Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Composite Robust Estimators for Linear Mixed Models

Claudio Agostinelli & Víctor J. Yohai

To cite this article: Claudio Agostinelli & Víctor J. Yohai (2015): Composite Robust Estimators for Linear Mixed Models, Journal of the American Statistical Association, DOI: 10.1080/01621459.2015.1115358

To link to this article: http://dx.doi.org/10.1080/01621459.2015.1115358

Composite Robust Estimators for Linear Mixed Models

Claudio Agostinelli and Víctor J. Yohai

Claudio Agostinelli is Associate Professor, Department of Mathematics, University of Trento, Trento, Italy (E–mail: claudio.agostinelli@unitn.it) and Víctor J. Yohai is Emeritus Professor, Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1, 1428 Buenos Aires, Argentina (E–mail: victoryohai@gmail.com).

This research was partially supported by the Italian-Argentinian project "Metodi robusti per la previsione del costo e della durata della degenza ospedaliera" funded inside the collaboration program of MINCYT-MAE AR14MO6.

Víctor Yohai research was also partially supported by Grants 20020130100279 from Universidad of Buenos Aires, PIP 112-2008-01-00216 and 112-2011-01-00339 from CONICET and PICT 2011-0397 from ANPCYT.

All statistical analysis were performed on SCSCF (www.dais.unive.it/scscf), a multiprocessor cluster system owned by Ca'Foscari University of Venice running under GNU/Linux.

This article has supplementary material online.

Abstract

The Classical Tukey-Huber Contamination Model (CCM) is a commonly adopted framework to describe the mechanism of outliers generation in robust statistics. Given a data set with n observations and p variables, under the CCM, an outlier is a unit, even if only one or a few values are corrupted. Classical robust procedures were designed to cope with this type of outliers. Recently, a new mechanism of outlier generation was introduced, namely the Independent Contamination Model (ICM), where the occurrences that each cell of the data matrix is an outlier are independent events and have the same probability. ICM poses new challenges to robust statistics since the percentage of contaminated rows dramatically increase with p, often reaching more than 50% whereas classical affine equivariant robust procedures have a breakdown point of 50% at most. For ICM we propose a new type of robust methods namely composite robust procedures which are inspired by the idea of composite likelihood, where low dimension likelihood, very often the likelihood of pairs, are aggregated in order to obtain a tractable approximation of the full likelihood. Our composite robust procedures are build on pairs of observations in order to gain robustness in the ICM. We propose composite τ-estimators for linear mixed models. Composite τ-estimators are proved to have a high breakdown point both in the CCM and ICM. A Monte Carlo study shows that while classical S-estimators can only cope with outliers generate by the CCM, the estimators proposed here are resistant to both CCM and ICM outliers.

Key words: Composite τ-estimators; Independent Contamination Model; Tukey-Huber Contamination Model; Robust estimation.

1 Introduction

The purpose of this paper is to find robust procedures for mixed linear models. This class of models includes, among others, ANOVA models with repeated measures, models with random

ACCEPTED MANUSCRIPT

nested design and models for studying longitudinal data. These models are generally based on the assumption that the data follow a normal distribution and therefore the parameters are estimated using the maximum likelihood principle. See for example, Searle et al. (1992). As it is well known, in general, the estimator obtained by maximum likelihood under the assumption that the data have a normal distribution is very sensitive to the presence of a small fraction of outliers in the sample. More than that, just one outlier may have an unbounded effect on this estimator. There are many robust estimators that have been proposed to avoid a large outlier influence. A large list of references of these proposals is available in Heritier et al. (2009). Copt and Victoria-Feser (2006) introduce a very interesting robust S-estimator for mixed linear models based on M-scales which has breakdown point equals to 0.5 under the classical contamination model. We can also mention Fellner (1986), Richardson and Welsh (1995), Stahel and Welsh (1997), Gill (2000), Jiang and Zhang (2001), Sinha (2004), Copt and Heritier (2006), Jacqmin-Gadda et al. (2007), Lachosa et al. (2009), Chervoneva and Vishnyakov (2011) and Koller (2013) which studied an SMDM-estimator. The procedure proposed in the last paper is implemented in the `R` package `robustlmm` (Koller, 2015).

However all these procedures aim at coping with outliers generated under the Classical (Tukey-Huber) Contamination Model (CCM), where some percentage of the units that compose the sample are replaced by outliers. Alqallaf et al. (2009) introduced another type of contamination model (called Independent Contamination Model, ICM) that may occur in multivariate data. Instead of contaminating a percentage of the units that compose the sample, the different cells of each unit may be independently contaminated. In this case, if the dimension of each unit is large, even a small fraction of cell contamination may lead to a large fraction of units with at least one contaminated cell. This type of contamination specially occurs when the different variables that compose each unit are measured from independent laboratories. Alqallaf et al. (2009) showed that for this type of contamination the breakdown point of affine equivariant procedures for multivariate location and covariance matrix tends to zero when the number of variables increases and therefore their degree of robustness is not satisfactory. A similar phenomenon occurs when dealing with mixed linear models. In particular the S-estimator procedure introduced in Copt and Victoria-Feser (2006) loses robustness for high dimensional data with independent contamination.

In this paper we propose a new class of robust estimators for linear mixed models. These estimators are based on a principle similar to the one used in the composite likelihood estimators proposed by Lindsay (1988). If a vector $\boldsymbol{y}$ of dimension p is observed, the composite likelihood estimators are based on the likelihood of all the subvectors of a dimension $p^* < p$. The estimators that we propose here are based on τ-scales of the Mahalanobis distances of two dimensional subvectors of $\boldsymbol{y}$. The τ-scale estimators were introduced by Yohai and Zamar (1988) and provides scales estimators which are simultaneously highly robust and highly efficient. Here we show that these estimators have a robust behavior for both contamination models: the classical contamination model and the independent contamination one. In particular, we will show that the breakdown point for the classical contamination model can be made close to 0.5, while for the independent contamination model can be made close to 0.25.

In Section 2 the linear mixed model is presented. We also describe two outlier contamination models: the classical and the independent contamination models. Section 3 describe the M- and τ-scales, briefly reviews the S-estimator for linear mixed models introduced in Copt and Victoria-Feser (2006) and define the composite τ-estimator for these models. Section 4 discusses the breakdown properties of composite τ-estimators under both contamination models. Section 5 states the continuity of the estimating functional associated to a composite τ-estimator. Section 6 states the consistency and asymptotic normality of the composite τ-estimator. Section 7 illustrates with real data set the advantages of the proposed estimator and in Section 8 we perform a Monte Carlo simulation that confirms that the proposed procedure has a robust behavior under both contamination models. Section 9 provides some concluding remarks. A Supplementary Material is available with the following sections. Section SM–1 provides details on the derivation of the estimating equations while Section SM–2 discusses computational aspects and algorithms. Section SM–3 contains the

ACCEPTED MANUSCRIPT

proofs of the statements reported in Section 4. In Section SM–4 we prove the Fisher-consistency of the composite τ-estimators and in Section SM–5 the continuity and consistency of these estimators. Section SM–6 provides complimentary results of the Monte Carlo experiments reported in Section 8 and Section SM–7 contains the `R` code for the example presented in Section 7. Finally, the `R` package `robustvarComp` implementing the proposed procedures is available in the Comprehensive R Archive Network.

2 Linear mixed models

Denote by $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the multivariate normal distribution of dimension p with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In case of fixed covariables it is assumed that n independent p-dimensional random vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ in $\mathbb{R}^p$ are observed and $\boldsymbol{y}_i$, $1 \le i \le n$, has distribution $N_p(\boldsymbol{\mu}_i(\boldsymbol{\beta}_0), \boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0))$, where

$$\begin{aligned} \boldsymbol{\mu}_i(\boldsymbol{\beta}) &= (\mu_{i1}(\boldsymbol{\beta}), \ldots, \mu_{ip}(\boldsymbol{\beta}))^\top \\ &= \boldsymbol{x}_i\boldsymbol{\beta}, \qquad 1 \le i \le n, \end{aligned} \tag{1}$$

$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are fixed $p \times k$ matrices and $\boldsymbol{\beta} \in \mathbb{R}^k$ is an unknown k-vector parameter. Moreover,

$$\boldsymbol{\Sigma}(\eta, \boldsymbol{\gamma}) = \eta(\boldsymbol{V}_0 + \sum_{j=1}^{J} \gamma_j \boldsymbol{V}_j), \tag{2}$$

where $\boldsymbol{V}_j$, $1 \le j \le J$ are $p \times p$ known matrices, $\boldsymbol{V}_0$ is the $p \times p$ identity, $\eta > 0$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)^\top \in \Gamma$ are unknown parameters, where

$$\Gamma = \{\boldsymbol{\gamma} \in \mathbb{R}^J : \ \boldsymbol{\Sigma}(1, \boldsymbol{\gamma}) \quad \text{is positive definite}\}.$$

In the case of random covariables, that is, when $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are i.i.d. random matrices, it is assumed that $\boldsymbol{y}_i | \boldsymbol{x}_i \sim N_p\left(\boldsymbol{\mu}_i(\boldsymbol{\beta}_0), \boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0)\right)$. This is equivalent to

$$\boldsymbol{y}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}_0) + \boldsymbol{u}_i, \tag{3}$$

ACCEPTED MANUSCRIPT

where $\boldsymbol{u}_i$ is independent of $\boldsymbol{x}_i$ with distribution $N_p(\mathbf{0}, \boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0))$. However, in Section 6, where we study the asymptotic properties of the proposed estimators, we weaken the assumption. In fact, we only require that $\boldsymbol{u}_i$ are independent of $\boldsymbol{x}_i$ and have elliptical distribution with center $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0)$.

This setup covers many linear mixed models, for instance those of the form

$$\boldsymbol{y}_i = \boldsymbol{x}_i \boldsymbol{\beta}_0 + \sum_{j=1}^{J} \boldsymbol{z}_j \boldsymbol{\zeta}_{ij} + \boldsymbol{\varepsilon}_i, \qquad 1 \leq i \leq n, \tag{4}$$

where the $\boldsymbol{x}_i$s are as before, $\boldsymbol{z}_j$, $1 \leq j \leq J$, are $p \times q_j$ known design matrices for the random effects, $\boldsymbol{\zeta}_{ij}$ are independent q_j-dimensional vectors with distribution $N_{q_j}(0, \sigma_{0j}^2 \boldsymbol{I}_{q_j})$, where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix and $\boldsymbol{\varepsilon}_i$ $(1 \leq i \leq n)$ are p-dimensional error vectors with distribution $N(0, \sigma_0^2 \boldsymbol{I}_p)$. Then, in this case we have $\eta_0 = \sigma_0^2, \boldsymbol{\gamma}_0 = (\gamma_{01}, \ldots, \gamma_{0J})^\top$ with $\gamma_{0j} = \sigma_{0j}^2/\sigma_0^2 > 0$, $\boldsymbol{V}_j = \boldsymbol{z}_j \boldsymbol{z}_j^\top$, $1 \leq j \leq J$.

2.1 Outlier contamination models

We are going to introduce some notation. Let $\boldsymbol{T}$ be a data set of size n corresponding to model (1)-(2), $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n)$, where $\boldsymbol{t}_i = (\boldsymbol{y}_i, \boldsymbol{x}_i) = (\boldsymbol{t}_{i1}, \ldots, \boldsymbol{t}_{ip})$, $\boldsymbol{y}_i \in \mathbb{R}^p$, $\boldsymbol{x}_i \in \mathbb{R}^{p \times k}$, $\boldsymbol{t}_{ij} = (y_{ij}, x_{ij1}, \ldots, x_{ijk})$, $1 \leq j \leq p$ and x_{ijh} is the value in the j-th row and h-th column of the matrix $\boldsymbol{x}_i$.

The classical contamination model (CCM) assume that the probability that $\boldsymbol{t}_i$, $1 \leq i \leq n$, is replaced by an outlier is a given number ε and these n events are independent.

Alqallaf et al. (2009) consider a different contamination model for multivariate data: the independent contamination model (ICM). This definition can be adapted to the mixed linear model as follows. The probability that $\boldsymbol{t}_{ij}$, $1 \leq i \leq n, 1 \leq j \leq p$ is replaced by an outlier is a given number ε and these $n \times p$ events are independent. Therefore the probability that at least one component $\boldsymbol{t}_{ij}$, $1 \leq j \leq p$ of $\boldsymbol{t}_i$ is an outlier is $1 - (1 - \varepsilon)^p$, and this number is close to one when p is large even if ε is small. For this reason estimators that have breakdown point 0.5 under the CCM may have breakdown tending to zero under the ICM. Alqallaf et al. (2009) show that this happens with the most popular high breakdown point equivariant estimators of multivariate location, e.g., S-estimators (Davies, 1987), Minimum Volume Ellipsoid (Rousseeuw, 1985), Minimum Covariance Determinant (Rousseeuw, 1985) or the Donoho-Stahel estimators (Donoho, 1982; Stahel, 1981).

3 Estimators for linear mixed models based on robust scales

In this Section we define the composite τ-estimator. At this aim we first describe the M- and τ-scales, and briefly review the S-estimator for linear mixed models introduced in Copt and Victoria-Feser (2006).

3.1 M- and τ-Scales

In general, a *scale* s of a sample $\boldsymbol{u} = (u_1, \ldots, u_n)$ is a measure of the absolute largeness of these observations. A general class of robust scales introduced by Huber are the M-scales. The M-scales are defined as follows: Let $\rho : \mathbb{R} \to \mathbb{R}^+$ be a function satisfying the following properties:

A1 $\rho(0) = 0$, **A2** $0 \leq u \leq v$ implies $\rho(u) \leq \rho(v)$, **A3** ρ is continuous, **A4** $\sup_v \rho(v) = 1$ and **A5** if $\rho(u) < 1$ and $0 \leq u < v$, then $\rho(u) < \rho(v)$.

Then, an *M-scale* $s(\boldsymbol{u})$ based on ρ is defined by the value s satisfying

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{u_i}{s}\right) = b, \tag{5}$$

where $0 < b < 1$.

We now define the family of τ-scales introduced by Yohai and Zamar (1988). A *τ-scale* is based on two functions ρ_1 and ρ_2 satisfying conditions A1-A5 and

A6 ρ_2 is continuously differentiable and if $\psi_2(v) = \rho_2'(v)$, then $2\rho_2(v) - \psi_2(v)v > 0$.

Let $s(\boldsymbol{u})$ be the M-scale defined by (5) with ρ_1 in place of ρ, then the τ-scale is defined by

$$\tau^2(\boldsymbol{u}) = s^2(\boldsymbol{u})\frac{1}{n}\sum_{j=1}^{n}\rho_2\left(\frac{u_j}{s(\boldsymbol{u})}\right). \tag{6}$$

Condition A6 implies that if $B(s) = s^2 E(\rho/s)$, then $B'(s) = s(2\rho(u/s) - (u/s)\rho(u/s)) > 0$, that is, $B(s)$ is increasing on s. The advantage of the τ-scales over the M-scales is that they make possible to define estimators that are simultaneously highly robust and highly efficient. See for example Yohai and Zamar (1988) for linear regression and Lopuhaä (1991) for multivariate location and scatter.

3.2 S-estimators

In this subsection we review the class of S-estimators for linear mixed models introduced by Copt and Victoria-Feser (2006) and define the composite τ-estimators for the same class of models.

Given p dimensional column vectors $\boldsymbol{y}$ and $\boldsymbol{\mu}$ and a $p\times p$ matrix $\boldsymbol{\Sigma}$ the square of the Mahalanobis distance is defined by.

$$m(\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}).$$

Given a squared matrix $\boldsymbol{A}$ we denote by $\boldsymbol{A}^* = \boldsymbol{A}/\det(\boldsymbol{A})^{1/p}$ where $\det(\boldsymbol{A})$ is the determinant of the matrix $\boldsymbol{A}$. Note that $\boldsymbol{\Sigma}^*(\eta, \boldsymbol{\gamma})$ depends only on $\boldsymbol{\gamma}$ and then will be denoted by $\boldsymbol{\Sigma}^*(\boldsymbol{\gamma})$.

Let ρ_1 be a function satisfying A1-A5. The S-estimator proposed by Copt and Victoria-Feser (2006) can be defined by

$$(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}) = \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} s(m(\boldsymbol{y}_1, \boldsymbol{x}_1\boldsymbol{\beta}, \boldsymbol{\Sigma}^*(\boldsymbol{\gamma}))^{1/2}, \ldots, m(\boldsymbol{y}_n, \boldsymbol{x}_n\boldsymbol{\beta}, \boldsymbol{\Sigma}^*(\boldsymbol{\gamma}))^{1/2}), \tag{7}$$

$$\widehat{\eta} = s(m(\boldsymbol{y}_1, \boldsymbol{x}_1\widehat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}(1, \widehat{\boldsymbol{\gamma}}))^{1/2}, \ldots, m_n(\boldsymbol{y}_n, \boldsymbol{x}_n\widehat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}(1, \widehat{\boldsymbol{\gamma}}))^{1/2})^2/s_0^2, \tag{8}$$

where s is the M-scale corresponding to ρ_1 defined by (5) and s_0 is defined by $E\left[\rho_1\left(\sqrt{v}/s_0\right)\right] = b$, where v has chi-square distribution with p degrees of freedom. We will call this estimator Copt and Victoria-Feser S-estimator (CVFS-estimator).

These estimators can be thought of as an extension of the S-estimators for multidimensional location and scatter matrix proposed by Davies (1987). Copt and Victoria-Feser (2006) choose as ρ_1 the function proposed in Rocke (1996) for S-estimators of scatter matrix and multivariate location. This function, which depends on the number of variables, improves the robustness of S-estimators for high dimensional data under the CCM.

Copt and Victoria-Feser (2006) show that under the CCM the asymptotic breakdown point of the estimator defined by (7) and (8) is $\varepsilon^* = \min(b, 1-b)$. Therefore if $b = 0.5$, we get $\varepsilon^* = 0.5$. However, as the S-estimators for multidimensional location and scatter, the breakdown point of these estimator under the ICM tends to 0 when $p \to \infty$.

3.3 Composite τ-estimators

We introduce now a new type of estimators which are robust under the ICM: the composite τ-estimators.

Given a vector $\boldsymbol{a} = (a_1, \ldots, a_p)^\top$, a $p \times p$ matrix $\boldsymbol{A}$ and a couple (j, l) of indices $(1 \leq j < l \leq p)$ we denote $\boldsymbol{a}^{jl} = (a_j, a_l)^\top$ and $\boldsymbol{A}_{jl}$ the submatrix

$$\boldsymbol{A}_{jl} = \begin{pmatrix} a_{jj} & a_{jl} \\ a_{lj} & a_{ll} \end{pmatrix}.$$

In a similar way, given a $p \times k$ matrix $\boldsymbol{x}$ we denote by $\boldsymbol{x}^{jl}$ the matrix of dimension $2 \times k$ built by using rows j and l of $\boldsymbol{x}$ and by $A^*_{jl} = A_{jl} / \det(A_{jl})^{1/2}$. Note that $\det(A^*_{jl}) = 1$.

We define pairwise squared Mahalanobis distances by

$$m_i^{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = m(\boldsymbol{y}_i^{jl}, \boldsymbol{\mu}_i^{jl}(\boldsymbol{\beta}), \boldsymbol{\Sigma}^*_{jl}(\boldsymbol{\gamma})).$$

Given ρ_1 an ρ_2 satisfying A1-A6, we define pairwise M- and τ-scales $s_{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and $\tau_{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ by

$$\frac{1}{n} \sum_{j=1}^{n} \rho_1 \left(\frac{m_j^{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma})^{1/2}}{s_{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma})} \right) = b \tag{9}$$

and

$$\tau^2_{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = s^2_{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \frac{1}{n} \sum_{j=1}^{n} \rho_2 \left(\frac{m_j^{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma})^{1/2}}{s_{jl}(\boldsymbol{\beta}, \boldsymbol{\gamma})} \right). \tag{10}$$

Put

$$L(\boldsymbol{\beta},\boldsymbol{\gamma}) = \sum_{j=1}^{p-1}\sum_{l=j+1}^{p}\tau_{jl}^2(\boldsymbol{\beta},\boldsymbol{\gamma}), \tag{11}$$

then we define the composite τ-estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ by

$$(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\gamma}}) = \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} L(\boldsymbol{\beta},\boldsymbol{\gamma}), \tag{12}$$

and the estimator $\widehat{\eta}$ of η_0 by solving

$$\frac{2}{p(p-1)n}\sum_{i=1}^{n}\sum_{j=1}^{p-1}\sum_{l=j+1}^{p}\rho_1\left(\frac{(\boldsymbol{y}_i^{jl}-\boldsymbol{x}_i^{jl}\widehat{\boldsymbol{\beta}})^\top\boldsymbol{\Sigma}_{jl}(\widehat{\eta},\widehat{\boldsymbol{\gamma}})^{-1}(\boldsymbol{y}_i^{jl}-\boldsymbol{x}_i^{jl}\widehat{\boldsymbol{\beta}})^{1/2}}{s_0}\right) = b, \tag{13}$$

where s_0 is defined by

$$E\left[\rho_1\left(\frac{\sqrt{v}}{s_0}\right)\right] = b, \tag{14}$$

and v has chi-square distribution with 2 degrees of freedom. As a particular case when $\rho_2 = \rho_1$ we have the class of composite S-estimators. We will not discuss further this special case.

In the example of Section 7 and in the Monte Carlo study of Section 8 we considered ρ_i, $i = 1, 2$ in the following family of functions ρ_c^* introduced by Muler and Yohai (2002)

$$\rho_c^*(v) = \begin{cases} \frac{v^2}{2ac^2} & v \le 2 \\ \frac{1}{a}\left(\frac{a_4}{8}\frac{v^8}{c^8} + \frac{a_3}{6}\frac{v^6}{c^6} + \frac{a_2}{4}\frac{v^4}{c^4} + \frac{a_1}{2}\frac{v^2}{c^2} + a_0\right) & 2 < v \le 3 \\ 1 & v > 3 \end{cases} \tag{15}$$

where $a_0 = 1.792$, $a_1 = -1.944$, $a_2 = 1.728$, $a_3 = -0.312$, $a_4 = 0.016$ and $a = 3.250$. The functions in this family are characterized by shapes close to those in the optimal family for regression M-estimators, obtained by Yohai and Zamar (1997). However, they are easier to compute. Notice that for any $\lambda > 0$, the τ-scale obtained with $\rho_1 = \rho_{\lambda c_1}^*$ and $\rho_2 = \rho_{\lambda c_2}^*$ is equal to the τ-scale corresponding to $\rho_1 = \rho_{c_1}^*$ and $\rho_2 = \rho_{c_2}^*$ divided by λ^2. Hence without loss of generality we can consider $c_1 = 1$. We found that by taking $c_2 = 1.64$ we obtain a good trade-off between robustness and efficiency, and this is the value of c_2 that we recommend to use.

Note that the composite τ-estimators is quite similar to the S-estimator proposed by Copt and Victoria-Feser (2006). The main differences are:

(i) We use a τ-scale instead of an M-scale to gain efficiency under the nominal model,

(ii) We work with partial Mahalanobis distances of all pairs of components of the residual vectors instead of using the Mahalanobis distances of complete residual vectors. The reason for this is to increase the breakdown point of the estimators under the ICM. Since the breakdown point decreases with the dimension, we use subvectors with the smallest possible dimension. Note that to estimate the covariance matrix $\boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0)$ it is necessary to work with subvectors of dimension at least two. We might have worked in higher dimension, but this would have decreased the breakdown point. Besides, in this case we would have a larger number of index combinations, increasing the computational complexity of the estimator. For these reasons, it seems preferable to consider pairs of variables.

(iii) Finally we use the ρ-function given in (15) instead of the Rockes' ρ. The Rocke ρ-function was designed to increase the robustness under CCM for high dimension. However the covariance matrix estimator based on the Rocke ρ-function is inefficient in dimension two. Therefore, since we worked with subvectors of dimension two, the ρ-function given in (15) is preferable.

It is easy to show that the composite τ-estimators are equivariant for regression transformations of the form $\boldsymbol{y}_i^* = \boldsymbol{y}_i + \boldsymbol{x}_i\boldsymbol{\delta}$ where $\boldsymbol{\delta}$ is a $k \times 1$ vector, or scale transformations of the form $\boldsymbol{y}_i^* = \zeta y_i$, where ζ is a scalar.

Computational aspects and algorithms are discussed in Section SM–2 of the Supplementary Material.

4 Breakdown point

Donoho and Huber (1983) introduced the concept of a finite sample breakdown point (FSBDP). For our case, let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\upsilon}} = (\widehat{\eta}, \widehat{\eta\boldsymbol{\gamma}})$ be estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\upsilon} = (\eta, \eta\boldsymbol{\gamma})$. Informally speaking, the

FSBDP of $\widehat{\boldsymbol{\beta}}$ is the smallest fraction of outliers that makes the estimator unbounded or becoming arbitrarily close to the border of the parameter space.

Let $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n)$ be a data set of size n corresponding to model (1)-(2) $\boldsymbol{t}_i = (\boldsymbol{y}_i, \boldsymbol{x}_i) = (\boldsymbol{t}_{i1}, \ldots, \boldsymbol{t}_{ip})$, $\boldsymbol{y}_i \in \mathbb{R}^p$, $\boldsymbol{x}_i \in \mathbb{R}^{p\times k}$ and $\boldsymbol{t}_{ij} = (y_{ij}, x_{ij1}, \ldots, x_{ijk})$. Let $\mathcal{T}_m^{(\mathrm{C})}$ be the set of all the samples $\check{\boldsymbol{T}} = (\check{\boldsymbol{t}}_1, \ldots, \check{\boldsymbol{t}}_n)$ with $\check{\boldsymbol{t}}_i = (\check{\boldsymbol{t}}_{i1}, \ldots, \check{\boldsymbol{t}}_{ip})^\top$ such that $\#\{i : \check{\boldsymbol{t}}_i = \boldsymbol{t}_i\} \geq n - m$. Given estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\upsilon}}$ we let

$$B_m^{(\mathrm{C})}(\boldsymbol{T}, \widehat{\boldsymbol{\beta}}) = \sup\{\|\widehat{\boldsymbol{\beta}}(\check{\boldsymbol{T}})\|, \check{\boldsymbol{T}} \in \mathcal{T}_m^{(\mathrm{C})}\},$$

$$B_m^{+(\mathrm{C})}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}}) = \sup\{\|\widehat{\boldsymbol{\upsilon}}(\check{\boldsymbol{T}})\|, \check{\boldsymbol{T}} \in \mathcal{T}_m^{(\mathrm{C})}\},$$

$$B_m^{-(\mathrm{C})}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}}) = \inf\{\|\widehat{\boldsymbol{\upsilon}}(\check{\boldsymbol{T}})\|, \check{\boldsymbol{T}} \in \mathcal{T}_m^{(\mathrm{C})}\},$$

where $\|\boldsymbol{x}\|$ denotes norm l_2 of $\boldsymbol{x}$.

Definition 1 *The finite sample breakdown point of $\widehat{\boldsymbol{\beta}}$ for classical contamination (FSBDPCC) at the sample $\boldsymbol{T}$ is defined by $\varepsilon^{(C)}(\boldsymbol{T}, \widehat{\boldsymbol{\beta}}) = m^*/n$ where $m^* = \min\{m : B_m^{(C)}(\boldsymbol{T}, \widehat{\boldsymbol{\beta}}) = \infty\}$ and the breakdown point of $\widehat{\boldsymbol{\upsilon}}$ by $\varepsilon^{(C)}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}}) = m^*/n$ where*

$$m^* = \min\{m : \frac{1}{B_m^{-(C)}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}})} + B_m^{+(C)}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}}) = \infty\}.$$

Let $\mathcal{T}_m^{(\mathrm{I})\cdot}$ be the set of all the samples $\check{\boldsymbol{T}} = (\check{\boldsymbol{t}}_1, \ldots, \check{\boldsymbol{t}}_n)$ such that $\#\{i : \check{\boldsymbol{t}}_{ij} = \boldsymbol{t}_{ij}\} \geq n - m$ for each j, $1 \leq j \leq p$. Given estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\upsilon}}$ we let

$$B_m^{(\mathrm{I})\cdot}(\boldsymbol{T}, \widehat{\boldsymbol{\beta}}) = \sup\{\|\widehat{\boldsymbol{\beta}}(\check{\boldsymbol{T}})\|, \check{\boldsymbol{T}} \in \mathcal{T}_m^{(\mathrm{I})}\},$$

$$B_m^{+(\mathrm{I})}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}}) = \sup\{\|\widehat{\boldsymbol{\upsilon}}(\check{\boldsymbol{T}})\|, \check{\boldsymbol{T}} \in \mathcal{T}_m^{(\mathrm{I})}\},$$

$$B_m^{-(\mathrm{I})}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}}) = \inf\{\|\widehat{\boldsymbol{\upsilon}}(\check{\boldsymbol{T}})\|, \check{\boldsymbol{T}} \in \mathcal{T}_m^{(\mathrm{I})}\}.$$

Definition 2 *The finite sample breakdown point for $\widehat{\boldsymbol{\beta}}$ under independent contamination (FSBD-PIC) at the sample $\boldsymbol{T}$ is defined by $\varepsilon^{(I)}(\boldsymbol{T}, \widehat{\boldsymbol{\beta}}) = m^*/n$ where $m^* = \min\{m : B_m^{(I)}(\boldsymbol{T}, \widehat{\boldsymbol{\beta}}) = \infty\}$ and the* *breakdown point of* $\widehat{\boldsymbol{\upsilon}}$ *by* $\varepsilon^{(I)}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}}) = m^*/n$ *where*

$$m^* = \min\{m : \frac{1}{B_m^{-(I)}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}})} + B_m^{+(I)}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}}) = \infty\}.$$

The following theorems, whose proofs can be found in Section SM–3 of the Supplementary Material, give lower bound for the breakdown points of composite τ-estimators under both the classical and the independent contamination models. Before to state these Theorems we need the following notation. Given a sample $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n)$ we define

$$h_{jl}(\boldsymbol{T}) = \max_{\|\boldsymbol{b}\|>0} \#\{i : \boldsymbol{x}_i^{jl}\boldsymbol{b} = \boldsymbol{0}\}, \tag{16}$$

$$h(\boldsymbol{T}) = \max_{jl} h_{jl}(\boldsymbol{T}), \tag{17}$$

$$h^*_{jl}(\boldsymbol{T}) = \max_{\|\boldsymbol{e}\|>0, \boldsymbol{b}} \#\{i : \boldsymbol{e}^\top(\boldsymbol{y}_i^{jl} - \boldsymbol{x}_i^{jl}\boldsymbol{b}) = \boldsymbol{0}\}, \tag{18}$$

$$h^*(\boldsymbol{T}) = \max_{jl} h^*_{jl}(\boldsymbol{T}), \tag{19}$$

$$d(\boldsymbol{T}) = h(\boldsymbol{T}) + h^*(\boldsymbol{T}). \tag{20}$$

Theorem 1 *Let* $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n)$, $\boldsymbol{t}_i = (\boldsymbol{y}_i, \boldsymbol{x}_i)$, $d(\boldsymbol{T})$ *as defined in (20). Assume that A1-A6 holds and let* $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\upsilon}})$ *be the composite* τ*-estimator for the model given by (1) and (2). Then a lower bound for* $\varepsilon^{(C)}(\boldsymbol{T}, \widehat{\boldsymbol{\beta}})$ *and for* $\varepsilon^{(C)}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}})$ *is given by* $\min((1-b) - (d(\boldsymbol{T})/n), b)$.

Note that taking $b = 0.5$, this lower bound is close to 0.5 for large n independently of p.

Theorem 2 *Let* $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n)$, $\boldsymbol{t}_i = (\boldsymbol{y}_i, \boldsymbol{x}_i)$, $d(\boldsymbol{T})$ *as defined in (20). Assume that A1-A6 holds and let* $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\upsilon}})$ *be the composite* τ*-estimator for the model given by (1) and (2). Then a lower bound for* $\varepsilon^{(I)}(\boldsymbol{T}, \widehat{\boldsymbol{\beta}})$ *and for* $\varepsilon^{(I)}(\boldsymbol{T}, \widehat{\boldsymbol{\upsilon}})$ *is given by* $0.5\min((1-b) - (d(\boldsymbol{T})/n), b)$.

In this case by taking $b = 0.5$, this lower bound is close to 0.25 for large n independently of p.

ACCEPTED MANUSCRIPT

5 Continuity

In this Section we study another robustness property of the composed τ-estimators. In Section SM–4 of the Supplemental Material we show that these estimators can be expressed as estimating functionals defined on the set of distributions of $(\boldsymbol{y}, \boldsymbol{x})$ applied to the empirical distribution F_n of $(\boldsymbol{y}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_n, \boldsymbol{x}_n)$ That is, we can write

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{T}_\beta(F_n), \quad \widehat{\boldsymbol{\gamma}} = \boldsymbol{T}_\gamma(F_n), \quad \widehat{\eta} = T_\eta(F_n), \quad \widehat{\boldsymbol{\upsilon}} = \boldsymbol{T}_\upsilon(F_n)$$

Denote by $\overset{D}{\to}$ convergence in distribution. The estimating functionals $\boldsymbol{T}_{\boldsymbol{\beta}}$, $\boldsymbol{T}_{\boldsymbol{\gamma}}$, T_η and $\boldsymbol{T}_{\boldsymbol{\upsilon}}$ have the following robustness property

Definition 3 *An estimating Functional $\boldsymbol{T} : \mathcal{F} \to \mathbb{R}^h$, where $\mathcal{F}$ is a subset of the distribution on $\mathbb{R}^k$ is continuous at a distribution H if given a sequence of distributions $H_n \overset{D}{\to} H$, then $\boldsymbol{T}(H_n) \to \boldsymbol{T}(H)$.*

Note that the continuity of an estimating functional implies that a small variation in the empirical distribution will produce a small variation in the estimator. Then, this property is closely related to the concept of qualitative robustness introduced by Hampel (1971) and is equivalent to the asymptotic qualitative robustness property defined in Papantoni-Kazakos and Gray (1979).

To prove the continuity of the composite τ-estimators we need the following additional assumptions.

A7 The matrix $\boldsymbol{x}$ is random and independent of the error term $\boldsymbol{u}$. Besides, the error vector $\boldsymbol{u}$ has an elliptical density of the form

$$f(\boldsymbol{u}) = \frac{f_0^*(\boldsymbol{u}^\top \boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0)^{-1}\boldsymbol{u})}{\det(\boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0))^{1/2}}, \tag{21}$$

where f_0^* is non increasing and is strictly decreasing in a neighborhood of 0; **A8** Let H_0 be the distribution of $\boldsymbol{x}$. Then for any $\boldsymbol{b} \in \mathbb{R}^k$, $\boldsymbol{\delta} \neq \boldsymbol{0}$ we have $P_{H_0}(\boldsymbol{x}\boldsymbol{b} \neq \boldsymbol{0}) > 0$ and for all pair $1 \leq j < l \leq p$ and all $\boldsymbol{b} \in \mathbb{R}^2$ we have $P_{H_0}(\boldsymbol{x}^{jl}\boldsymbol{b} = \boldsymbol{0}) < 1 - b$ and **A9** (Identification condition) If $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}^*$ for all α we have $\boldsymbol{\Sigma}(1, \boldsymbol{\gamma}) \neq \boldsymbol{\Sigma}(\alpha, \boldsymbol{\gamma}^*)$.

An important family of distributions satisfying A7 is the multivariate normal, in this case,

$$f_0^*(z) = (2\pi)^{-p/2} \exp(-z/2). \tag{22}$$

The following theorem establishes the continuity at the nominal model of the estimating functionals corresponding to composite τ-estimators.

Theorem 3 *Assume (i) ρ_1 satisfies A1-A5, (ii) ρ_2 satisfies A1-A6, (iii) $(\mathbf{y}, \mathbf{x})$ satisfy $\mathbf{y} = \mathbf{x}\boldsymbol{\beta}_0 + \mathbf{u}$ and A7- A9 hold. Then, if F_0 is the distribution of $(\mathbf{y}, \mathbf{x})$, the composite τ-estimating functionals $\mathbf{T}_{\boldsymbol{\beta}}$, $\mathbf{T}_{\boldsymbol{\gamma}}$, T_η and $\mathbf{T}_{\boldsymbol{\nu}}$ are continuous at F_0.*

6 Asymptotics

Consistency and asymptotic distribution of the composite τ-estimators are discussed in the next two subsections.

6.1 Consistency

Notice that, contamination causes asymptotic bias, and therefore consistency cannot be ensured for any robust procedure in the presence of outliers. For this reason consistency to the true parameters only occurs under the nominal model. The following Theorem states the consistency of composite τ-estimators.

Theorem 4 *Let $(\mathbf{y}_i, \mathbf{x}_i)$, $1 \leq i \leq n$, be i.i.d. random samples of a distribution F_0. Assume (i) ρ_1 satisfies A1-A5, (ii) ρ_2 satisfies A1-A6, (iii) $\mathbf{y}_i = \mathbf{x}_i \boldsymbol{\beta}_0 + \mathbf{u}_i$ and A7, A8 and A9 hold. Then, the composite τ-estimators $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\eta}$ satisfy $\lim_{n\to\infty} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$ a.s. and $\lim_{n\to\infty} \widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}_0$ a.s.. Moreover, if f_0^* is given by (22) we have $\lim_{n\to\infty} \widehat{\eta} = \eta_0$ a.s. too.*

Note that for the consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ is not necessary that $\mathbf{u}$ be multivariate normal.

6.2 Asymptotic normality

The following Theorem states the asymptotic normality of composite τ-estimators. We need the following additional assumptions

A10 Let H_0 be the distribution of $\boldsymbol{x}$. Then H_0 has finite second moments and $E_{H_0}(\boldsymbol{x}\boldsymbol{x}^\top)$ is non–singular. **A11** The functions ρ_i, $i = 1, 2$ are twice differentiable.

Theorem 5 *Let $\boldsymbol{\lambda}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ and $\widehat{\boldsymbol{\lambda}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}})$ be the composite τ-estimator. Consider the same assumptions as in Theorem 4, A10 and A11. Then, we have*

$$\sqrt{n}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0) \stackrel{D}{\to} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\lambda}_0}\right),$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} = E\left[\nabla^2_{\boldsymbol{\lambda}} L(\boldsymbol{\lambda})\right]^{-1} E\left[\nabla_{\boldsymbol{\lambda}} L(\boldsymbol{\lambda}) \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{\lambda})^\top\right] \left(E\left[\nabla^2_{\boldsymbol{\lambda}} L(\boldsymbol{\lambda})\right]^{-1}\right)^\top,$$

and $\nabla_{\boldsymbol{\lambda}} L(\boldsymbol{\lambda})$ and $\nabla^2_{\boldsymbol{\lambda}} L(\boldsymbol{\lambda})$ are the gradient and Hessian matrix of $L(\boldsymbol{\lambda})$ respectively.

We do not give the proof of Theorem 5. However, it can be obtained using standard delta method arguments, see for example Theorem 10.9 in Maronna et al. (2006). This Theorem allows to define Wald tests for null hypothesis and confidence intervals for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, but not for η. However in most practical applications, like testing for the existence of fixed or random effects, the interest is centered in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Table 4 in Section 8 shows that the actual coverage of the confidence intervals obtained using this asymptotic result is close to the nominal level for the Monte Carlo setting we explore.

7 Example

Hereafter we present an application of the estimators introduced here to a real data set. The example is a prospective longitudinal study of children with disorder of neural development. In this data set, outliers are present in the couples rather than in the units and the composite τ-estimator provides a different analysis with respect to maximum likelihood and classical robust procedures.

7.1 Autism

The data used in this example were collected by researchers at the University of Michigan (Anderson et al., 2009) as part of a prospective longitudinal study of 214 children and they are analyzed, among others, also in West et al. (2007). The children were divided into three diagnostic groups at the age of 2 years old: autism, pervasive developmental disorder (PDD), and nonspectrum children. The study was designed to collect information on each child at ages 2, 3, 5, 9, and 13 years, although not all children were measured at each age. Among the objectives of the study there was assessing the relative influence of the initial diagnostic category (autism or PDD), language proficiency at age 2, and other covariates on the developmental trajectories of the socialization of these children. Study participants were children who had consecutive referrals to one of two autism clinics before the age of 3 years. Social development was assessed at each age using the Vineland Adaptive Behavior Interview survey form, a parent-reported measure of socialization. The dependent variable, vsae (Vineland Socialization Age Equivalent), was a combined score that included assessments of interpersonal relationships, play/leisure time activities, and coping skills. Initial language development was assessed using the Sequenced Inventory of Communication Development (SICD) scale; children were placed into one of three groups (sicdegp, $s_{(1)}, s_{(2)}, s_{(3)}$, where $s_{(k)}$ is the indicator function of the k group) based on their initial SICD scores on the expressive language subscale at age 2. We consider the subset of $n = 41$ children for which all measurements are available. We analyze this data using a regression model with random coefficients where vsae is explained by intercept, age, age^2 and sicdegp as a factor variable plus interaction among the age related variables and sicdegp. Hereafter, the variable age is shifted by 2. Let y_{ij} be the value of the

ACCEPTED MANUSCRIPT

i-th vsae for the j-th ages value a_j, then it is assumed that for $1 \le i \le 41$, $1 \le j \le 5$ we have

$$\begin{aligned} y_{ij} &= b_{i1} + b_{i2}a_j + b_{i3}a_j^2 \\ &\quad + \beta_4 s_{(1)i} + \beta_5 s_{(2)i} \\ &\quad + \beta_6 a_j \times s_{(1)i} + \beta_7 a_j \times s_{(2)i} + \beta_8 a_j^2 \times s_{(1)i} + \beta_9 a_j^2 \times s_{(2)i} + \varepsilon_{ij}, \end{aligned}$$

where (b_{i1}, b_{i2}, b_{i3}) are i.i.d. random coefficients with mean $(\beta_1, \beta_2, \beta_3)$ and covariance matrix

$$\mathbf{\Sigma}_b = \begin{pmatrix} \sigma_{11} & \sigma_{1a} & \sigma_{1a^2} \\ \sigma_{1a} & \sigma_{aa} & \sigma_{aa^2} \\ \sigma_{1a^2} & \sigma_{aa^2} & \sigma_{a^2a^2} \end{pmatrix},$$

$\beta_4, \ldots, \beta_9$ are fixed coefficients and the ε_{ij} are i.i.d. random errors independent of the random coefficients with zero mean and variance $\sigma_{\varepsilon\varepsilon}$. Then, the model could be rewritten in term of (1) and (2) with $p = 5$, $n = 41$, $J = 6$ and $k = 9$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{i5})^\top$, and the matrix $\boldsymbol{x}_i$ is

$$\boldsymbol{x}_i = (\boldsymbol{j}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{s}_{(1)i}, \boldsymbol{s}_{(2)i}, \boldsymbol{a} \times \boldsymbol{s}_{(1)i}, \boldsymbol{a} \times \boldsymbol{s}_{(2)i}, \boldsymbol{b} \times \boldsymbol{s}_{(1)i}, \boldsymbol{b} \times \boldsymbol{s}_{(2)i}),$$

where, $\boldsymbol{j}$ is a 5-vector of ones, $\boldsymbol{a} = (a_1, a_2, a_3, a_4, a_5)^\top$, which corresponds to age and $\boldsymbol{b} = \boldsymbol{a}^2$ which corresponds to age^2, $\boldsymbol{s}_{(k)i}$ is a 5-vector with all the elements equal to $s_{(k)i}$ and $k = 1, 2$. The variance and covariance structure $\mathbf{\Sigma}(\eta, \boldsymbol{\gamma}) = \eta(I + \sum_{j=1}^{J} \gamma_j \boldsymbol{V}_j)$ has the following components $\boldsymbol{V}_1 = \boldsymbol{j}\boldsymbol{j}^\top$, $\boldsymbol{V}_2 = \boldsymbol{a}\boldsymbol{a}^\top$, $\boldsymbol{V}_3 = \boldsymbol{b}\boldsymbol{b}^\top$, $\boldsymbol{V}_4 = \boldsymbol{j}\boldsymbol{a}^\top + \boldsymbol{a}\boldsymbol{j}^\top$, $\boldsymbol{V}_5 = \boldsymbol{j}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{j}^\top$ and $\boldsymbol{V}_6 = \boldsymbol{a}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{a}^\top$. $\eta = \sigma_{\varepsilon\varepsilon}$ is the scale of the error term, $\gamma_1 = \sigma_{11}/\sigma_{\varepsilon\varepsilon}$, $\gamma_2 = \sigma_{aa}/\sigma_{\varepsilon\varepsilon}$, $\gamma_3 = \sigma_{a^2a^2}/\sigma_{\varepsilon\varepsilon}$, $\gamma_4 = \sigma_{1a}/\sigma_{\varepsilon\varepsilon}$, $\gamma_5 = \sigma_{1a^2}/\sigma_{\varepsilon\varepsilon}$ and $\gamma_6 = \sigma_{aa^2}/\sigma_{\varepsilon\varepsilon}$.

We estimate the parameters using different methods: restricted maximum likelihood (ML), the Copt and Victoria Feser S-estimator (CVFS-estimator) described in Section 3 as defined in Copt and Victoria-Feser (2006) using a Rocke ρ function with asymptotic rejection point equals to $\alpha = 0.1$, The SMDM estimator as defined in Koller (2013) using a direct approximation for computing the consistency factors and smoothed Huber ψ functions with $c = 1.345$, and our composite τ-estimator with ρ in the family given by (15) with $c_1 = 1$ and $c_2 = 1.64$. A review of the CVFS estimator is available in Section 3.2. The SMDM estimator is an extension of the Huber's Proposal II approach where both the error term and the random effects are "huberized". We use the implementation of the SMDM available in the `R` package `robustlmm` (Koller, 2015).

Table 1 reports the estimators and the inference for the fixed term parameters using different methods, while Table 2 reports the estimators of the random effect terms. ML, CVFS and SMDM provide similar results, while discrepancies are present with the composite τ method. The main differences are on the estimation of the random effects terms, both in size (error variance component) and shape (correlation components). Composite τ assigns part of the total variance to the random components while the other methods assign it to the error term. In fact, variances estimated by composite τ are in general larger than that estimated via the other methods; composite τ suggests negative correlation between intercept and age, while ML, CVFS and SMDM suggest positive correlation. Composite τ provides small estimates compared to the other methods for the error variance. These discrepancies reflects mainly on the inference for the fixed term coefficients where the variable sicdegp is significant using composite τ but is not using ML, CVFS and SMDM procedures. Interactions between age^2 and sicdegp is highly non significant using composite τ and SMDM while it is somewhat significant using CVFS.

To go more deeply into the reasons of differences between composite robust procedure and classic robust procedure results, we investigate cell, couple and row outliers. For a given dimension $1 \leq q \leq p$ we define as q-dimension outliers those q-dimension observations such that the corresponding squared Mahalanobis distance is greater than a quantile order α of a chi-square distribution with q degree of freedom. In particular we call cell, couple and row outliers respectively the 1-dimension, 2-dimension and p-dimension outliers. Composite τ procedure identifies 33 couple outliers out of 410 couples (8%) at $\alpha = 0.999$. The affected rows, with at least one couple outliers, are 12 out of 41. This means that the CVFS and SMDM procedures have to deal with a data set with a level of contamination about 29%. The `R` code to replicate the analysis is available in section SM–7 of the Supplementary Material.

8 Monte Carlo simulations

In this section we describe the results of a Monte Carlo study whose aim is to illustrate the performance of the new procedure in the classical contamination model (CCM) and in the independent contamination model (ICM).

8.1 Model and setting

We consider a 2-way crossed classification with interaction linear mixed model

$$y_{fgh} = \boldsymbol{x}_{fgh}^{\top}\boldsymbol{\beta}_0 + a_f + b_g + c_{fg} + e_{fgh}, \tag{23}$$

where $f = 1, \ldots, F$, $g = 1, \ldots, G$, and $h = 1, \ldots, H$. Here, we set $F = 2$, $G = 2$ and $H = 3$ which leads to $p = F \times G \times H = 12$. $\boldsymbol{x}_{fgh}$ is a $(k + 1) \times 1$ vector where the last k components are from a standard multivariate normal and the first component is identically equal to 1, $\boldsymbol{\beta}_0 = (0, 2, 2, 2, 2, 2)^{\top}$ is $(k + 1) \times 1$ vector of the fixed parameters with $k = 5$. The random variables a_f, b_g and c_{fg} are the random effects which are normally distributed with variances σ_a^2, σ_b^2, and σ_c^2. Arranging the y_{fgh} in lexicon order (ordered by h within g within f) we obtain the vector $\boldsymbol{y}$ of dimension p and in the similar way the $p \times k$ matrix $\boldsymbol{x}$ obtained arranging $\boldsymbol{x}_{fgh}$. Similarly, we set $\boldsymbol{a} = (a_1, \ldots, a_F)^{\top}$, $\boldsymbol{b} = (b_1, \ldots, b_G)^{\top}$ and $\boldsymbol{c} = (c_{11}, \ldots, c_{FG})^{\top}$, that is, $\boldsymbol{a} \sim N_F(\boldsymbol{0}, \sigma_a^2 \boldsymbol{I}_F)$ and similar for $\boldsymbol{b}$ and $\boldsymbol{c}$, while $\boldsymbol{e} = (e_{111}, \ldots, e_{FGH})^{\top} \sim N_p(\boldsymbol{0}, \sigma_e^2 \boldsymbol{I}_p)$. Hence $\boldsymbol{y}$ is a p multivariate normal with mean $\boldsymbol{\mu} = \boldsymbol{x}\boldsymbol{\beta}$ and variance matrix $\boldsymbol{\Sigma}_0 = \eta_0(\boldsymbol{V}_0 + \sum_{j=0}^{J} \gamma_j \boldsymbol{V}_j)$, where $\boldsymbol{V}_0 = \boldsymbol{I}_p$, $\boldsymbol{V}_1 = \boldsymbol{I}_F \otimes \boldsymbol{J}_G \otimes \boldsymbol{J}_H$, $\boldsymbol{V}_2 = \boldsymbol{J}_F \otimes \boldsymbol{I}_G \otimes \boldsymbol{J}_H$, and $\boldsymbol{V}_3 = \boldsymbol{J}_F \otimes \boldsymbol{J}_G \otimes \boldsymbol{I}_H$; $\otimes$ is the Kronecker product and $\boldsymbol{J}$ is a matrix of ones with appropriate dimension. We took $\sigma_a^2 = \sigma_b^2 = 1/16$ and $\sigma_c^2 = 1/8$. Then $\boldsymbol{\gamma}_0 = (\gamma_{10}, \gamma_{20}, \gamma_{30})^{\top} = (\sigma_a^2/\sigma_e^2, \sigma_b^2/\sigma_e^2, \sigma_c^2/\sigma_e^2)^{\top} = (1/4, 1/4, 1/2)^{\top}$ and $\eta_0 = \sigma_e^2 = 1/4$. We consider a sample of size $n = 100$ and four levels of contamination $\varepsilon = 0, 5, 10$ and 15%. In the CCM $n \times \varepsilon$ observations are contaminated by replacing all the elements of the vector $\boldsymbol{y}$ by observations from $\boldsymbol{y}_0 \sim N_p(\boldsymbol{x}_0\boldsymbol{\beta}_0 + \boldsymbol{\omega}_0, \boldsymbol{\Sigma})$ and the corresponding components of $\boldsymbol{x}$ are replaced by the components of $\boldsymbol{x}_0$. The first column of $\boldsymbol{x}_0$ is identically equal to 1 while the last k columns are from $N_{p\times k}(\boldsymbol{\phi}_0, 0.005^2\boldsymbol{I}_{p\times k})$ and all the components of $\boldsymbol{\phi}_0$ equal to 1 in the case of low leverage outliers (lev1) or to 20 for large leverage outliers (lev20). $\boldsymbol{\omega}_0$ is a p-vector of constants all equal to ω_0. In the ICM we replace $n \times p \times \varepsilon$ cells, randomly chosen in the $n \times p = 1200$ values of $\boldsymbol{y}$ by others obtained as in the previous case. In each case we take a grid of values for ω_0 so than we can estimate the least favorable case. For each combination of these factors we compute the CVFS-estimator described in Copt and Victoria-Feser (2006) with Rocke ρ function with asymptotic rejection probability set to 0.01, the composite τ-estimator with ρ_1 and ρ_2 in the family given by (15) with constants $c_1 = 1$ and $c_2 = 1.64$ respectively and the SMDM estimator introduced by Koller (2013). For each case we run 500 Monte Carlo replications.

8.2 Measures of performance

Let $(\boldsymbol{y}, \boldsymbol{x})$ be an observation independent of the sample $(\boldsymbol{y}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_n, \boldsymbol{x}_n)$ used to compute $\widehat{\boldsymbol{\beta}}$ and let $\widehat{\boldsymbol{y}} = \boldsymbol{x}\widehat{\boldsymbol{\beta}}$ be the predicted value of $\boldsymbol{y}$ using $\boldsymbol{x}$. Then the square Mahalanobis distance between $\widehat{\boldsymbol{y}}$ and $\boldsymbol{y}$ using the matrix $\boldsymbol{\Sigma}_0$ is

$$\begin{aligned} m(\widehat{\boldsymbol{y}}, \boldsymbol{y}, \boldsymbol{\Sigma}_0) &= (\widehat{\boldsymbol{y}} - \boldsymbol{y})^\top \boldsymbol{\Sigma}_0^{-1} (\widehat{\boldsymbol{y}} - \boldsymbol{y}) \\ &= (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \boldsymbol{x}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{x} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &+ (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0). \end{aligned}$$

Since $\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0$ is independent of $\boldsymbol{x}$ and has covariance matrix $\boldsymbol{\Sigma}_0$, putting $A = E(\boldsymbol{x}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{x})$ we have

$$\begin{aligned} E\left[m(\widehat{\boldsymbol{y}}, \boldsymbol{y}, \boldsymbol{\Sigma}_0)\right] &= E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top A (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] \\ &+ \operatorname{trace}(\boldsymbol{\Sigma}_0^{-1} (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0)(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta}_0)^\top) \\ &= E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top A (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] + p. \end{aligned}$$

Then, to evaluate an estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ by its prediction performance we can use

$$E\left[m(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0, A)\right] = E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top A(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right]. \tag{24}$$

Let N be the number of replications in the simulation study, and let $\widehat{\boldsymbol{\beta}}_j$, $1 \leq j \leq N$ be the value of $\widehat{\boldsymbol{\beta}}$ at the j-th replication, then we can estimate $E\left[m(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0, A)\right]$ by the Mean Square Mahalanobis distance

$$\text{MSMD} = \frac{1}{N}\sum_{j=1}^{N} m(\widehat{\boldsymbol{\beta}}_j, \boldsymbol{\beta}_0, \boldsymbol{A}).$$

It is easy to prove that as in this case $\boldsymbol{x}$ is a $p \times k$ matrix where the cells are independent $N(0, 1)$ random variables, then $\boldsymbol{A} = \text{trace}(\boldsymbol{\Sigma}_0^{-1})\boldsymbol{I}_k$.

Given two covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_0$, one way to measure how close are $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_0$ is by the Kullback-Leibler divergence between two normal distributions with the same mean and covariance matrices equal to $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_0$ given by

$$\text{KLD}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_0) = \text{trace}\left(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}\right) - \log\left(\det(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1})\right) - p. \tag{25}$$

Since $(\eta_0, \boldsymbol{\gamma}_0)$ determines $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\eta_0, \boldsymbol{\gamma}_0)$, that is, the covariance matrix of $\boldsymbol{y}$ given $\boldsymbol{x}$, one way to measure the performance of an estimator $(\widehat{\eta}, \widehat{\boldsymbol{\gamma}})$ of $(\eta_0, \boldsymbol{\gamma}_0)$ is by

$$E\left[\text{KLD}(\boldsymbol{\Sigma}(\widehat{\eta}, \widehat{\boldsymbol{\gamma}}), \boldsymbol{\Sigma}_0)\right].$$

Let $(\widehat{\eta}_j, \widehat{\boldsymbol{\gamma}}_j), 1 \leq j \leq N$, be the value of $(\widehat{\eta}, \widehat{\boldsymbol{\gamma}})$ at the j-th replication, then we can estimate $E\left[\text{KLD}(\boldsymbol{\Sigma}(\widehat{\eta}, \widehat{\boldsymbol{\gamma}}), \boldsymbol{\Sigma}_0)\right]$ by the Mean Kullback-Leibler Divergence

$$\text{MKLD} = \frac{1}{N}\sum_{j=1}^{N} \text{KLD}(\boldsymbol{\Sigma}(\widehat{\eta}_j, \widehat{\boldsymbol{\gamma}}_j), \boldsymbol{\Sigma}_0).$$

8.3 Results

We summarize hereafter some of the results obtained from the simulations. Table 3 reports the relative efficiency of the CVFS-, SMDM-, and composite τ-estimators with respect to the maximum likelihood in absence of contamination. The efficiency of estimators of $\boldsymbol{\beta}_0$ will be measured by the MSMD ratio while the efficiency of an estimator of $(\eta_0, \boldsymbol{\gamma}_0)$ by the MKLD ratio. For the setting under consideration, the efficiency of the composite τ- estimator is slightly higher of that of the CVFS-estimator for $\boldsymbol{\beta}_0$ and lower of the SMDM-estimator while is considerable higher for $(\eta_0, \boldsymbol{\gamma}_0)$ parameters with respect both competitors.

Table 4 reports the actual coverage of confidence intervals based on the asymptotic distribution of the composite τ-estimator with nominal coverage level of 0.95. The actual coverage seems reasonable close to the nominal level.

We report the results under 10% of both types outlier contamination: classical and independent. Figure 1 reports the behavior of the MSMD as a function of ω_0 while Figure 2 reports the behavior of MKLD. For easy of comparison, Table 5 reports the maximum values of MSMD and MKLD in the range of the Monte Carlo setting. Since similar behavior is observed for negative values of ω_0, these results are not reported.

Analogous behavior was observed for the case 5% and 15% which are not reported. The composite τ-estimator is very competitive with the CVFS- and SMDM-estimators under the classical contamination model, in fact, in the low leverage case (lev1) the maximum values of MSMD and MKLD of the composite τ-estimator are only slightly larger than those of the CVFS-estimator and smaller than those of the SMDM-estimator. Instead for the high leverage case (lev20) the values of MSDM are essentially the same for the CVFS- and the composite τ-estimators, while the maximum value of MKLD is smaller for the composite τ-estimator. The SMDM-estimator seems to breakdown with high leverage points. In the independent contamination model the composite τ-estimator clearly outperforms the CVFS- and the SMDM-estimators. In fact, while the MSMD and MKLD of the composite τ-estimator are always bounded by a small value, the MSMD and MKLD of the CVFS-estimator always show an unbounded behavior, while the SMDM-estimator shows a bounded, but large value for low leverage case (lev1) and an unbounded behavior for the high leverage case (lev20). Mean Square Errors, Biases and Standard Errors for the three estimators

ACCEPTED MANUSCRIPT

and all parameters are available in Section SM–6 of the Supplementary Material. These results confirm the conclusions obtained by the MSMD and MKLD measures of performance. R code to run this Monte Carlo experiment is available as Supplementary Material.

9 Conclusions

The independent contamination model presents new challenging problems for robust statistics. Robust estimators developed for the Tukey-Huber CCM show non-robust behavior under the ICM, in particular their breakdown point converges to zero as the dimension p increases. Furthermore, affine equivariance, so useful for achieving CCM robustness, becomes an obstacle under ICM. We introduce a new class of robust estimators, namely composite τ-estimators which are based on τ-scales of the Mahalanobis distances of two dimensional subvectors of $\boldsymbol{y}$ using the same idea from the composite likelihood. We apply them in linear mixed models estimation. Our methods provide fairly high resistance against both CCM and ICM outliers with breakdown point close to 0.5 and 0.25 respectively.

10 Supplementary Material

Supplementary Material with the derivation of the estimating equations, discussion on computational aspects and algorithms, proofs of theorems and R code for the example and the Monte Carlo experiment is available online.

An R package `robustvarComp` is available in the Comprehensive R Archive Network at cran.r-project.org/web/packages/robustvarComp/index.html. The package implements composite S-estimators and τ-estimators and the CVFS estimator for linear mixed models.

References

F. Alqallaf, S. Van Aelst, R. H. Zamar, and V. J. Yohai. Propagation of outliers in multivariate data. *The Annals of Statistics*, 37(1):311–331, 2009.

D. K. Anderson, R. S. Oti, C. Lord, and K. Welch. Patterns of growth in adaptive social abilities among children with autism spectrum disorders. *Journal of Abnormal Child Psychology*, 37: 1019–1034, 2009.

I. Chervoneva and M. Vishnyakov. Constrained s-estimators for linear mixed effects models with covariance components. *Statistics in Medicine*, 30(14):1735–1750, 2011.

S. Copt and S. Heritier. MM-estimation and inference in mixed linear models. Technical Report 2006.01, Faculté des Sciences Économiques et Sociales, Université de Genève, 2006.

S. Copt and M. P. Victoria-Feser. High breakdown inference in the mixed linear model. *Journal of American Statistical Association*, 101:292–300, 2006.

P. L. Davies. Asymptotic behaviour of S-Estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, 15(3):1269–1292, 1987.

D. L. Donoho. Breakdown properties of multivariate location estimators. *Qualifying paper, Harvard University, Boston*, 1982.

D. L. Donoho and P. J. Huber. The notion of breakdown point. A Festschrift for Erich L. Lehmann, 1983.

W.H. Fellner. Robust estimation of variance components. *Technometrics*, 28:51–60, 1986.

P. S. Gill. A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine*, 19: 975–987, 2000.

F.R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42(6):1887–1896, 1971.

S. Heritier, E. Cantoni, S. Copt, and M. P. Victoria-Feser. *Robust methods in Biostatistics*. John Wiley & Sons, 2009.

H. Jacqmin-Gadda, S. Sibillot, C. Proust, J. Molina, and R. Thiébaut. Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics & Data Analysis*, 51: 5142–5154, 2007.

J. Jiang and W. Zhang. Robust estimation in generalised linear mixed models. *Biometrika*, 88(3): 753–765, 2001.

M. Koller. *Robust Estimation of Linear Mixed Models*. PhD thesis, ETH Zürich, 2013.

M. Koller. robustlmm: Robust linear mixed effects models, July 2015. URL `cran.r-project.org/web/packages/robustlmm/index.html`.

V. H. Lachosa, D. K. Deyb, and V. G. Canchoc. Robust linear mixed models with skew-normal independent distributions from a bayesian perspective. *Journal of Statistical Planning and Inference*, 139:4098–4110, 2009.

B. G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–39, 1988. URL `www.stat.psu.edu/ bgl/center/tr/PUB88a.pdf`.

H.P. Lopuhaä. Multivariate τ-estimators for location and scatter. canadian journal of statistics. *Canadian Journal of Statistics*, 19:307–321, 1991.

R. A. Maronna, R. D. Martin, and V. J. Yohai. *Robust Statistics. Theorey and Methods*. Wiley, 2006.

ACCEPTED MANUSCRIPT

N. Muler and V. J. Yohai. Robust estimators for ARCH processes. *Journal of Time Series Analysis*, 23:341–375, 2002.

P. Papantoni-Kazakos and R.M. Gray. Robustness of estimators on stationary observations. *The Annals of Probability*, 7(6):989–1002, 1979.

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

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

P. Rousseeuw. Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, and Wertz. W., editors, *Mathematical Statistics and Applications*, pages 283–297. Reidel Publishing Company (co-published with Akadémiai e Kiadó, Budapest), 1985.

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

S. K. Sinha. Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association*, 99(466):451–460, 2004.

W. A. Stahel. *Robuste Schätzungen: infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen*. PhD thesis, ETH Zürich, 1981.

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

B. T. West, K. B. Welch, and A. T. Galecki. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman & Hall/CRC, 2007.

V. J. Yohai and R. H. Zamar. High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, 83:460–413, 1988.

V. J. Yohai and R. H. Zamar. Optimal locally robust m-estimates of regression. *Journal of Statistical Planning and Infererence*, 64:309–323, 1997.

ACCEPTED MANUSCRIPT

Table 1: Autism data set. Estimated fixed term parameters by Maximum Likelihood, CVFS-, SMDM-, and composite τ-estimators. P-values are reported between squared parenthesis.

Method	Int.	a	a^2	$s_{(1)}$	$s_{(2)}$	$a \times s_{(1)}$	$a \times s_{(2)}$	$a^2 \times s_{(1)}$	$a^2 \times s_{(2)}$
ML	12.847	6.851	−0.062	−5.245	−2.154	−6.345	−4.512	0.133	0.236
	[0.000]	[0.000]	[0.579]	[0.041]	[0.325]	[0.000]	[0.000]	[0.447]	[0.122]
CVFS	10.934	7.162	−0.108	−4.457	−0.107	−5.770	−4.995	0.094	0.419
	[0.000]	[0.001]	[0.667]	[0.050]	[0.957]	[0.002]	[0.000]	[0.688]	[0.011]
SMDM	12.347	6.020	0.001	−5.192	−2.173	−5.190	−3.870	0.046	0.151
	[0.000]	[0.000]	[0.992]	[0.010]	[0.213]	[0.000]	[0.000]	[0.781]	[0.300]
composite τ	12.145	6.308	−0.089	−5.216	−4.213	−5.361	−3.851	0.082	0.061
	[0.000]	[0.000]	[0.329]	[0.000]	[0.012]	[0.000]	[0.001]	[0.578]	[0.677]

ACCEPTED MANUSCRIPT

Table 2: Autism data set. Estimated random term parameters by Maximum Likelihood, CVFS-, SMDM-, and composite τ-estimators. In round parenthesis the estimated standard errors for CVFS- and composite τ-estimators. Standard errors for ML and SMDM are not available in the used software.

Method	σ_{11}	σ_{aa}	$\sigma_{a^2a^2}$	σ_{1a}	σ_{1a^2}	σ_{aa^2}	$\sigma_{\varepsilon\varepsilon}$
ML	2.647	2.329	0.102	0.774	0.430	−0.038	51.355
CVFS	9.456	3.386	0.222	2.158	1.062	−0.350	22.207
	(44.938)	(11.761)	(0.531)	(13.621)	(1.095)	(1.695)	–
SMDM	5.745	0.092	0.115	0.727	0.813	0.103	25.385
composite τ	9.357	9.680	0.051	−4.024	−0.003	−0.327	5.152
	(5.208)	(3.240)	(0.019)	(2.642)	(0.230)	(0.195)	–

Table 3: Relative efficiency with respect to Maximum Likelihood measured by MSMD ratio for the fixed terms and by MKLD for the random terms for CVFS-, SMDM-, and composite τ- estimators.

Method	MSMD EFF.	MKLD EFF.
CVFS	0.705	0.453
SMDM	0.955	0.147
composite τ	0.799	0.820

ACCEPTED MANUSCRIPT

Table 4: Empirical coverage of confidence intervals based on the asymptotic distribution for the fixed and random term parameters. Nominal level is 0.95. Results are based on 500 Monte Carlo replications.

β_0	β_1	β_2	β_3	β_4	β_5	σ_a^2	σ_b^2	σ_c^2
0.962	0.948	0.946	0.946	0.944	0.928	0.924	0.930	0.924

ACCEPTED MANUSCRIPT

Table 5: Maximum values of MSDM and MKLD in Figures 1 and 2 respectively for CVFS-, SMDM-, and composite τ-estimators.

		CCM		ICM	
	Method	lev1	lev20	lev1	lev20
MSDM	CVFS	0.34	4.43	2406.85	116.08
	SMDM	0.89	23190.69	9.62	23190.24
	composite τ	0.79	4.29	3.17	4.29
MKLD	CVFS	0.20	1.06	5819.80	85.28
	SMDM	0.62	79.43	14.04	79.31
	composite τ	0.43	0.74	2.09	1.19

ACCEPTED MANUSCRIPT

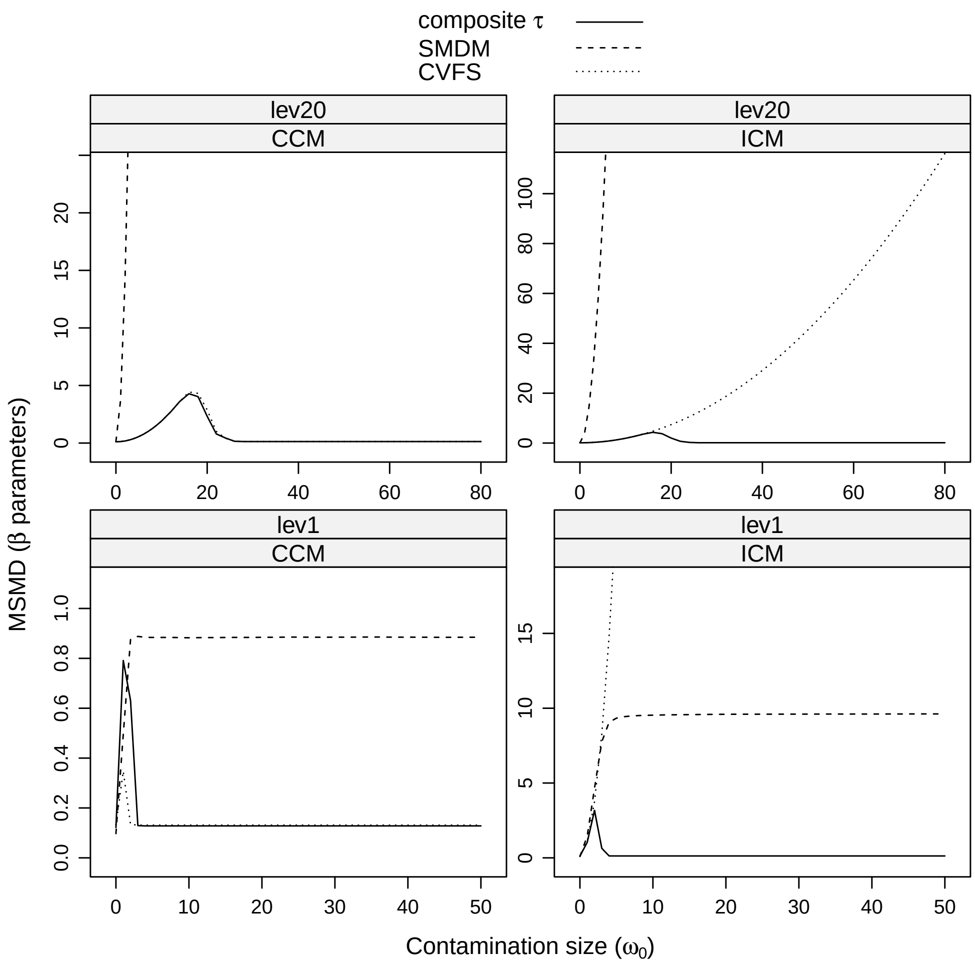

Figure 1: MSMD performance of the CVFS-, SMDM-, and composite τ-estimators of $\boldsymbol{\beta_0}$ under 10% of outlier contamination.

ACCEPTED MANUSCRIPT

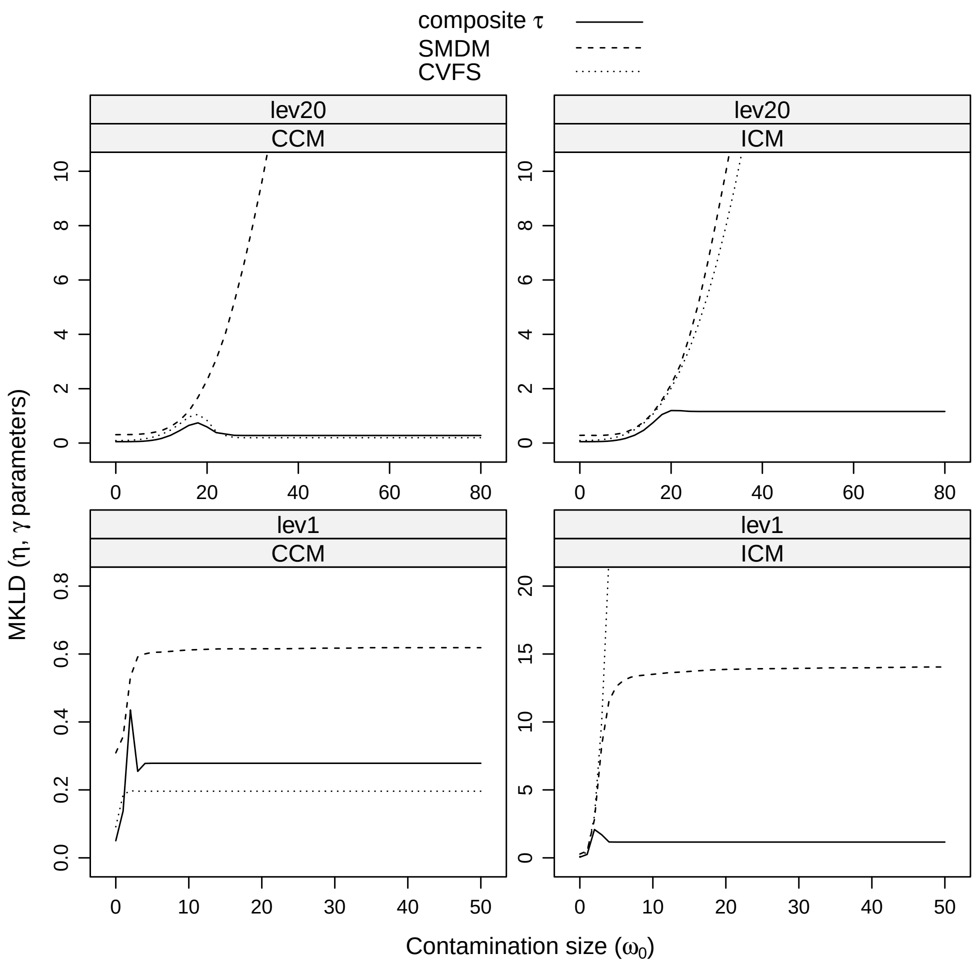

Figure 2: MKLD performance of the CVFS-, SMDM-, and composite τ-estimators of $(\eta_0, \boldsymbol{\gamma_0})$ under 10% of outlier contamination.

ACCEPTED MANUSCRIPT