G(\theta)$ has a partial derivative with respect to $\boldsymbol{\alpha}$ that is continuous at any $\boldsymbol{\theta}$ and its partial derivative is nothing else than:

$$E_{F_{\boldsymbol{\alpha}, \boldsymbol{\Sigma}}} \left(\frac{\partial}{\partial \boldsymbol{\alpha}} \Psi_H(\mathbf{y}, \mathbf{x}; \boldsymbol{\alpha}) \right) = -E_{F_{\boldsymbol{\alpha}, \boldsymbol{\Sigma}}}(\Psi_H s^T) = -e_1 \boldsymbol{\Gamma} = -\mathbf{M}$$

Now repeat the same developments used in the proof of theorem 3.2 in that particular setting. We then obtain that $\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ is asymptotically normal with covariance matrix $\mathbf{H} = \mathbf{M}^{-1} \mathbf{Q} \mathbf{M}^{-T}$ where $\mathbf{Q} = E_{F_{\alpha, \boldsymbol{\Sigma}}}(\Psi_H \Psi_H^T) = e_2 \boldsymbol{\Gamma}$. The simpler form of $\mathbf{M}$, $\mathbf{Q}$ and finally $\mathbf{H}$ is a straightforward consequence of the structure of Ψ_H and the elliptical property of $F_{\alpha, \boldsymbol{\Sigma}}$. This result can be seen as a direct extension of Theorem 3.2 in Lopuhaä (1992). When a specific structure is imposed on the covariance matrix as in (6) the proof is unchanged provided that $\hat{\boldsymbol{\Sigma}}_S$ is a strongly consistent (robust and affine-equivariant) estimate of $\boldsymbol{\Sigma}$. In practice $\mathbf{M}$, $\mathbf{Q}$ and $\mathbf{V}$ have to be estimated to compute the eigenvalues defining the asymptotic distribution given in Proposition 2. We used the fully empirical version of each matrix as estimates, e.g. $\widehat{\mathbf{M}} = \frac{1}{n}\sum_{i=1}^{n} \Psi_H\big(y_i, x_i; \boldsymbol{\alpha}\big) s^T\big(y_i, x_i; \boldsymbol{\alpha}\big)$ as estimate of $\mathbf{M}$. The corresponding robust estimate of $\boldsymbol{\alpha}$ was plugged-in to obtain numerical values.

Proof of proposition 2

We will give here only a sketch of the proof as it follows easily from the asymptotic results obtained by Heritier and Ronchetti (1994) for the likelihood-ratio type test in a general parametric model. The only slight difference arises from the presence of the nuisance parameter $\boldsymbol{\Sigma}$ required to fully specify the model. To overcome this problem assume first that all the variances σ_j's are known and redefine estimators and related LRT tests accordingly. Then, the parameter of interest is only $\boldsymbol{\alpha}$ and Proposition 3 Part a. in Heritier and Ronchetti (1994) applies directly yielding the proposed weighted sum of q independent $\chi^2(1)$ distributions as asymptotic distribution of nS_n^2 under the null hypothesis. Now, just substitute (6) by $\widehat{\boldsymbol{\Sigma}}_S$. As it is a strongly consistent estimate of (6) that is furthermore estimated independently of $\widehat{\boldsymbol{\alpha}}$ asymptotically, the asymptotic distribution remains unchanged, which completes the proof.

References

Cantoni, E. and E. Ronchetti (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association 96*(455), 1022–1030.

Copt, S. and M.-P. Victoria-Feser (2005). High breakdown inference for mixed linear models. *To appear in Journal of the American Statistical Association.*

Davies, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics 15*, 1269–1292.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust statistics: The approach based on influence functions.* John Wiley and Sons.

Heritier, S. and E. Ronchetti (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association 89*, 897–904.

Huggins, R. M. and R. G. Staudte (1994). Variance components models for dependent cell populations. *Journal of the American Statistical Association 89*, 19–29.

Lopuhaä, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics 17*, 1662–1683.

Lopuhaä, H. P. (1992). High efficient estimators of multivariate location with high breakdown point. *The Annals of Statistics 20*, 398–413.

Maronna, R. A. and R. H. Zamar (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics 44*(4), 307–317.

Moy, G. and P. Mounoud (2003). Object recognition in young adult : Is priming with pantomimes possible ? In *Catalogue des abstracts : 8ème congrès de la société suisse de psychologie (SSP)*, Bern, Switzerland.

Patterson, H. D. and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58*, 545–554.

Pinheiro, J. C., C. Liu, and Y. N. Wu (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics 10*(2), 249–276.

Potthoff, R. F. and S. N. Roy (1964). A generalized multivariate analysis of variance model useful especially for growth curve problem. *Biometrika 51*, 313–326.

Richardson, A. M. (1997). Bounded influence estimation in the mixed linear model. *Journal of the American Statistical Association 92*, 154–161.

Richardson, A. M. and A. H. Welsh (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics 51*, 1429–1439.

Robinson, J., E. Ronchetti, and G. A. Young (2003). Saddlepoint approximations and tests based on multivariate m-estimates. *The Annals of statistics 24*, 1154–1169.

Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *The Annals of Statistics 24*, 1327–1345.

Ronchetti, E. (1982). Robust alternatives to the F-test for the linear model (STMA V24 1026). In W. Grossmann, G. C. Pflug, and W. Wertz (Eds.), *Probability and Statistical Inference*, pp. 329–342. D. Reidel Publishing Co.

Ronchetti, E. and F. Trojani (2001). Robust inference with GMM estimators. *Journal of Econometrics 101*(1), 37–69.

Rousseeuw, P. and V. Yohai (1984). Robust regression by means of S-estimators. In W. H. J. Franke and D. Martin (Eds.), *Robust and Nonlinear Time Series Analysis*, New-York, pp. 256–272. Springer-Verlag.

Salibian-Barrera, M. and R. H. Zamar (2002). Bootrapping robust estimates of regression. *The Annals of Statistics 30*(2), 556–582.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance components.* John Wiley and Sons.

Stahel, W. A. and A. Welsh (1997). Approaches to robust estimation in the simplest variance components model. *Journal of Statistical Planning and Inference 57*, 295–319.

Van Aelst, S. and G. Willems (2005). Multivariate regression s-estimators for robust estimation and inference. *submitted*.

Welsh, A. and A. Richardson (1997). Approaches to the robust estimation of mixed models. In G. Maddala and C. Rao (Eds.), *Handbook of Statistics Vol. 15*, pp. 343–385. Holland: Elsevier Science B.V.

Woodruff, D. L. and D. M. Rocke (1994). Computable robust estimation of multivariate location and shape in high dimension using compound estimators. *Journal of the American Statistical Association 89*, 888–896.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics 15*, 642–656.

NHMRC Clinical Trial Centre, University of Sydney

E-mail: scopt@ctc.usyd.edu.au

NHMRC Clinical Trial Centre, University of Sydney

E-mail: stephane.heritier@ctc.usyd.edu.au

constant c_0 for $\varepsilon^* = 50\%$										
p	1	2	3	4	5	6	7	8	9	10
c_0	1.56	2.66	3.45	4.09	4.65	5.14	5.59	6.01	6.40	6.77
constant c_1 for 95% efficiency										
c_1	4.68	5.12	5.51	5.82	6.10	6.37	6.60	6.83	7.04	7.25

Table 1: Values for c_0 and c_1 for Tukey's biweight.

	$\epsilon = 0\%$	$\epsilon = 2\%$	$\epsilon = 5\%$	$\epsilon = 8\%$
Classical *LRT*	5.32%	7.45%	21.12%	48.01%
Robust *LRT* (Huber)	5.29%	4.98%	5.33%	5.23%
Robust *LRT* (MM)	5.18%	5.21%	5.13%	4.98%
F-test	5.04%	7.01%	19.78%	47.92%

Table 2: Proportion of times the null hypothesis is rejected.

	$\epsilon = 0\%$	$\epsilon = 1\%$	$\epsilon = 3\%$
Classical Wald	5.12%	27.45%	82.32%
Robust Wald (Huber)	5.79%	12.12%	52.33%
Robust Wald (MM)	5.15%	4.89%	5.21%
F-test	5.17%	13.65%	46.58%

Table 3: Proportion of times the null hypothesis is rejected.

	Classical LRT	Robust LRT (M)	Robust LRT (MM)	F-test
DE	0.1905	**0.0041**	**0.0043**	0.1883
PT	**0.0014**	**0.0021**	**0.0028**	**0.0018**
DE:PT	0.8997	0.9251	0.9139	0.9232

Table 4: p-values for the different test statistics.

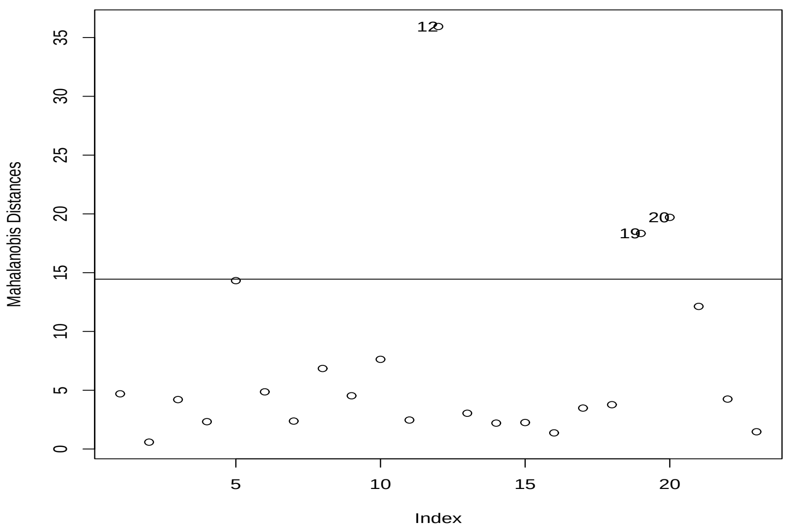

Figure 1: Robust Mahalanobis Distances.

Publications récentes du Département d'économétrie

pouvant être obtenues à l'adresse suivante :

Université de Genève
UNI MAIL
A l'att. de Mme Caroline Schneeberger
Département d'économétrie
40, Bd du Pont-d'Arve
CH - 1211 Genève 4
ou sur
INTERNET : http//www.unige.ch/ses/metri/cahiers

2005.04 KRISHNAKUMAR Jaya and David NETO, Testing Unit Root in Threshold Cointegration, Novembre 2005, 25 pages.

2005.03 VAN BAALEN Brigitte, Tobias MÜLLER, Social Welfare effects of tax-benefit reform under endogenous participation and unemployment, Février 2005, 42 pages.

2005.02 CZELLAR Véronika, G. Andrew KAROLYI, Elvezio RONCHETTI, Indirect Robust Estimation of the Short-term Interest Rate Process, Mars 2005, 29 pages.

2005.01 E. CANTONI, C. FIELD, J. MILLS FLEMMING, E. RONCHETTI, Longitudinal Variable Selection by Cross-Validation in the Case of Many Covariates, Février 2005, 17 pages.

2004.15 KRISHNAKUMAR Jaya, Marc-Jean MARTIN, Nils SOGUEL, Application of Granger Causality Tests to Revenue and Expenditure of Swiss Cantons, Décembre 2004, 27 pages.

2004.14 KRISHNAKUMAR Jaya, Gabriela FLORES, Sudip Ranjan BASU, Spatial Distribution of Welfare Across States and Different Socio-Economic Groups in Rural and Urban India, Mai 2004, 66 pages.

2004.13 KRISHNAKUMAR Jaya, Gabriela FLORES, Sudip Ranjan BASU, Demand 2004System Estimations and Welfare Comparisons : Application to Indian Household Data, Mai 2004, 70 pages.

2004.12 KRISHNAKUMAR Jaya, Going beyond functionings to capabilities: an econometric model to explain and estimate capabilities, Août 2004, 29 pages.

2004.11 MÜLLER Tobias, RAMSES Abul Naga, KOLODZIEJCZYK Christophe, The Redistributive Impact of Altrnative Income Maintenance Schemes : A Microsimulation Study using Swiss Data, Août 2004, 40 pages.

2004.10 GHIGLINO Christian, DUC François, Expectations in an OG Economy, Août 2004, 21 pages.