# Highly Efficient Estimators with High Breakdown Point for Linear Models with Structured Covariance Matrices

Hendrik Paul Lopuhaä

*Delft University of Technology, Mekelweg 4, Delft, 2628 CD, The Netherlands*

## ARTICLE INFO

## ABSTRACT

A unified approach is provided for a method of estimation of the regression parameter in balanced linear models with a structured covariance matrix that combines a high breakdown point with high asymptotic efficiency at models with multivariate normal errors. Of main interest are linear mixed effects models, but our approach also includes several other standard multivariate models, such as multiple regression, multivariate regression, and multivariate location and scatter. Sufficient conditions are provided for the existence of the estimators and corresponding functionals, strong consistency and asymptotic normality is established, and robustness properties are derived in terms of breakdown point and influence function. All the results are obtained for general identifiable covariance structures and are established under mild conditions on the distribution of the observations, which goes far beyond models with elliptically contoured densities. Some results are new and others are more general than existing ones in the literature. In this way, results on high breakdown estimation with high efficiency in a wide variety of multivariate models are completed and improved.

## 1. Introduction

Linear models are widely used and provide a versatile approach for analyzing correlated responses, such as longitudinal data, growth data or repeated measurements. In such models, each subject $i$, $i = 1, \ldots, n$, is observed at $k_i$ occasions, and the vector of responses $\mathbf{y}_i$ is assumed to arise from the model

$$\mathbf{y}_i = \mathbf{X}_i \beta + \mathbf{u}_i,$$

where $\mathbf{X}_i$ is the design matrix for the $i$th subject and $\mathbf{u}_i$ is a vector whose covariance matrix can be used to model the correlation between the responses. One possibility is the linear mixed effects model, in which the random effects together with the measurement error yields a specific covariance structure depending on a vector $\theta$ consisting of some unknown covariance parameters. Other covariance structures may arise, for example if the $\mathbf{u}_i$ are the outcome of a time series. See e.g., Jennrich and Schluchter (1986) or Fitzmaurice et al. (2011), for several possible covariance structures.

Maximum likelihood estimation of $\beta$ and $\theta$ has been studied, e.g., in Hartley and Rao (1967); Rao (1972); Laird and Ware (1982), see also Fitzmaurice et al. (2011); Demidenko (2013). To be resistant against outliers, robust methods have been

---

*E-mail address:* H.P.Lopuhaa@tudelft.nl

investigated for the linear mixed effects models, e.g., in Pinheiro et al. (2001); Copt and Victoria-Feser (2006); Copt and Heritier (2007); Heritier et al. (2009); Koller (2013); Chervoneva and Vishnyakov (2014); Agostinelli and Yohai (2016). This mostly concerns S-estimators, originally introduced in the multiple regression context by Rousseeuw and Yohai (1984) and extended to multivariate location and scatter in Davies (1987); Lopuhaä (1989); Fishbone (2021), to multivariate linear regression in Van Aelst and Willems (2005), and to linear mixed effects models in Copt and Victoria-Feser (2006); Heritier et al. (2009); Chervoneva and Vishnyakov (2011, 2014). A unified approach to S-estimation in balanced linear models with structured covariances can be found in Lopuhaä et al. (2022).

S-estimators are well known smooth versions of the minimum volume ellipsoid estimator (see Rousseeuw (1985)) that are highly resistant against outliers and are asymptotically normal at $\sqrt{n}$-rate. Unfortunately, the choice of the tuning constant corresponding to an S-estimator, forces a trade-off between robustness and efficiency. For this reason, remedies have been developed that retain the high breakdown point of the S-estimator and improve the efficiency of the regression estimator in a second step. One possibility are MM-estimators, introduced by Yohai (1987) in the multiple regression setup. Extensions to multivariate location and scatter can be found in Lopuhaä (1992); Tatsuoka and Tyler (2000); Salibián-Barrera et al. (2006); Fishbone (2021). An extension to linear mixed effects models was discussed in Copt and Heritier (2007) and to multivariate linear regression by Kudraszow and Maronna (2011). An application of MM-estimation to emitter localization can be found in Park and Chang (2021).

We will extend the approaches in Lopuhaä (1992) and Copt and Heritier (2007) to balanced linear models with structured covariance matrices, and postpone MM-estimation for unbalanced models to a future manuscript. The balanced setup is already quite flexible and includes several specific multivariate statistical models. Of main interest are high breakdown estimators with high normal efficiency for linear mixed effects models, but our approach also includes high breakdown estimators in several other standard multivariate models, such as multiple regression, multivariate linear regression, and multivariate location and scatter. We provide sufficient conditions for the existence of the estimators and corresponding functionals, establish their asymptotic properties, such as consistency and asymptotic normality, and derive their robustness properties in terms of breakdown point and influence function. All results are obtained for a large class of identifiable covariance structures, and are established under very mild conditions on the distribution of the observations, which goes far beyond models with elliptically contoured densities. In this way, some of our results are new and others are more general than existing ones in the literature.

The paper is organized as follows. In Section 2, we explain the model in detail and provide some examples of standard multivariate models that are included in our setup. In Section 3 we define the regression M-estimator and M-functional and in Section 4 we give conditions under which they exist. In Section 5 we establish continuity of the regression M-functional, which is then used to obtain consistency of the regression M-estimator. Section 6 deals with the breakdown point. Section 7 provides the preparation for Sections 8 and 9, in which we determine the influence function and establish asymptotic normality. Finally, in Section 10, we investigate the performance of the estimators by means of a simulation and an application to data from a trial on the treatment of lead-exposed children.

## 2. Balanced linear models with structured covariances

We consider independent observations $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$, for which we assume the following model

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{y}_i \in \mathbb{R}^k$ contains repeated measurements for the $i$-th subject, $\boldsymbol{\beta} \in \mathbb{R}^q$ is an unknown parameter vector, $\mathbf{X}_i \in \mathbb{R}^{k \times q}$ is a known design matrix, and the $\mathbf{u}_i \in \mathbb{R}^k$ are unobservable independent mean zero random vectors with covariance matrix $\mathbf{V} \in \mathrm{PDS}(k)$, the class of positive definite symmetric $k \times k$ matrices. The model is balanced in the sense that all $\mathbf{y}_i$ have the same dimension. Furthermore, we consider a structured covariance matrix, that is, the matrix $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$ is a known function of unknown covariance parameters combined in a vector $\boldsymbol{\theta} \in \mathbb{R}^l$. We first discuss some examples that are covered by this setup.

An important case of interest is the (balanced) linear mixed effects model. For a general formulation covered by our setup, see Lopuhaä et al. (2022). A specific example is the model

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{j=1}^{r} \mathbf{Z}_j \gamma_{ij} + \epsilon_i, \quad i = 1, \ldots, n, \tag{2}$$

considered in Copt and Heritier (2007). This model arises from $\mathbf{u}_i = \sum_{j=1}^{r} \mathbf{Z}_j \gamma_{ij} + \epsilon_i$, for $i = 1, \ldots, n$, where the $\mathbf{Z}_j$'s are known $k \times g_j$ design matrices and the $\gamma_{ij} \in \mathbb{R}^{g_j}$ are independent mean zero random variables with covariance matrix $\sigma_j^2 \mathbf{I}_{g_j}$, for $j = 1, \ldots, r$, independent from $\epsilon_i$, which has mean zero and covariance matrix $\sigma_0^2 \mathbf{I}_k$. In this case, $\mathbf{V}(\boldsymbol{\theta}) = \sum_{j=1}^{r} \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j^T + \sigma_0^2 \mathbf{I}_k$ and $\boldsymbol{\theta} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_r^2)$.

Another example of (1), is the multivariate linear regression model

$$\mathbf{y}_i = \mathbf{B}^T \mathbf{x}_i + \mathbf{u}_i, \quad i = 1, \ldots, n, \tag{3}$$

considered in Kudraszow and Maronna (2011), where $\mathbf{B} \in \mathbb{R}^{q \times k}$ is a matrix of unknown parameters, $\mathbf{x}_i \in \mathbb{R}^q$ is known, and $\mathbf{u}_i$, for $i = 1, \ldots, n$, are independent mean zero random variables with covariance matrix $\mathbf{V}(\boldsymbol{\theta}) = \mathbf{C} \in \mathrm{PDS}(k)$. In this case, the

vector of unknown covariance parameters is given by

$$\boldsymbol{\theta} = \text{vech}(\mathbf{C}) = (c_{11}, \ldots, c_{1k}, c_{22}, \ldots, c_{kk})^T \in \mathbb{R}^{\frac{1}{2}k(k+1)}, \tag{4}$$

The model can be obtained as a special case of (1), by taking $\mathbf{X}_i = \mathbf{x}_i^T \otimes \mathbf{I}_k$ and $\boldsymbol{\beta} = \text{vec}(\mathbf{B}^T)$, where $\text{vec}(\cdot)$ is the $k^2$-vector that stacks the columns of a matrix. Clearly, the multiple linear regression model considered in Yohai (1987) is a special case with $k = 1$.

Also the multivariate location-scale model, as considered in Lopuhaä (1992) (see also Salibián-Barrera et al. (2006); Tatsuoka and Tyler (2000)), can be obtained as a special case of (1), by taking $\mathbf{X}_i = \mathbf{I}_k$, the $k \times k$ identity matrix. In this case, $\boldsymbol{\beta} \in \mathbb{R}^k$ is the unknown location parameter and covariance matrix $\mathbf{V}(\boldsymbol{\theta}) = \mathbf{C} \in \text{PDS}(k)$, with $\boldsymbol{\theta}$ as in (4).

Model (1) also includes examples, for which $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are generated by a time series. An example is the case where $\mathbf{u}_i$ has a covariance matrix with elements $v_{st} = \sigma^2 \rho^{|s-t|}$, for $s, t = 1, \ldots, n$. This arises when the $\mathbf{u}_i$'s are generated by an autoregressive process of order one. The vector of unknown covariance parameters is $\boldsymbol{\theta} = (\sigma^2, \rho) \in (0, \infty) \times [-1, 1]$. A general stationary process leads to $v_{st} = \theta_{|s-t|+1}$, for $s, t = 1, \ldots, n$, in which case $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T \in \mathbb{R}^k$, where $\theta_{|s-t|+1}$ represents the autocovariance over lag $|s - t|$.

Throughout the manuscript we will assume that the parameter $\boldsymbol{\theta}$ is identifiable in the sense that, $\mathbf{V}(\boldsymbol{\theta}_1) = \mathbf{V}(\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. This is true for all examples mentioned above.

## 3. Definitions

The definition of MM-estimators involves the use of a single real-valued function $\rho$, or the use of multiple real-valued functions $\rho_0$ and $\rho_1$. Moreover, depending on the specific statistical model of interest, the breakdown behavior of the corresponding MM-estimator may depend on whether the $\rho$-functions are bounded or unbounded. Since we intend to include both possibilities, we first discuss them both.

### 3.1. Bounded and unbounded $\rho$-functions

Yohai (1987) defines the regression MM-estimator in multiple stages. By means of a function $\rho_0$, an M-estimator of scale is determined from residuals, that are obtained from an initial high breakdown regression estimator. Given the M-estimator of scale, a final regression M-estimator is determined by means of a function $\rho_1$. The conditions imposed on the two $\rho$-functions are similar to the following conditions.

(R-BND) $\rho$ is symmetric around zero with $\rho(0) = 0$ and $\rho$ is continuous at zero. There exists a finite constant $c > 0$, such that $\rho$ is strictly increasing on $[0, c]$ and constant on $[c, \infty)$; put $a = \sup \rho = \rho(c)$.

In addition, the two $\rho$-functions are related. Suitable tuning of the bounded function $\rho_0$ ensures a high breakdown point of the scale M-estimator, and by imposing the relationship between $\rho_0$ and $\rho_1$, the final regression M-estimator inherits the high breakdown point from the scale M-estimator. Typical choices for bounded $\rho_0$ and $\rho_1$ that satisfy (R-BND), can be determined from Tukey's biweight, defined as

$$\rho_B(s; c) = \begin{cases} \dfrac{s^2}{2} - \dfrac{s^4}{2c^2} + \dfrac{s^6}{6c^4}, & |s| \le c, \\ \dfrac{c^2}{6} & |s| > c, \end{cases} \tag{5}$$

by taking $\rho_0(d) = \rho_B(d; c_0)$ and $\rho_1(d) = \rho_B(d; c_1)$, where the cut-off constants are chosen such that $0 < c_0 < c_1 < \infty$. The cut-off constant $c_0$ can be tuned such that the MM-estimator inherits the breakdown point of the initial regression estimator, whereas the constant $c_1$ can be tuned such that the MM-estimator has high efficiency at the model with Gaussian errors.

In Lopuhaä (1992) this idea has been extended to multivariate location and scatter by determining a location M-estimator after first obtaining a high breakdown covariance estimator. After rescaling the observations with the initial covariance estimator, a location M-estimator is obtained by minimizing an object function that involves only a single $\rho$-function that satisfies the following condition.

(R-UNB) $\rho$ is symmetric, $\rho(0) = 0$ and $\rho(s) \to \infty$, as $s \to \infty$. The functions $\rho'$ and $u(s) = \rho'(s)/s$ are continuous, $\rho' \ge 0$ on $[0, \infty)$ and there exists a $s_0$ such that $\rho'$ is nondecreasing on $(0, s_0)$ and nonincreasing on $(s_0, \infty)$.

In view of the results found by Huber (1984), an unbounded $\rho$-function is used in Lopuhaä (1992) to avoid that the breakdown point of the location M-estimator depends on the configuration of the sample, which is the case for bounded $\rho$-functions. With an unbounded $\rho$-function, the location M-estimator is shown to inherit the breakdown point of the initial covariance estimator. A typical choice of an unbounded $\rho$-function that satisfies (R-UNB) is

$$\rho_H(s; c) = \begin{cases} \dfrac{s^2}{2}, & |s| \le c, \\ -\dfrac{c^2}{2} + c|s|, & |s| > c, \end{cases} \tag{6}$$

whose derivative $\psi_H = \rho_H'$ is a bounded monotone function known as Huber's $\psi$-function. The constant $c$ can be tuned such that the location M-estimator has high efficiency at the multivariate normal distribution.

Tatsuoka and Tyler (2000) and Salibián-Barrera et al. (2006), propose a different version of location MM-estimators also using bounded $\rho$-functions. Instead of using the entire covariance matrix as auxiliary statistic, they estimate the shape of the scatter matrix along with location parameter and only use a univariate auxiliary estimator for the scale of the scatter matrix. In Salibián-Barrera et al. (2006) it is shown that the location and shape estimators in the second step inherit the breakdown point of the initial estimators used in the first step. Kudraszow and Maronna (2011) use a similar version for multivariate linear regression and also establish that the regression and shape estimators in the second step inherit the breakdown point of the initial estimators used in the first step. Copt and Heritier (2007) treat regression MM-estimators in the context of linear mixed effects models. They allow both bounded and unbounded $\rho$-functions and briefly discuss the pros and cons, but do not explicitly derive the breakdown point.

Extending the approach in Lopuhaä (1992) and Copt and Heritier (2007) to the regression parameter $\beta$ in the current setup (1) seems straightforward. First obtain a high breakdown structured covariance estimator and determine a regression M-estimator from the re-scaled observations. However, in order to make sure that the resulting M-estimator inherits the breakdown point from the initial covariance estimator, the use of an unbounded $\rho$-function, as in Lopuhaä (1992), does not seem to be suitable. The presence of the design matrices $\mathbf{X}_i$ in the object function to be minimized, makes things more complex than for multivariate location. Alternatively, one could minimize a single object function based on a bounded $\rho$-function. However, in view of the results in Huber (1984), in this case it seems difficult to ensure that the resulting regression M-estimator inherits the breakdown point from the initial covariance estimator.

### 3.2. The regression M-estimator and corresponding M-functional

We start by representing our observations as points in $\mathbb{R}^k \times \mathbb{R}^{kq}$ in the following way. For $r = 1, \ldots, k$, let $\mathbf{x}_r^T$ denote the $r$-th row of the $k \times q$ matrix $\mathbf{X}$, so that $\mathbf{x}_r \in \mathbb{R}^q$. We represent the pair $\mathbf{s} = (\mathbf{y}, \mathbf{X})$ as an element in $\mathbb{R}^k \times \mathbb{R}^{kq}$ defined by $\mathbf{s}^T = (\mathbf{y}^T, \mathbf{x}_1^T, \ldots, \mathbf{x}_k^T)$. In this way our observations can be represented as $\mathbf{s}_1, \ldots, \mathbf{s}_n$, with $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i) \in \mathbb{R}^k \times \mathbb{R}^{kq}$.

We will show that an approach similar to Yohai (1987), using two bounded $\rho$-functions that are suitably related turns out to be helpful. For the moment, we intend to include both bounded as well as unbounded $\rho$-functions in our approach. In order to do so, the estimator for $\beta$ is defined in two stages as follows.

**Definition 1.** Let $\mathbf{V}_{0,n}$ be a (high breakdown) positive definite symmetric covariance estimator. For a function $\rho_1 : \mathbb{R} \to [0, \infty)$, define $\beta_{1,n}$ as the vector that minimizes

$$R_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left( \sqrt{(\mathbf{y}_i - \mathbf{X}_i\beta)^T \mathbf{V}_{0,n}^{-1} (\mathbf{y}_i - \mathbf{X}_i\beta)} \right). \tag{7}$$

At this point, $\rho_1$ can be either bounded or unbounded. Later on, we will further specify under what conditions on $\rho_1$, several properties hold for $\beta_{1,n}$. Note that one may choose any initial (high breakdown) covariance estimator, but in our setup we typically think of a structured covariance estimator $\mathbf{V}_{0,n} = \mathbf{V}(\theta_{0,n})$, where $\theta_{0,n}$ is an initial estimator for the vector of covariance parameters. This means that $\mathbf{V}_{0,n}$ is not necessarily affine equivariant, and similarly for $\beta_{1,n}$. However, it is not difficult to see that $\beta_{1,n}$ is regression equivariant, i.e.,

$$\beta_{1,n}(\{(\mathbf{y}_i + \mathbf{X}_i\mathbf{b}, \mathbf{X}_i), i = 1, \ldots, n\}) = \beta_{1,n}(\{(\mathbf{y}_i, \mathbf{X}_i), i = 1, \ldots, n\}) + \mathbf{b},$$

for all $\mathbf{b} \in \mathbb{R}^q$. The corresponding functional is defined similarly.

**Definition 2.** Let $\mathbf{V}_0(P)$ be a positive definite symmetric covariance functional. For a function $\rho_1 : \mathbb{R} \to [0, \infty)$, define $\beta_1(P)$ as the vector that minimizes

$$R_P(\beta) = \int \rho_1 \left( \sqrt{(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}_0(P)^{-1} (\mathbf{y} - \mathbf{X}\beta)} \right) dP(\mathbf{y}, \mathbf{X}). \tag{8}$$

The functional $\beta_1(P)$ is regression equivariant in the sense that

$$\beta_1(P_{\mathbf{y}+\mathbf{Xb},\mathbf{X}}) = \beta_1(P_{\mathbf{y},\mathbf{X}}) + \mathbf{b},$$

for all $\mathbf{b} \in \mathbb{R}^q$, where $P_{\mathbf{y},\mathbf{X}}$ denotes the distribution of $(\mathbf{y}, \mathbf{X})$. Clearly, if one takes $P = \mathbb{P}_n$, the empirical measure of the sample $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$, then $\beta_1(\mathbb{P}_n) = \beta_{1,n}$.

As before, $\mathbf{V}_0(P)$ can be any covariance functional, but in our setup we typically think of a structured covariance $\mathbf{V}_0(P) = \mathbf{V}(\theta_0(P))$, where $\theta_0(P)$ is an initial functional representing the vector of covariance parameters. An example of an estimator $\theta_{0,n}$ and corresponding functional $\theta_0(P)$ that yield a high breakdown structured covariance estimator $\mathbf{V}(\theta_{0,n})$, is the S-estimator $\theta_{0,n}$ and its corresponding functional proposed in Lopuhaä et al. (2022).

Definitions 1 and 2 coincide with the ones for the multivariate location M-estimator in Lopuhaä (1992), when we choose $\mathbf{X}_i = \mathbf{I}_k$ and $\mathbf{V}(\theta) = \mathbf{C} \in \text{PDS}(k)$, with $\theta$ as in (4). For the multiple linear regression model (3), it follows that if $\beta_{1,n}$ exists, then it satisfies score equation (2.6) and equation (2.7) in Yohai (1987). Similarly, for the linear mixed effects model (2), it follows that if $\beta_{1,n}$ exists, then it satisfies a score equation similar to equation (8) in Copt and Heritier (2007).

We should emphasize that score equations like (2.6) in Yohai (1987) and (8) in Copt and Heritier (2007) are useful to obtain asymptotic properties, but they do not guarantee that $\beta_{1,n}$ inherits the breakdown point of the estimators used in the first step. Breakdown behavior is typically established from the minimization problem in Definition 1 itself. If this minimization problem has a solution $\beta_{1,n}$ and if this solution inherits the high breakdown point from $\mathbf{V}_{0,n}$, then $\beta_{1,n}$ will be a zero of the corresponding score equation with a high breakdown point. But just being a zero of the score equation does not ensure a high breakdown point. Indeed, the breakdown point of the MM-estimators in the multiple linear regression model Yohai (1987) and the multivariate location-scale model Lopuhaä (1992), have been obtained from the respective minimization problems, and similarly for the MM-estimators in Salibián-Barrera et al. (2006) and Kudraszow and Maronna (2011). For the MM-estimator in the linear mixed effects model Copt and Heritier (2007), the robustness properties have not been investigated. In view of the fact that the use of bounded or unbounded $\rho$-functions may lead to different breakdown behavior, the breakdown point of MM-estimators for the linear mixed effects model considered in Copt and Heritier (2007) will be investigated in Section 6.

## 4. Existence

Consider the functional $\beta_1(P)$, as defined in Definition 2. We will establish existence of $\beta_1(P)$, where we allow both bounded and unbounded $\rho_1$. Existence of the corresponding estimator $\beta_{1,n}$ will follow from this. Also of interest is the special case in which $P$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density of the form

$$f_{\mu,\Sigma}(\mathbf{y}) = \det(\Sigma)^{-1/2} h\big((\mathbf{y} - \mu)^T \Sigma^{-1}(\mathbf{y} - \mu)\big), \tag{9}$$

with $\mu = \mathbf{X}\beta \in \mathbb{R}^k$ and $\Sigma = \mathbf{V}(\theta) \in \text{PDS}(k)$, and $h : [0, \infty) \to [0, \infty)$. For the linear mixed effects model in Copt and Heritier (2007), it is assumed that $\mathbf{y} \mid \mathbf{X}$ has a multivariate normal distribution, which is a special case of (9) with $h(t) = (2\pi)^{-k/2} \exp(-t/2)$.

For bounded $\rho_1$, we want to rule out the pathological case, where $P$ has all of its mass outside the ellipsoid centered around the origin with covariance structure $\mathbf{V}_0(P)$ and radius $c_1$. To this end we require the following condition on $P$.

(A) Suppose that

$$R_P(\mathbf{0}) = \int \rho_1\left(\sqrt{\mathbf{y}^T \mathbf{V}_0(P)^{-1} \mathbf{y}}\right) dP(\mathbf{y}, \mathbf{X}) < \sup \rho_1.$$

Clearly, if $\mathbf{y} \mid \mathbf{X} \sim \mathcal{N}(\mathbf{X}\beta, \Sigma)$ and $\rho_1$ satisfies (R-BND), this condition is trivially fulfilled. We then have the following theorem for bounded $\rho_1$.

**Theorem 1.** *Let $\rho_1 : \mathbb{R} \to [0, \infty)$ satisfy condition (R-BND), and suppose that $\mathbf{X}$ has full rank with probability one.*

(i) *If $P$ satisfies (A), then there is at least one vector $\beta_1(P)$ that minimizes $R_P(\beta)$.*
(ii) *When $P$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density from (9) with parameters $\mu = \mathbf{X}\beta$ and $\Sigma$, and if $\mathbf{V}_0(P) = \Sigma$, then $R_P(\mathbf{b}) \geq R_P(\beta)$, for all $\mathbf{b} \in \mathbb{R}^q$. When $h$ in (9) and $-\rho_1$ have a common point of decrease, then $R_P(\mathbf{b})$ is uniquely minimized by $\beta_1(P) = \beta$.*

**Proof.** (i) Let $0 < \lambda_1 < \infty$ be the largest eigenvalue of $\mathbf{V}_0(P)$, and let $\lambda_k(\mathbf{X}^T\mathbf{X}) > 0$ denote the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}$. Let $\|\cdot\|$ denote the Euclidean norm. Then we have that

$$\sqrt{(\mathbf{y} - \mathbf{Xb})^T \mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{Xb})} \geq \frac{\|\mathbf{y} - \mathbf{Xb}\|}{\sqrt{\lambda_1}} \geq \frac{\|\mathbf{Xb}\| - \|\mathbf{y}\|}{\sqrt{\lambda_1}}$$
$$\geq \frac{1}{\sqrt{\lambda_1}}\left(\|\mathbf{b}\|\sqrt{\lambda_k(\mathbf{X}^T\mathbf{X})} - \|\mathbf{y}\|\right). \tag{10}$$

Then by dominated convergence and (R-BND), it follows that

$$\lim_{\|\mathbf{b}\| \to \infty} R_P(\mathbf{b}) = \int \lim_{\|\mathbf{b}\| \to \infty} \rho_1\left(\sqrt{(\mathbf{y} - \mathbf{Xb})^T \mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{Xb})}\right) dP(\mathbf{y}, \mathbf{X}) = \sup \rho_1. \tag{11}$$

According to condition (A), this means that there exists a constant $M > 0$, such that

$$R_P(\mathbf{b}) > R_P(\mathbf{0}), \quad \text{for all } \|\mathbf{b}\| > M. \tag{12}$$

Therefore, for minimizing $R_P(\mathbf{b})$ we may restrict ourselves to the set $K = \{\mathbf{b} \in \mathbb{R}^q : \|\mathbf{b}\| \leq M\}$. By dominated convergence and (R-BND), it also follows that $R_P(\mathbf{b})$ is continuous on the compact set $K$, and therefore it must attain at least one minimum $\beta_1(P)$.

(ii) Write

$$R_P(\mathbf{b}) = \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_{\mathbf{y}\mid\mathbf{X}}\left[\rho_1\left(\sqrt{(\mathbf{y} - \mathbf{Xb})^T \Sigma^{-1}(\mathbf{y} - \mathbf{Xb})}\right)\right]\right].$$

By change of variables $\mathbf{y} = \boldsymbol{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}$, the inner conditional expectation can be written as

$$\int \rho_1\big(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \mathbf{Xb})\|\big)f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{y})\,\mathrm{d}\mathbf{y} = \int \rho_1\big(\|\mathbf{z} - \boldsymbol{\Sigma}^{-1/2}\mathbf{X}(\mathbf{b} - \boldsymbol{\beta})\|\big)h(\mathbf{z}^T\mathbf{z})\,\mathrm{d}\mathbf{z}.$$

Next, we apply Lemma 4 from Davies (1987) to the functions $\xi(d) = 1 - \rho_1(\sqrt{d})/a_0$ and $g = h$ and taking $\Lambda = \mathbf{I}_k$. Since $-\rho_1$ and $h$ have a common point of decrease, for all $\mathbf{X}$, it follows that

$$\int \rho_1\big(\|\mathbf{z} - \boldsymbol{\Sigma}^{-1/2}\mathbf{X}(\mathbf{b} - \boldsymbol{\beta})\|\big)h(\mathbf{z}^T\mathbf{z})\,\mathrm{d}\mathbf{z} \le \int \rho_1(\|\mathbf{z}\|)h(\mathbf{z}^T\mathbf{z})\,\mathrm{d}\mathbf{z},$$

with a strict inequality unless $\boldsymbol{\Sigma}^{-1/2}\mathbf{X}(\mathbf{b} - \boldsymbol{\beta}) = \mathbf{0}$, i.e., unless $\mathbf{b} = \boldsymbol{\beta}$, since $\mathbf{X}$ has full rank with probability one. Finally, with the same change of variables $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, the right hand side can be written as

$$\int \rho_1(\|\mathbf{z}\|)h(\mathbf{z}^T\mathbf{z})\,\mathrm{d}\mathbf{z} = \int \rho_1\big(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|\big)f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{y})\,\mathrm{d}\mathbf{y}.$$

After taking expectations $\mathbb{E}_{\mathbf{X}}$, we conclude that $R_P(\mathbf{b}) \le R_P(\boldsymbol{\beta})$, with a strict inequality, unless $\mathbf{b} = \boldsymbol{\beta}$. This proves the theorem. $\qquad\square$

For bounded $\rho_1$, the function $R_P(\boldsymbol{\beta})$ in (8) is well defined. This is not necessarily true for unbounded $\rho_1$. However, this will be the case when $P$ has a first moment. For unbounded $\rho_1$ we have the following result.

**Theorem 2.** *Let $\rho_1 : \mathbb{R} \to [0, \infty)$ satisfy condition (R-UNB). Suppose that $\mathbb{E}_P\|\mathbf{s}\| < \infty$ and that $\mathbf{X}$ has full rank with probability one.*

  *(i) For every $\boldsymbol{\beta} \in \mathbb{R}^q$ fixed, $R_P(\boldsymbol{\beta}) < \infty$.*
  *(ii) There is at least one vector $\boldsymbol{\beta}_1(P)$ that minimizes $R_P(\boldsymbol{\beta})$. When $\rho_1$ is also strictly convex, then $\boldsymbol{\beta}_1(P)$ is uniquely defined.*
  *(iii) When $P$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density from (9) with parameters $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, and if $\mathbf{V}_0(P) = \boldsymbol{\Sigma}$, then $R_P(\mathbf{b}) \ge R_P(\boldsymbol{\beta})$, for all $\mathbf{b} \in \mathbb{R}^q$. When $h$ in (9) is strictly decreasing, then $R_P(\mathbf{b})$ is uniquely minimized by $\boldsymbol{\beta}_1(P) = \boldsymbol{\beta}$.*

**Proof.** Let $0 < \lambda_k \le \lambda_1 < \infty$ be the smallest and largest eigenvalue of $\mathbf{V}_0(P)$.
(i) Condition (R-UNB) implies that $\rho_1(s) \le \rho_1(s_0)$, for $s \in [0, s_0]$, and that for $s > s_0$,

$$\rho_1(s) = \int_0^{s_0} \rho_1'(t)\,\mathrm{d}t + \int_{s_0}^s \rho_1'(t)\,\mathrm{d}t \le \rho_1(s_0) + (s - s_0)\rho_1'(s_0). \tag{13}$$

Hence, for $\|\mathbf{V}_0(P)^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\| > s_0$, we have that

$$\rho_1\big(\|\mathbf{V}_0(P)^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|\big) \le \rho_1(s_0) + \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|\lambda_k^{-1/2}\rho_1'(s_0) - s_0\rho_1'(s_0)$$
$$\le \rho_1(s_0) + (\|\mathbf{y}\| + \|\mathbf{X}\| \cdot \|\boldsymbol{\beta}\|)\lambda_k^{-1/2}\rho_1'(s_0). \tag{14}$$

Since $\mathbb{E}_P\|\mathbf{s}\| < \infty$, we find that for any $\boldsymbol{\beta} \in \mathbb{R}^q$ fixed,

$$\int \rho_1\big(\|\mathbf{V}_0(P)^{-1/2}(\mathbf{y} - \mathbf{Xb})\|\big)\,\mathrm{d}P(\mathbf{s}) \le \rho_1(s_0) + (\mathbb{E}_P\|\mathbf{y}\| + \|\boldsymbol{\beta}\|\mathbb{E}_P\|\mathbf{X}\|)\lambda_k^{-1/2}\rho_1'(s_0) < \infty,$$

which proves part (i).
(ii) We first argue that for minimizing $R_P(\boldsymbol{\beta})$, we can restrict ourselves to a compact set. Note that $R_P(\mathbf{0}) < \infty$, according to part (i). Now, suppose that $\|\boldsymbol{\beta}\| > M$. Then from (10),

$$\sqrt{(\mathbf{y} - \mathbf{Xb})^T\mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{Xb})} \ge \frac{1}{\sqrt{\lambda_1}}\Big(\|\mathbf{b}\|\sqrt{\lambda_k(\mathbf{X}^T\mathbf{X})} - \|\mathbf{y}\|\Big) \ge \frac{\sqrt{M}}{2\sqrt{\lambda_1}}, \tag{15}$$

on the set

$$A_M = \Big\{(\mathbf{y}, \mathbf{X}) \in \mathbb{R}^{k+qk} : \lambda_k(\mathbf{X}^T\mathbf{X}) \ge 1/M;\ \|\mathbf{y}\| \le \sqrt{M}/2\Big\}. \tag{16}$$

Since $\mathbf{X}$ has full rank with probability one, $P(A_M) \to 1$ and $\rho_1(\sqrt{M}/(2\sqrt{\lambda_1})) \to \infty$, as $M \to \infty$, according to (R-UNB). This implies that for $M$ sufficiently large,

$$R_P(\boldsymbol{\beta}) \ge \rho_1(\sqrt{M}/(2\sqrt{\lambda_1}))P(A_M) > R_P(\mathbf{0}).$$

Therefore, that there exists a constant $M > 0$, such that for minimizing $R_P(\boldsymbol{\beta})$ we may restrict ourselves to the compact set $K = \{\boldsymbol{\beta} \in \mathbb{R}^q : \|\boldsymbol{\beta}\| \le M\}$. Since $\mathbb{E}_P\|\mathbf{s}\| < \infty$, from (14) and dominated convergence, it follows that $R_P(\boldsymbol{\beta})$ is continuous on $K$ and therefore it must attain at least one minimum $\boldsymbol{\beta}_1(P)$ on the compact set $K$. It is easily seen that strict convexity of $\rho_1$ implies strict convexity of $R_P$, which means that $\boldsymbol{\beta}_1(P)$ is unique.
(iii) Because $\boldsymbol{\beta}_1(\cdot)$ is regression equivariant, we may assume that $\boldsymbol{\beta} = \mathbf{0}$. Write

$$R_P(\mathbf{b}) = \mathbb{E}_{\mathbf{X}}\big[\mathbb{E}_{\mathbf{y}|\mathbf{X}}\big[\rho_1\big(\|\mathbf{V}_0(P)^{-1/2}(\mathbf{y} - \mathbf{Xb})\|\big)\big]\big].$$

Since $\mathbf{z} = \mathbf{V}_0(P)^{-1/2}\mathbf{y} = \mathbf{\Sigma}^{-1/2}\mathbf{y}$ has an elliptically contoured density with parameters $(\mathbf{0}, \mathbf{I}_k)$, the inner conditional expectation can be written as

$$\iint \left\{ 0 \leq s \leq \rho_1(\|\mathbf{z} - \mathbf{V}_0(P)^{-1/2}\mathbf{X}\mathbf{b}\|) \right\} f(\|\mathbf{z}\|) \, ds \, d\mathbf{z}.$$

From here on, we can copy the proof of Theorem 2.1 in Lopuhaä (1992) and conclude that

$$\mathbb{E}_{\mathbf{y}|\mathbf{x}}\left[ \rho_1\left( \|\mathbf{V}_0(P)^{-1/2}(\mathbf{y} - \mathbf{X}\mathbf{b})\| \right) \right] \geq 0, \quad \mathbf{X} - \text{a.s.} \tag{17}$$

It follows that $R_P(\mathbf{b}) \geq 0 = R_P(\mathbf{0})$. When $f$ is strictly decreasing, similar to the proof of Theorem 2.1 in Lopuhaä (1992), it follows that inequality (17) is strict, which yields $R_P(\mathbf{b}) > 0 = R_P(\mathbf{0})$. □

A direct consequence of Theorems 1 and 2 is the existence of $\beta_{1,n}$.

**Corollary 1.** *Let* $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$ *be a sample, such that* $\mathbf{X}_i$ *has full rank for each* $i = 1, \ldots, n$.

(i) *If* $\rho_1 : \mathbb{R} \to [0, \infty)$ *satisfies conditions (R-BND) and* $R_n(\mathbf{0}) < \sup \rho_1$, *then there exists at least one* $\beta_{1,n}$ *that minimizes* $R_n(\beta)$.

(ii) *If* $\rho_1 : \mathbb{R} \to [0, \infty)$ *satisfies condition (R-UNB), then there exists at least one* $\beta_{1,n}$ *that minimizes* $R_n(\beta)$. *When* $\rho_1$ *is also strictly convex, then* $\beta_{1,n}$ *is uniquely defined.*

**Proof.** Take $P$ equal to the empirical measure $\mathbb{P}_n$ of the sample $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$. If $\mathbf{X}_i$ has full rank, for each $i = 1, \ldots, n$, then if $\rho_1 : \mathbb{R} \to [0, \infty)$ satisfies either (R-BND) or condition (R-UNB), the corollary follows from Theorems 1(i) and 2(ii). □

The condition $R_n(\mathbf{0}) < \sup \rho_1$ is not very restrictive. It rules out the pathological case of all observations being outside the ellipsoid centered around the origin with covariance structure $\mathbf{V}_{0,n}$ and radius $c_1$.

Existence of regression MM-estimators was not considered in Copt and Heritier (2007) for linear mixed effects models or in Yohai (1987) for multiple linear regression. Their existence now follows from Corollary 1. Existence of location MM-estimators, as obtained in Lopuhaä (1992), now also follows from Corollary 1 as a special case of part (ii). Existence of a slightly different MM-estimator has been established in Tatsuoka and Tyler (2000) for multivariate location, and in Kudraszow and Maronna (2011) for multivariate linear regression.

## 5. Continuity and Consistency

Consider a sequence $P_t$, $t \geq 0$, of probability measures on $\mathbb{R}^k \times \mathbb{R}^{kq}$ that converges weakly to $P$, as $t \to \infty$. By continuity of the functional $\beta_1(P)$ we mean that $\beta_1(P_t) \to \beta_1(P)$, as $t \to \infty$. An example of such a sequence is the sequence of empirical measures $\mathbb{P}_n$, $n = 1, 2, \ldots$, that converges weakly to $P$, almost surely. Continuity of the functional $\beta_1(P)$ for this sequence would then mean that the estimator $\beta_{1,n}$ is consistent, i.e., $\beta_{1,n} = \beta_1(\mathbb{P}_n) \to \beta_1(P)$, almost surely. Furthermore, continuity of $\beta_1(P)$ also provides a first step in deriving the influence function, in the sense that $\beta_1(P_{\epsilon, \mathbf{s}_0}) \to \beta_1(P)$, as $\epsilon \downarrow 0$, where

$$P_{\epsilon, \mathbf{s}_0} = (1 - \epsilon)P + \epsilon \delta_{\mathbf{s}_0}, \tag{18}$$

with $\delta_{\mathbf{s}_0}$ representing the Dirac measure at $\mathbf{s}_0 = (\mathbf{y}_0, \mathbf{X}_0)$.

When $\rho_1$ is bounded, we can obtain continuity of the functional $\beta_1(P)$ for general weakly convergent sequences $P_t$, $t \geq 0$. When $\rho_1$ is unbounded, this becomes more complicated, but we can still establish continuity for the sequence of empirical measures $\mathbb{P}_n$, $n = 1, 2, \ldots$, and for the sequence $P_{\epsilon, \mathbf{s}_0}$, for $\epsilon \downarrow 0$. For bounded $\rho_1$ we have the following theorem.

**Theorem 3.** *Let* $P_t$, $t \geq 0$ *be a sequence of probability measures on* $\mathbb{R}^k \times \mathbb{R}^{kq}$ *that converges weakly to* $P$, *as* $t \to \infty$. *Suppose that* $\rho_1 : \mathbb{R} \to [0, \infty)$ *satisfies (R-BND) and suppose that* $P$ *is such that (A) holds and that* $\mathbf{X}$ *has full rank with probability one. Suppose that for* $t$ *sufficiently large,* $\mathbf{V}_0(P_t)$ *exists and that*

$$\lim_{t \to \infty} \mathbf{V}_0(P_t) = \mathbf{V}_0(P). \tag{19}$$

*Then for* $t$ *sufficiently large, there exists at least one* $\beta_1(P_t)$ *that minimizes* $R_{P_t}(\beta)$. *If* $\beta_1(P)$ *is the unique minimizer of* $R_P(\beta)$, *then for any sequence* $\beta_1(P_t)$, $t \geq 0$, *it holds that*

$$\lim_{t \to \infty} \beta_1(P_t) = \beta_1(P).$$

**Proof.** Similar to Lemma B.1 in Lopuhaä et al. (2022), one can show that

$$\lim_{t \to \infty} \int \rho_1(d(\mathbf{s}, \beta_t, \mathbf{V}_t)) \, dP_t(\mathbf{s}) = \int \rho_1(d(\mathbf{s}, \beta_L, \mathbf{V}_L)) \, dP(\mathbf{s}), \tag{20}$$

for any sequence $(\beta_t, \mathbf{V}_t) \to (\beta_L, \mathbf{V}_L)$, where

$$d^2(\mathbf{s}, \beta, \mathbf{V}) = (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta). \tag{21}$$

In particular, this yields that, for every $\beta \in \mathbb{R}^q$ fixed, it holds that $R_{P_t}(\beta) \to R_P(\beta)$, as $t \to \infty$. We first show that there exists $M > 0$, such that for minimizing $R_{P_t}(\beta)$, we can restrict ourselves to $\|\beta\| \leq M$ for $t$ sufficiently large. Consider the set $A_M$

defined in (16). Due to condition (A) and the fact that **X** has full rank, with probability one, for any $\eta > 0$, we can find an $M > 0$, such that $\rho_1(\sqrt{M}/(2\sqrt{\lambda_1}))P(A_M) > R_P(\mathbf{0}) + 2\eta$. On the other hand, for any $\eta > 0$, we have

$$|R_{P_t}(\mathbf{0}) - R_P(\mathbf{0})| \leq \eta, \tag{22}$$

for $t$ sufficiently large. If $\beta$ minimizes $R_{P_t}(\beta)$, for $t$ sufficiently large, we must have $\|\beta\| \leq M$, since otherwise, according to (15),

$$R_{P_t}(\beta) \geq \rho_1(\sqrt{M}/(2\sqrt{\lambda_1}))P(A_M) > R_P(\mathbf{0}) + 2\eta \geq R_{P_t}(\mathbf{0}) + \eta > R_{P_t}(\mathbf{0}).$$

Hence, for minimizing $R_{P_t}(\beta)$, we can restrict to the compact set $K = \{\beta \in \mathbb{R}^q : \|\beta\| \leq M\}$. Furthermore, as in the proof of Theorem 1, the function $R_{P_t}(\beta)$ is continuous on the compact set $K$, and must therefore attain a minimum $\beta_1(P_t)$.

According to Theorem 1 there exists at least one $\beta_1(P)$ that minimizes $R_P(\beta)$. Now, suppose that $\beta_1(P)$ is unique. Because $\beta_1(P)$ is regression equivariant, we may assume that $\beta_1(P) = \mathbf{0}$. For the sake of brevity, let us write $\beta_{1,t} = \beta_1(P_t)$, $\mathbf{V}_{0,t} = \mathbf{V}_0(P_t)$, and $R_t = R_{P_t}$. From (19) it follows that for $t$ sufficiently large,

$$0 < \lambda_k(\mathbf{V}_0(P))/4 \leq \lambda_k(\mathbf{V}_{0,t}) \leq \lambda_1(\mathbf{V}_{0,t}) \leq 4\lambda_1(\mathbf{V}_0(P)) < \infty. \tag{23}$$

Now, consider a sequence $\{(\beta_{1,t}, \mathbf{V}_{0,t})\}$, such that $\|\beta_{1,t}\| \leq M$ and $\mathbf{V}_{0,t}$ satisfies (23). Then the sequence $\{(\beta_{1,t}, \mathbf{V}_{0,t})\}$ lies in a compact set, so it has a convergent subsequence $(\beta_{1,t_j}, \mathbf{V}_{0,t_j}) \to (\beta_{1,L}, \mathbf{V}_0(P))$. According to (20), it follows that

$$\lim_{j \to \infty} R_{t_j}(\beta_{1,t_j}) = \lim_{j \to \infty} \int \rho_1\big(d(\mathbf{s}, \beta_{1,t_j}, \mathbf{V}_{0,t_j})\big) \, dP_{t_j}(\mathbf{s})$$
$$= \int \rho_1(d(\mathbf{s}, \beta_{1,L}, \mathbf{V}_0(P))) \, dP(\mathbf{s}) = R_P(\beta_{1,L}).$$

Now, suppose that $\beta_{1,L} \neq \mathbf{0}$. Then, since $R_P(\beta)$ is uniquely minimized at $\beta = \mathbf{0}$, this would mean that there exists $\eta > 0$, such that together with (22),

$$R_{t_j}(\beta_{1,t_j}) > R_P(\beta_{1,l}) + 2\eta \geq R_P(\mathbf{0}) + 2\eta \geq R_{t_j}(\mathbf{0}) + \eta > R_{t_j}(\mathbf{0}),$$

for $t_j$ sufficiently large, This would mean that $\beta_{1,t_j}$ is not the minimizer of $R_{t_j}(\beta)$. We conclude that $\beta_{1,L} = \mathbf{0}$, which proves the theorem. $\square$

There are several examples of covariance functionals that satisfy (19), such as the Minimum Covariance Determinant functional (see Cator and Lopuhaä (2012)) and the covariance S-functional (see Lopuhaä (1989)), including the Minimum Volume Ellipsoid functional. For a structured covariance functional $\mathbf{V}(\theta_0(P))$ to satisfy (19), it is required that the mapping $\theta \mapsto \mathbf{V}(\theta)$ is continuous. This is true for all the examples mentioned in Section 2. In addition, the functional $\theta_0(P)$ needs to be continuous. An example is the S-functional $\theta_0(P)$ defined in Lopuhaä et al. (2022).

A direct corollary of $\beta_1(P)$ being continuous, is the consistency of the estimator $\beta_{1,n}$.

**Corollary 2.** *Suppose that $\rho_1 : \mathbb{R} \to [0, \infty)$ satisfies (R-BND) and suppose that $P$ is such that (A) holds and that **X** has full rank with probability one. Suppose that $\mathbf{V}_{0,n} \to \mathbf{V}_0(P)$, with probability one. Then for $n$ sufficiently large, there is at least one $\beta_{1,n}$ that minimizes $R_n(\beta)$, with probability one. If $\beta_1(P)$ is the unique minimizer of $R_P(\beta)$, then for any sequence $\beta_{1,n}, n = 1, 2 \ldots$, it holds that*

$$\lim_{n \to \infty} \beta_{1,n} = \beta_1(P),$$

*with probability one.*

**Proof.** We apply Theorem 3 to the sequence $\mathbb{P}_n, n = 1, 2, \ldots$, of probability measures, where $\mathbb{P}_n$ is the empirical measure corresponding to $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$. According to the Portmanteau Theorem (e.g., see Theorem 2.1 in Billingsley (1968)), $\mathbb{P}_n$ converges weakly to $P$, with probability one. The corollary then follows from Theorem 3. $\square$

For unbounded $\rho_1$, we cannot obtain continuity of the functional $\beta_1(P)$, for all sequences $\{P_t\}$ that converge weakly to $P$. However, we can establish strong consistency for the estimator $\beta_{1,n}$.

**Theorem 4.** *Let $\rho_1 : \mathbb{R} \to [0, \infty)$ satisfy (R-UNB). Suppose that $\mathbb{E}_P \|\mathbf{s}\| < \infty$ and that **X** has full rank with probability one. Suppose that $\mathbf{V}_{0,n} \to \mathbf{V}_0(P)$, with probability one, and let $\beta_{1,n}$ minimize $R_n(\beta)$. If $\beta_1(P)$ is the unique minimizer of $R_P(\beta)$, then*

$$\lim_{n \to \infty} \beta_{1,n} = \beta_1(P),$$

*with probability one.*

**Proof.** For the sake of brevity, write $\mathbf{V}_0$ instead of $\mathbf{V}_0(P)$. Since $\mathbf{V}_{0,n} \to \mathbf{V}_0$, with probability one, there exists $0 < L_1 = \lambda_k(\mathbf{V}_0)/4 \leq 4\lambda_1(\mathbf{V}_0) = L_2 < \infty$, such that, for $n$ sufficiently large, all eigenvalues of $\mathbf{V}_{0,n}$ are between $L_1$ and $L_2$ with probability one. Let $h(\mathbf{s}; \beta, \mathbf{V}) = \rho_1(\|\mathbf{V}^{-1/2}(\mathbf{y} - \mathbf{X}\beta)\|)$ and define

$$H(\beta, \mathbf{V}) = \int h(\mathbf{s}; \beta, \mathbf{V}) \, dP(\mathbf{s}),$$

$$H_n(\boldsymbol{\beta}, \mathbf{V}) = \int h(\mathbf{s}; \boldsymbol{\beta}, \mathbf{V}) \, d\mathbb{P}_n(\mathbf{s}).$$

For $M > 0$, consider the class of functions $\mathcal{F} = \{h(\cdot; \boldsymbol{\beta}, \mathbf{V}) : \|\boldsymbol{\beta}\| \leq M; \lambda_k(\mathbf{V}) \geq L_1\}$. Then, according to (13) and (14), the class $\mathcal{F}$ has envelope

$$\rho_1(s_0) + (\|\mathbf{y}\| + M\|\mathbf{X}\|)a_1^{-1/2}\rho_1'(s_0),$$

which is integrable, due to $\mathbb{E}_P\|\mathbf{s}\| < \infty$. Hence, by dominated convergence, $H(\boldsymbol{\beta}, \mathbf{V})$ is continuous on the set $K_M = \{(\boldsymbol{\beta}, \mathbf{V}) : \|\boldsymbol{\beta}\| \leq M; \lambda_k(\mathbf{V}) \geq L_1\}$.

Moreover, the graphs of functions in $\mathcal{F}$ have polynomial discrimination. This can be shown similar to the proof of Lemma B.6 in Lopuhaä et al. (2022). From Theorem 24 in Pollard (1984), we may then conclude

$$\sup_{(\boldsymbol{\beta}, \mathbf{V}) \in K_M} |H_n(\boldsymbol{\beta}, \mathbf{V}) - H(\boldsymbol{\beta}, \mathbf{V})| \to 0, \tag{24}$$

with probability one. As a first consequence, we find that

$$|R_P(\mathbf{0}) - R_n(\mathbf{0})| \leq |H_n(\mathbf{0}, \mathbf{V}_{0,n}) - H(\mathbf{0}, \mathbf{V}_{0,n})| + |H(\mathbf{0}, \mathbf{V}_{0,n}) - H(\mathbf{0}, \mathbf{V}_0)| \to 0, \tag{25}$$

with probability one, due to (24) and continuity of $H(\boldsymbol{\beta}, \mathbf{V})$. Next, we argue there exists $M > 0$, such that for $n$ sufficiently large $\|\boldsymbol{\beta}_{1,n}\| \leq M$. Since $\mathbb{E}_P\|\mathbf{s}\| < \infty$, as in the proof of Theorem 2(i), this ensures that

$$R_P(\mathbf{0}) = \int \rho_1\left(\|\mathbf{V}_0^{-1/2}\mathbf{y}\|\right) dP(\mathbf{s}) < \infty.$$

Next, consider the set $A_M$ defined in (16) and choose $M > 0$, such that

$$\rho_1(\sqrt{M}/(2\sqrt{\lambda_1}))P(A_M) > R_P(\mathbf{0}).$$

Then, for $n$ sufficiently large, we must have $\|\boldsymbol{\beta}_{1,n}\| \leq M$, since otherwise, according to (15),

$$R_n(\boldsymbol{\beta}_{1,n}) \geq \rho_1(\sqrt{M}/(2\sqrt{\lambda_1}))\mathbb{P}_n(A_M) \to \rho_1(\sqrt{M}/(2\sqrt{\lambda_1}))P(A_M) > R_P(\mathbf{0}),$$

as $n \to \infty$, with probability one, which would imply that for $n$ sufficiently large, $R_n(\boldsymbol{\beta}_{1,n}) > R_n(\mathbf{0})$, with probability one.

Now suppose that $\boldsymbol{\beta}_1(P)$ is the unique minimizer of $R_P(\boldsymbol{\beta})$. Because $\boldsymbol{\beta}_1(P)$ is regression equivariant, we may assume that $\boldsymbol{\beta}_1(P) = \mathbf{0}$. This means that for any $\delta > 0$, there exist $\alpha > 0$, such that

$$\inf_{\|\boldsymbol{\beta}\| > \delta} \int \rho_1\left(\frac{\|\mathbf{V}_0^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|}{1 + \alpha}\right) dP(\mathbf{s}) > R_P(\mathbf{0}).$$

Because $\mathbf{V}_{0,n}^{-1/2}\mathbf{V}_0^{1/2} \to \mathbf{I}_k$, with probability one, we can choose $n$ sufficiently large such that $\lambda_k(\mathbf{V}_{0,n}^{-1/2}\mathbf{V}_0^{1/2}) \geq 1/(1+\alpha)^2$. Then, since

$$\|\mathbf{V}_{0,n}^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2 = \|\mathbf{V}_{0,n}^{-1/2}\mathbf{V}_0^{1/2}\mathbf{V}_0^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2$$

$$\geq \lambda_k(\mathbf{V}_{0,n}^{-1/2}\mathbf{V}_0^{1/2})\|\mathbf{V}_0^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2 \geq \frac{\|\mathbf{V}_0^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2}{(1+\alpha)^2},$$

we find

$$\inf_{\|\boldsymbol{\beta}\| > \delta} H_n(\boldsymbol{\beta}, \mathbf{V}_{0,n}) \geq \inf_{\|\boldsymbol{\beta}\| > \delta} \int \rho_1\left(\frac{\|\mathbf{V}_0^{-1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|}{1 + \alpha}\right) dP(\mathbf{s}) > R_P(\mathbf{0}).$$

Furthermore, $|R_P(\mathbf{0}) - H_n(\mathbf{0}, \mathbf{V}_{0,n})| \to 0$, with probability one, as $n \to \infty$, according to (25). Hence, for $n$ sufficiently large, we would find that for all $\delta > 0$,

$$\inf_{\|\boldsymbol{\beta}\| > \delta} H_n(\boldsymbol{\beta}, \mathbf{V}_{0,n}) > H_n(\mathbf{0}, \mathbf{V}_{0,n}).$$

Therefore, for all $\delta > 0$, we must have $\|\boldsymbol{\beta}_{1,n}\| \leq \delta$, for $n$ sufficiently large, with probability one. This means $\boldsymbol{\beta}_{1,n} \to \mathbf{0}$, with probability one. □

Asymptotic properties of the MM-estimator for linear mixed effects models in Copt and Heritier (2007) was only considered for the simple model with a fixed design matrix and normal errors. As a consequence of their Theorem 1, the MM-estimator would be weakly consistent. However, the requirement that the covariance estimator used in the first step is consistent, seems to be missing. Our Corollary 2 and Theorem 4 establish consistency for the MM-estimator for a larger class of linear mixed effects models and under very mild conditions on the distribution $P$. Theorem 4 is equivalent to Theorem 3.1 in Lopuhaä (1992) for multivariate location and scatter. Corollary 2 extends this result to bounded $\rho$-functions. Furthermore, Corollary 2 is obtained under conditions on the distribution $P$, that are much weaker than the ones for the MM-estimators considered in Salibián-Barrera et al. (2006) and Kudraszow and Maronna (2011), which restrict themselves to distributions with an elliptically contoured density.

## 6. Global robustness: breakdown point

Consider a collection of points $\mathcal{S}_n = \{\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i), i = 1, \ldots, n\} \subset \mathbb{R}^k \times \mathbb{R}^{kq}$. To emphasize the dependence on the collection $\mathcal{S}_n$, we sometimes denote the estimators in Definition 1 by $\beta_{1,n}(\mathcal{S}_n)$ and $\mathbf{V}_{0,n}(\mathcal{S}_n)$. To investigate the global robustness of $\beta_{1,n}$, we compute that finite-sample (replacement) breakdown point. For a given collection $\mathcal{S}_n$ the finite-sample breakdown point (see Donoho and Huber (1983)) of regression estimator $\beta_{1,n}$ is defined as the smallest proportion of points from $\mathcal{S}_n$ that one needs to replace in order to carry the estimator over all bounds. More precisely,

$$\epsilon_n^*(\beta_{1,n}, \mathcal{S}_n) = \min_{1 \le m \le n} \left\{ \frac{m}{n} : \sup_{\mathcal{S}_m'} \left\| \beta_{1,n}(\mathcal{S}_n) - \beta_{1,n}(\mathcal{S}_m') \right\| = \infty \right\}, \tag{26}$$

where the minimum runs over all possible collections $\mathcal{S}_m'$ that can be obtained from $\mathcal{S}_n$ by replacing $m$ points of $\mathcal{S}_n$ by arbitrary points in $\mathbb{R}^k \times \mathbb{R}^{kq}$. The finite sample (replacement) breakdown point of a covariance estimator $\mathbf{V}_{0,n}$ at a collection $\mathcal{S}_n$, is defined as

$$\epsilon_n^*(\mathbf{V}_{0,n}, \mathcal{S}_n) = \min_{1 \le m \le n} \left\{ \frac{m}{n} : \sup_{\mathcal{S}_m'} \text{dist}(\mathbf{V}_{0,n}(\mathcal{S}_n), \mathbf{V}_{0,n}(\mathcal{S}_m')) = \infty \right\}, \tag{27}$$

with $\text{dist}(\cdot, \cdot)$ defined as $\text{dist}(\mathbf{A}, \mathbf{B}) = \max \left\{ |\lambda_1(\mathbf{A}) - \lambda_1(\mathbf{B})|, \left| \lambda_k(\mathbf{A})^{-1} - \lambda_k(\mathbf{B})^{-1} \right| \right\}$, where the minimum runs over all possible collections $\mathcal{S}_m'$ that can be obtained from $\mathcal{S}_n$ by replacing $m$ points of $\mathcal{S}_n$ by arbitrary points in $\mathbb{R}^k \times \mathbb{R}^{kq}$. So the breakdown point of $\mathbf{V}_{0,n}$ is the smallest proportion of points from $\mathcal{S}_n$ that one needs to replace in order to make the largest eigenvalue of $\mathbf{V}_{0,n}(\mathcal{S}_m')$ arbitrarily large (explosion), or to make the smallest eigenvalue of $\mathbf{V}_{0,n}(\mathcal{S}_m')$ arbitrarily small (implosion). When we estimate a structured covariance matrix $\mathbf{V}(\theta)$ by $\mathbf{V}(\theta_{0,n})$, we need to specify what the breakdown point of $\theta_{0,n}$ is. Since, the estimator $\theta_{0,n}$ determines the covariance estimator $\mathbf{V}(\theta_{0,n})$, it seems natural to let the breakdown point of $\theta_{0,n}$ correspond to the breakdown point of the covariance estimator (see also Lopuhaä et al. (2022)).

The breakdown behavior of $\beta_{1,n}$ depends on whether the function $\rho_1$ in Definition 1 is bounded or unbounded. Huber (1984) pointed put that the breakdown point of location M-estimators constructed with a bounded $\rho$-function not only depends on the function $\rho$, but also on the configuration of the sample. Depending on the configuration of the sample, the breakdown point can be any value between 0 and $1/2$. For this reason, the multivariate location M-estimator (see Lopuhaä (1992)) is constructed with an unbounded $\rho$-function. In this way, the location M-estimator inherits the breakdown point of the initial covariance estimator.

Unfortunately, the use of an unbounded function $\rho_1$ in Definition 1, does not seem suitable for the breakdown behavior of the regression MM-estimator. The presence of the design matrices $\mathbf{X}_i$ makes things more complicated than in the multivariate location case. Nevertheless, for unbounded $\rho_1$, we can establish a result similar to the one in Lopuhaä (1992), when all design matrices are the same. An example is of course when all $\mathbf{X}_i = \mathbf{I}_k$ as in the location-scale model, but another example occurs in linear mixed effects models for which all subjects have the same the design matrix $\mathbf{X}$ representing particular contrasts for the fixed effects.

**Proposition 1.** *Suppose that $\rho_1$ satisfies (R-UNB). Let $\mathcal{S}_n \subset \mathbb{R}^p$ be a collection of $n$ points $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$, $i = 1, \ldots, n$. Suppose that $\mathbf{X}_i = \mathbf{X}$, for all $i = 1, \ldots, n$, where $\mathbf{X}$ is fixed and has full rank. Then for any $\beta_{1,n}$ that minimizes $R_n(\beta)$, it holds that*

$$\epsilon_n^*(\beta_{1,n}, \mathcal{S}_n) \ge \epsilon_n^*(\mathbf{V}_{0,n}, \mathcal{S}_n).$$

**Proof.** For $\mathbf{t} \in \mathbb{R}^k$, let

$$\widetilde{R}_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \left\{ \rho_1 \left( \sqrt{(\mathbf{y}_i - \mathbf{t})^T \mathbf{V}_{0,n}^{-1} (\mathbf{y}_i - \mathbf{t})} \right) - \rho_1 \left( \sqrt{\mathbf{y}_i^T \mathbf{V}_{0,n}^{-1} \mathbf{y}_i} \right) \right\},$$

be the object function for the location M-estimator in Lopuhaä (1992). Then we can write

$$R_n(\beta) = R_n(\mathbf{0}) + \widetilde{R}_n(\mathbf{X}\beta).$$

Because $\mathbf{X}$ has full rank, $\beta_{1,n}$ minimizes $R_n(\mathbf{b})$ if and only if $\mathbf{t}_{1,n} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \beta_{1,n}$ minimizes $\widetilde{R}_n(\mathbf{t})$. As $\mathbf{X}$ is considered to be fixed, this means that $\beta_{1,n}$ breaks down precisely when $\mathbf{t}_{1,n}$ does. Hence from Theorem 4.1 in Lopuhaä (1992) we conclude that $\epsilon_n^*(\beta_{1,n}, \mathcal{S}_n) \ge \epsilon_n^*(\mathbf{V}_{0,n}, \mathcal{S}_n)$. □

The use of a bounded function $\rho_1$ in Definition 1 also does not seem very suitable, in view of the results found by Huber (1984). However, the approach followed by Yohai (1987), which relates the bounded function $\rho_1$ to another bounded function $\rho_0$ in the first stage, turns out to be adequate. Let $\beta_{0,n}$ be an initial regression estimate, such that together with the covariance estimate $\mathbf{V}_{0,n}$, it holds that

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \sqrt{(\mathbf{y}_i - \mathbf{X}_i \beta_{0,n})^T \mathbf{V}_{0,n}^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta_{0,n})} \right) = b_0, \tag{28}$$

for a function $\rho_0$ that satisfies (R-BND) and suppose that $\rho_1$ satisfies (R-BND), such that

$$\frac{\rho_1(s)}{a_1} \le \frac{\rho_0(s)}{a_0}. \tag{29}$$

Next, we proceed as in Definition 1, i.e., define $\beta_{1,n}$ as the vector that minimizes (7). Estimates $\beta_{0,n}$ and $\mathbf{V}_{0,n}$ can be any two (high breakdown) estimators satisfying (28). However, natural candidates for our setup are the S-estimates $(\beta_{0,n}, \theta_{0,n})$, with $\mathbf{V}_{0,n} = \mathbf{V}(\theta_{0,n})$, defined in Lopuhaä et al. (2022) by means of the function $\rho_0$.

In order to formulate the breakdown point of $\beta_{1,n}$ using bounded $\rho$-functions, we first need to discuss the following. Recall that $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$ are represented as points in $\mathbb{R}^k \times \mathbb{R}^{kq}$. Note however, that for linear models with intercept the first column of each $\mathbf{X}_i$ consists of 1's. This means that the points $(\mathbf{y}_i, \mathbf{X}_i)$ are concentrated in a lower dimensional subset of $\mathbb{R}^k \times \mathbb{R}^{kq}$. A similar situation occurs when all $\mathbf{X}_i$ are equal to the same design matrix. In view of this, define $\mathcal{X} \subset \mathbb{R}^{kq}$ as the subset with the lowest dimension $p = \dim(\mathcal{X}) \leq kq$ satisfying $P(\mathbf{X} \in \mathcal{X}) = 1$. Hence, $P$ is concentrated on the subset $\mathbb{R}^k \times \mathcal{X}$ of $\mathbb{R}^k \times \mathbb{R}^{kq}$, which may be of lower dimension $k + p$ than $k + kq$. Let $\mathcal{S}_n = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, with $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$ be a collection of $n$ points in $\mathbb{R}^k \times \mathcal{X}$. Define

$$\kappa(\mathcal{S}_n) = \text{maximal number of points of } \mathcal{S}_n \text{ lying on the same hyperplane in } \mathbb{R}^k \times \mathcal{X}. \tag{30}$$

For example, if the distribution $P$ is absolutely continuous, then $\kappa(\mathcal{S}_n) \leq k + p$ with probability one. We then have the following theorem.

**Theorem 5.** *Suppose that $(\beta_{0,n}, \mathbf{V}_{0,n})$ satisfy (28), with $\rho_0$ such that (29) holds, where $\rho_0$ and $\rho_1$ satisfy (R-BND). Let $\mathcal{S}_n \subset \mathbb{R}^k \times \mathcal{X}$ be a collection of $n$ points $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$, $i = 1, \ldots, n$. Let $r_0 = b_0/a_0$ and suppose that $0 < r_0 \leq (n - k(\mathcal{S}_n))/(2n)$, where $k(\mathcal{S}_n)$ is defined by (30). Then for any $\beta_{1,n}$ that minimizes $R_n(\beta)$, it holds that*

$$\epsilon_n^*(\beta_{1,n}, \mathcal{S}_n) \geq \min\left(\epsilon_n^*(\mathbf{V}_{0,n}, \mathcal{S}_n), \frac{\lceil nr_0 \rceil}{n}\right).$$

**Proof.** Suppose we replace $m$ points, where $m$ is such that

$$m \leq \lceil nr_0 \rceil - 1 \quad \text{and} \quad m \leq n\epsilon_n^*(\mathbf{V}_{0,n}, \mathcal{S}_n) - 1.$$

Let $\mathcal{S}_m'$ be the corrupted collection of points. Write $\beta_{0,m} = \beta_{0,n}(\mathcal{S}_m')$, and $\mathbf{V}_{0,m} = \mathbf{V}_{0,n}(\mathcal{S}_m')$. Then $\mathbf{V}_{0,m}$ does not break down, so that there exist constants $0 < L_1 \leq L_2 < \infty$, not depending on $\mathcal{S}_m'$ such that $0 < L_1 \leq \lambda_k(\mathbf{V}_{0,m}) \leq \lambda_1(\mathbf{V}_{0,m}) \leq L_2 < \infty$. For any $\beta \in \mathbb{R}^q$, define the cylinder $\mathcal{C}_{1,m}(\beta) = \left\{(\mathbf{y}, \mathbf{X}) \in (\mathbb{R})^k \times (\mathbb{R})^p : (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}_{0,m}^{-1}(\mathbf{y} - \mathbf{X}\beta) \leq c_1^2\right\}$. Consider the function

$$R_m(\beta) = \frac{1}{n} \sum_{\mathbf{s}_i \in \mathcal{S}_m'} \rho_1\left(\sqrt{(\mathbf{y}_i - \mathbf{X}_i\beta)^T \mathbf{V}_{0,m}^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)}\right),$$

for the corrupted sample $\mathcal{S}_m'$. For any $\beta$ that minimizes $R_m(\beta)$, it holds $R_m(\beta) \leq R_m(\beta_{0,m})$. Therefore, for such $\beta$, according to (29) and (28), we have that

$$R_m(\beta) \leq \frac{1}{n} \sum_{\mathbf{s}_i \in \mathcal{S}_m'} \rho_1\left(\sqrt{(\mathbf{y}_i - \mathbf{X}_i\beta_{0,m})^T \mathbf{V}_{0,m}^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta_{0,m})}\right)$$

$$\leq \frac{a_1}{a_0} \frac{1}{n} \sum_{\mathbf{s}_i \in \mathcal{S}_m'} \rho_0\left(\sqrt{(\mathbf{y}_i - \mathbf{X}_i\beta_{0,m})^T \mathbf{V}_{0,m}^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta_{0,m})}\right) = r_0 a_1.$$

Let $\mathbb{P}_m'$ be the empirical measure corresponding to the corrupted collection $\mathcal{S}_m'$. Then it holds that

$$\mathbb{P}_m'(\mathcal{C}_{1,m}(\beta)) = \frac{1}{n} \sum_{\mathbf{s}_i \in \mathcal{S}_m'} \mathbb{1}\{\mathbf{s}_i \in \mathcal{C}_{1,m}(\beta)\}$$

$$\geq 1 - \frac{1}{na_1} \sum_{\mathbf{s}_i \in \mathcal{S}_m'} \rho_1\left(\sqrt{(\mathbf{y}_i - \mathbf{X}_i\beta)^T \mathbf{V}_{0,m}^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)}\right) \geq 1 - r_0.$$

It follows that the cylinder $\mathcal{C}_{1,m}(\beta)$ must contain at least $\lceil n - nr_0 \rceil$ number of points from the corrupted collection $\mathcal{S}_m'$. Furthermore, since $r_0 \leq (n - k(\mathcal{S}_n))/(2n)$, for any such subset of $\mathcal{S}_m'$ it holds that it contains $\lceil n - nr_0 \rceil - m = n - \lfloor nr_0 \rfloor - \lceil nr_0 \rceil + 1 \geq k(\mathcal{S}_n) + 1$ points of the original collection $\mathcal{S}_n$. Let $J_0$ be a subset of $k(\mathcal{S}_n) + 1$ points from the original collection $\mathcal{S}_n$ contained in $\mathcal{C}_{1,m}(\beta)$. By definition, $k(\mathcal{S}_n) + 1$ original points cannot be on the same hyperplane, so that

$$\gamma_n = \inf_{J \subset \mathcal{S}_n} \inf_{\|\gamma\|=1} \max_{\mathbf{s} \in J} \|\mathbf{X}\gamma\| > 0.$$

where the first infimum runs over all subsets $J \subset \mathcal{S}_n$ of $k(\mathcal{S}_n) + 1$ points. By definition of $\gamma_n$, there exists an original point $\mathbf{s}_0 \in J_0 \subset \mathcal{S}_n \cap \mathcal{C}_{1,m}(\beta)$, such that

$$\|\beta\| = \|\mathbf{X}_0\beta\| \times \frac{\|\beta\|}{\|\mathbf{X}_0\beta\|} \leq \frac{1}{\gamma_n} \|\mathbf{X}_0\beta\|.$$

Because $\mathbf{s}_0 \in \mathcal{C}_{1,m}(\beta)$, it follows that

$$\|\mathbf{y}_0 - \mathbf{X}_0\beta\|^2 \leq (\mathbf{y}_0 - \mathbf{X}_0\beta)^T \mathbf{V}_{0,n}^{-1}(\mathbf{y}_0 - \mathbf{X}_0\beta)\lambda_1(\mathbf{V}_{0,m}) \leq c_1^2 L_2,$$

and because $\mathbf{s}_0 \in \mathcal{S}_n$, we have that

$$\|\mathbf{X}_0\boldsymbol{\beta}\| \leq c_1\sqrt{L_2} + \max_{(\mathbf{y}_i,\mathbf{X}_i)\in\mathcal{S}_n}\|\mathbf{y}_i\| < \infty.$$

We conclude that for minimizing $R_m(\boldsymbol{\beta})$ we can restrict ourselves to a compact set $K_n$, only depending on the original collection $\mathcal{S}_n$. Firstly, since $R_m(\boldsymbol{\beta})$ is continuous, this implies there exists at least one $\boldsymbol{\beta}_{1,n}(\mathcal{S}'_m)$, which minimizes $R_m(\boldsymbol{\beta})$. Secondly, since any $\boldsymbol{\beta}_{1,n}(\mathcal{S}'_m)$ must be in $K_n$, which only depends on the original collection $\mathcal{S}_n$, the estimate $\boldsymbol{\beta}_{1,n}(\mathcal{S}'_m)$ does not break down. □

Theorem 5 is comparable to Theorem 1 in Salibián-Barrera et al. (2006) and Theorem 3 in Kudraszow and Maronna (2011) for MM-estimators for multivariate location and scatter and for multivariate linear regression, respectively. The breakdown point for MM-estimators for linear mixed effects models has only been discussed in Copt and Heritier (2007). They conjecture that the exact value can be derived using the technique in Van Aelst and Willems (2005), but do not pursue a rigorous derivation. Together with Proposition 1, the result in Theorem 5 provides sufficient conditions for the MM-estimators in the linear mixed effects model used in Copt and Heritier (2007), to inherit the breakdown point from the initial covariance estimate.

It can be shown that if $\mathbf{V}_{0,n}$ satisfies (28) for some $\boldsymbol{\beta}_{0,n}$, it must have a breakdown point that is less than or equal to $\lceil nr_0\rceil/n$ (e.g., see the proof of Theorem 4 in Lopuhaä et al. (2022)). This means that from Theorem 5, we have $\epsilon_n^*(\boldsymbol{\beta}_{1,n},\mathcal{S}_n) \geq \epsilon_n^*(\mathbf{V}_{0,n})$. Moreover, if $(\boldsymbol{\beta}_{0,n},\theta_{0,n})$ are S-estimators, as defined in Lopuhaä et al. (2022), such that $\mathbf{V}_{0,n} = \mathbf{V}(\theta_{0,n})$, then $\epsilon_n^*(\mathbf{V}_{0,n}) = \lceil nr_0\rceil/n$ according to Theorem 4 in Lopuhaä et al. (2022), so that

$$\epsilon_n^*(\boldsymbol{\beta}_{1,n},\mathcal{S}_n) \geq \frac{\lceil nr_0\rceil}{n}.$$

The largest possible value of the breakdown point occurs when $r_0 = (n - \kappa(\mathcal{S}_n))/(2n)$, in which case $\lceil nr_0\rceil/n = \lceil(n - \kappa(\mathcal{S}_n))/2\rceil/n = \lfloor(n - \kappa(\mathcal{S}_n) + 1)/2\rfloor/n$. When the collection $\mathcal{S}_n$ is in general position, then $\kappa(\mathcal{S}_n) = k + p$. In that case the breakdown point is at least $\lfloor(n - k - p + 1)/2\rfloor/n$. When all $\mathbf{X}_i$ are equal to the same $\mathbf{X}$, one has $p = 0$ and $\kappa(\mathcal{S}_n) = k$. In that case, the breakdown point is at least $\lfloor(n - k + 1)/2\rfloor/n$. This coincides with the maximal breakdown point for affine equivariant estimators for $k \times k$ covariance matrices (see Davies (1987)).

## 7. Score equations

Recall the definition of the functional $\boldsymbol{\beta}_1(P)$ in Section 3, which minimizes $R_P(\boldsymbol{\beta})$. Then $\boldsymbol{\beta}_1(P)$ is also a solution of $\partial R_P(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0}$. In order to allow changing the order of integration and differentiation in $R_P(\boldsymbol{\beta})$, we require an additional condition on $\rho_1$.

(R-CD1) $\rho_1$ is continuously differentiable and $u_1(s) = \rho_1'(s)/s$ is continuous,

If $\rho_1$ satisfies (R-CD1), then

$$\frac{\partial}{\partial\boldsymbol{\beta}}\rho_1\left(\sqrt{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}\right) = u_1(d_0)\mathbf{X}^T\mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

where $d_0 = d(\mathbf{s}, \boldsymbol{\beta}, \mathbf{V}_0(P))$, as defined in (21). This means that

$$\left\|\frac{\partial}{\partial\boldsymbol{\beta}}\rho_1\left(\sqrt{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}\right)\right\| = \frac{|u_1(d_0)d_0| \cdot \|\mathbf{X}\|}{\sqrt{\lambda_k(\mathbf{V}_0(P))}}.$$

When $\rho_1$ satisfies either (R-BND) or (R-UNB), the function $u_1(s)s$ is uniformly bounded. This means that in both cases the right hand side is bounded by a constant times $\|\mathbf{X}\|$. Hence, if $\mathbb{E}_P\|\mathbf{X}\| < \infty$, then by dominated convergence

$$\frac{\partial}{\partial\boldsymbol{\beta}}\int\rho_1\left(\sqrt{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}\right)\mathrm{d}P(\mathbf{y}, \mathbf{X})$$
$$= \int\frac{\partial}{\partial\boldsymbol{\beta}}\rho_1\left(\sqrt{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}\right)\mathrm{d}P(\mathbf{y}, \mathbf{X})$$
$$= \int u_1(d_0)\mathbf{X}^T\mathbf{V}_0(P)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\,\mathrm{d}P(\mathbf{y}, \mathbf{X}).$$

We conclude that, if $\mathbb{E}_P\|\mathbf{X}\| < \infty$, the functional $\boldsymbol{\beta}_1(P)$ satisfies score equation

$$\int\Psi_1(\mathbf{s}, \boldsymbol{\beta}, \mathbf{V}_0(P))\,\mathrm{d}P(\mathbf{y}, \mathbf{X}) = \mathbf{0}, \tag{31}$$

where

$$\Psi_1(\mathbf{s}, \boldsymbol{\beta}, \mathbf{V}) = u_1(d)\mathbf{X}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{32}$$

with $d = d(\mathbf{s}, \boldsymbol{\beta}, \mathbf{V})$, as defined in (21).

Score equation (31) coincides with equation (3.8) in Lopuhaä (1992) for the multivariate location-scale model. If $P$ is the empirical measure $\mathbb{P}_n$ corresponding to $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$, then (31) coincides with equation (2.6) for the multiple regression model in Yohai (1987) and with equation (8) for the linear mixed effects model (2) in Copt and Heritier (2007). Furthermore, for the empirical measure $\mathbb{P}_n$, equation (31) is also similar to equation (16) for the location MM-estimator in Salibián-Barrera et al. (2006) and to equation (2.10) for the multivariate linear regression MM-estimator in Kudraszow and Maronna (2011).

Let

$$\Lambda_1(\beta, \mathbf{V}) = \int \Psi_1(\mathbf{s}, \beta, \mathbf{V}) \, dP(\mathbf{s}). \tag{33}$$

A vector $\mathbf{b} \in \mathbb{R}^q$ is called a *point of symmetry* of $P$, if for almost all $\mathbf{X}$, it holds that

$$P(\mathbf{Xb} + A \mid \mathbf{X}) = P(\mathbf{Xb} - A \mid \mathbf{X}),$$

for all measurable sets $A \subset \mathbb{R}^k$, where for $\lambda \in \mathbb{R}$ and $\mathbf{b} \in \mathbb{R}^q$, $\mathbf{Xb} + \lambda A$ denotes the set $\{\mathbf{X} + \lambda \mathbf{y} : \mathbf{y} \in A\}$. If $\mathbf{b}$ is a point of symmetry of $P$, it has the property that

$$\Lambda_1(\mathbf{b}, \mathbf{V}) = \mathbf{0} \quad \text{for all non-singular symmetric matrices } \mathbf{V}. \tag{34}$$

This will become very useful in determining asymptotic properties, such as the influence function for $\beta_1(P)$ and asymptotic normality of $\beta_{1,n}$. Note that if $P$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density as defined in (9), then the vector $\mu = \mathbf{X}\beta$ is a point of symmetry.

## 8. Local robustness: the influence function

For $0 < \epsilon < 1$ and $\mathbf{s}_0 = (\mathbf{y}_0, \mathbf{X}_0) \in \mathbb{R}^k \times \mathbb{R}^{kq}$ fixed, consider $P_{\epsilon, \mathbf{s}_0} = (1 - \epsilon)P + \epsilon \delta_{\mathbf{s}_0}$, as defined in (18). The *influence function* of the functional $\beta_1(\cdot)$ at probability measure $P$, is defined as

$$\mathrm{IF}(\mathbf{s}_0 \beta_1, P) = \lim_{\epsilon \downarrow 0} \frac{\beta_1((1 - \epsilon)P + \epsilon \delta_{\mathbf{s}_0}) - \beta_1(P)}{\epsilon}, \tag{35}$$

if this limit exists (see Hampel (1974)).

We intend to include both bounded and unbounded functions $\rho_1$ in Definition 2. For bounded $\rho_1$, it follows from Theorem 3 that, under suitable conditions, the functional $\beta_1(P)$ is continuous. In particular, this means that

$$\lim_{\epsilon \downarrow 0} \beta_1(P_{\epsilon, \mathbf{s}_0}) = \beta_1(P). \tag{36}$$

For unbounded $\rho_1$, the functional $\beta_1(P)$ is not necessarily continuous, but we can still establish (36).

**Lemma 1.** *Let $\rho_1 : \mathbb{R} \to [0, \infty)$ satisfy (R-UNB). Suppose that $\mathbb{E}_P \|\mathbf{s}\| < \infty$ and that $\mathbf{X}$ has full rank with probability one. Suppose that $\mathbf{V}_0(P_{\epsilon, \mathbf{s}_0})$ exists and that $\mathbf{V}_0(P_{\epsilon, \mathbf{s}_0}) \to \mathbf{V}_0(P)$, as $\epsilon \downarrow 0$. Suppose that $\beta_1(P_{\epsilon, \mathbf{s}_0})$ minimizes $R_{P_{\epsilon, \mathbf{s}_0}}(\beta)$. If $\beta_1(P)$ is the unique minimizer of $R_P(\beta)$, then*

$$\lim_{\epsilon \downarrow 0} \beta_1(P_{\epsilon, \mathbf{s}_0}) = \beta_1(P).$$

**Proof.** Define the functions $h(\mathbf{s}; \beta, \mathbf{V})$ and $H(\beta, \mathbf{V})$ as in the proof of Theorem 4, and let

$$H_{\epsilon, \mathbf{s}_0}(\beta, \mathbf{V}) = \int h(\mathbf{s}; \beta, \mathbf{V}) \, dP_{\epsilon, \mathbf{s}_0}(\mathbf{s}).$$

Let $K_M$ be the same set of pairs $(\beta, \mathbf{V})$ as in the proof of Theorem 4. Then for $(\beta, \mathbf{V}) \in K_M$,

$$|h(\mathbf{s}; \beta, \mathbf{V})| \le \rho_1(s_0) + (\|\mathbf{y}\| + M\|\mathbf{X}\|)L_1^{-1/2}\rho_1'(s_0).$$

Instead of (24), we now have

$$\sup_{(\beta, \mathbf{V}) \in K_M} |H_{\epsilon, \mathbf{s}_0}(\beta, \mathbf{V}) - H(\beta, \mathbf{V})| \le \epsilon \left( \sup_{(\beta, \mathbf{V}) \in K_M} |H(\beta, \mathbf{V})| + \sup_{(\beta, \mathbf{V}) \in K_M} h(\mathbf{s}_0; \beta, \mathbf{V}) \right).$$

Because $\mathbb{E}_P \|\mathbf{s}\| < \infty$, the first supremum on the right hand side is bounded and similarly, the second supremum is bounded by a constant depending on $\mathbf{s}_0$ and $\rho_1$. Therefore,

$$\sup_{(\beta, \mathbf{V}) \in K_M} |H_{\epsilon, \mathbf{s}_0}(\beta, \mathbf{V}) - H(\beta, \mathbf{V})| \to 0, \quad \text{as } \epsilon \downarrow 0.$$

From here on, one can mimic the proof of Theorem 4 and show that $\beta_1(P_{\epsilon, \mathbf{s}_0}) \to \beta_1(P)$. $\qquad \square$

Now, that we have established (36) for both bounded and unbounded $\rho_1$, we have the following general result for the influence function.

**Theorem 6.** *Suppose that $\rho_1$ either satisfies (R-BND) and (R-CD1) or satisfies (R-UNB), and suppose that $\mathbb{E}_P\|\mathbf{s}\| < \infty$. Let $\beta_1(P_{\epsilon,\mathbf{s}_0})$ and $\beta_1(P)$ be a solution to the minimization problem in Definition 2 at $P_{\epsilon,\mathbf{s}_0}$ and $P$, respectively. Suppose that $(\beta_1(P_{\epsilon,\mathbf{s}_0}), \mathbf{V}_0(P_{\epsilon,\mathbf{s}_0})) \to (\beta_1(P), \mathbf{V}_0(P))$, as $\epsilon \downarrow 0$, and suppose that $\beta_1(P)$ is a point of symmetry of $P$. Suppose that $\Lambda_1$, as defined in (33), has a partial derivative $\partial \Lambda_1/\partial \beta$ that is continuous at $(\beta_1(P), \mathbf{V}_0(P))$ and that $\mathbf{D}_1 = (\partial \Lambda_1/\partial \beta)(\beta_1(P), \mathbf{V}_0(P))$ is non-singular. Then for $\mathbf{s}_0 \in \mathbb{R}^p$,*

$$\text{IF}(\mathbf{s}_0; \beta_1, P) = -\mathbf{D}_1^{-1}\Psi_1(\mathbf{s}_0, \beta_1(P), \mathbf{V}_0(P)),$$

*where $\Psi_1$ is defined in (32).*

**Proof.** Denote

$$\boldsymbol{\xi}_{\epsilon,\mathbf{s}_0} = (\beta_{1,\epsilon,\mathbf{s}_0}, \mathbf{V}_{0,\epsilon,\mathbf{s}_0}) = (\beta_1(P_{\epsilon,\mathbf{s}_0}), \mathbf{V}_0(P_{\epsilon,\mathbf{s}_0}))$$
$$\boldsymbol{\xi}_P = (\beta_{1,P}, \mathbf{V}_{0,P}) = (\beta_1(P), \mathbf{V}_0(P))$$

Since $\mathbf{V}_{0,\epsilon,\mathbf{s}_0} \to \mathbf{V}_{0,P}$ and the fact that $\mathbb{E}_P\|\mathbf{s}\| < \infty$, it follows from Theorem 3 and Lemma 1 that there exists $\beta_1(P_{\epsilon,\mathbf{s}_0})$ minimizing $R_P(\beta)$ at $P = P_{\epsilon,\mathbf{s}_0}$. According to Section 7, this means that $\boldsymbol{\xi}_{\epsilon,\mathbf{s}_0}$ satisfies the score equation (31) for the regression M-functional $\beta_1$ at $P_{\epsilon,\mathbf{s}_0}$, that is

$$\int \Psi_1(\mathbf{s}, \boldsymbol{\xi}_{\epsilon,\mathbf{s}_0})\, dP_{\epsilon,\mathbf{s}_0}(\mathbf{s}) = \mathbf{0}.$$

We decompose as follows

$$\mathbf{0} = \int \Psi_1(\mathbf{s}, \boldsymbol{\xi}_{\epsilon,\mathbf{s}_0})\, dP_{\epsilon,\mathbf{s}_0}(\mathbf{s})$$
$$= (1 - \epsilon)\int \Psi_1(\mathbf{s}, \boldsymbol{\xi}_{\epsilon,\mathbf{s}_0})\, dP(\mathbf{s}) + \epsilon \Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_{\epsilon,\mathbf{s}_0}) = (1 - \epsilon)\Lambda_1(\boldsymbol{\xi}_{\epsilon,\mathbf{s}_0}) + \epsilon \Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_{\epsilon,\mathbf{s}_0}).$$

We first determine the order of $\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}$, as $\epsilon \downarrow 0$. Because $\boldsymbol{\xi} \mapsto \Psi_1(\mathbf{s}_0, \boldsymbol{\xi})$ is continuous, it follows that $\Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_{\epsilon,\mathbf{s}_0}) = \Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_P) + o(1)$, as $\epsilon \downarrow 0$. Furthermore, because $\Lambda_1$ has a partial derivative $\partial \Lambda_1/\partial \beta$ that is continuous at $\boldsymbol{\xi}_P = (\beta_{1,P}, \mathbf{V}_{0,P})$, we have that

$$\Lambda_1(\boldsymbol{\xi}_{\epsilon,\mathbf{s}_0}) = \Lambda_1(\beta_{1,P}, \mathbf{V}_{0,\epsilon,\mathbf{s}_0}) + \frac{\partial \Lambda_1}{\partial \beta}(\beta_{1,P}, \mathbf{V}_{0,\epsilon,\mathbf{s}_0})(\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}) + o(\|\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}\|)$$
$$= \Lambda_1(\beta_{1,P}, \mathbf{V}_{0,\epsilon,\mathbf{s}_0}) + \Big(\mathbf{D}_1 + o(1)\Big)\Big(\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}\Big) + o(\|\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}\|).$$

Since $\beta_{1,P}$ is a point of symmetry of $P$, according to (34) it holds that $\Lambda_1(\beta_{1,P}, \mathbf{V}_{0,\epsilon,\mathbf{s}_0}) = \mathbf{0}$. It follows that

$$\mathbf{0} = (1 - \epsilon)\mathbf{D}_1\Big(\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}\Big) + o(\|\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}\|) + o(\epsilon) + \epsilon \Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_P).$$

Because $\mathbf{D}_1$ is non-singular and $\Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_P)$ is fixed, this implies $\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P} = O(\epsilon)$. After inserting this in the previous equality, it follows that

$$\mathbf{0} = (1 - \epsilon)\mathbf{D}_1\Big(\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}\Big) + \epsilon \Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_P) + o(\epsilon)$$
$$= \mathbf{D}_1\Big(\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}\Big) + \epsilon \Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_P) + o(\epsilon).$$

We conclude

$$\frac{\beta_{1,\epsilon,\mathbf{s}_0} - \beta_{1,P}}{\epsilon} = -\mathbf{D}_1^{-1}\Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_P) + o(1).$$

This means that the limit of the left hand side exists and

$$\text{IF}(\mathbf{s}_0; \beta_1, P) = \lim_{\epsilon \downarrow 0}\frac{\beta_1((1 - \epsilon)P + \epsilon \delta_{\mathbf{s}_0}) - \beta_1(P)}{\epsilon} = -\mathbf{D}_1^{-1}\Psi_1(\mathbf{s}_0, \boldsymbol{\xi}_P).$$

$\square$

When $P$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density (9) we can obtain a more detailed expression for the influence function. This requires the following additional condition on the function $\rho_1$.

(R-CD2) $\rho_1$ is twice continuously differentiable.

We then have the following corollary.

**Corollary 3.** *Suppose that $P$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density $f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ from (9), with $(\mathbf{X}\beta_1(P), \mathbf{V}_0(P)) = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\beta_1(P_{\epsilon,\mathbf{s}_0})$ and $\beta_1(P)$ be a solution to the minimization problem in Definition 2 at $P_{\epsilon,\mathbf{s}_0}$ and $P$, respectively, and suppose that*

$(\beta_1(P_{\epsilon,\mathbf{s}_0}), \mathbf{V}_0(P_{\epsilon,\mathbf{s}_0})) \to (\beta_1(P), \mathbf{V}_0(P))$, *as $\epsilon \downarrow 0$. Suppose that $\rho_1 : \mathbb{R} \to [0, \infty)$ either satisfies (R-BND), (R-CD1) and (R-CD2), or satisfies (R-UNB) and (R-CD2), such that $u_1(s)$ and $u_1'(s)s$ are bounded. Let*

$$\alpha_1 = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[ \left(1 - \frac{1}{k}\right) \frac{\rho_1'(\|\mathbf{z}\|)}{\|\mathbf{z}\|} + \frac{1}{k}\rho_1''(\|\mathbf{z}\|) \right], \tag{37}$$

*and suppose that $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho_1''(\|\mathbf{y}\|)\right] > 0$. If $\mathbf{X}$ has full rank with probability one, then for $\mathbf{s}_0 = (\mathbf{y}_0, \mathbf{X}_0)$ we have*

$$\text{IF}(\mathbf{s}_0, \beta_1, P) = \frac{u_1(d_0)}{\alpha_1}\left( \mathbb{E}\left[\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}\right]\right)^{-1}\mathbf{X}_0^T\mathbf{\Sigma}^{-1}(\mathbf{y}_0 - \mathbf{X}_0\beta_1(P)),$$

*where $d_0 = d(\mathbf{s}_0, \beta_1(P), \mathbf{\Sigma})$, as defined in (21).*

**Proof.** Consider $\partial \Lambda_1 / \partial \beta$. We have

$$\frac{\partial \Psi_1(\mathbf{s}, \beta, \mathbf{V})}{\partial \beta} = -\frac{u_1'(d)}{d}\mathbf{X}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)^T\mathbf{V}^{-1}\mathbf{X} - u_1(d)\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}, \tag{38}$$

*where $d = d(\mathbf{s}, \beta, \mathbf{V})$, as defined in (21). Since $u_1(s)$ and $u_1'(s)s = \rho_1''(s) - u_1(s)$ are bounded, similar to the proof of Lemma B.3 in Lopuhaä et al. (2022), it follows that for $(\beta, \mathbf{V})$ in the neighborhood $N$ of $(\beta_1(P), \mathbf{V}_0(P))$, it holds that*

$$\frac{\partial \Lambda_1(\beta, \mathbf{V})}{\partial \beta} = \int \frac{\partial \Psi_1(\mathbf{s}, \beta, \mathbf{V})}{\partial \beta} \, dP(\mathbf{s}), \tag{39}$$

*and that $\partial \Lambda_1 / \partial \beta$ is continuous at $(\beta_1(P), \mathbf{V}_0(P))$. Then, similar to the first part of the proof of Lemma 2 in Lopuhaä et al. (2022), it can be shown that*

$$\mathbf{D}_1 = \frac{\partial \Lambda_1(\beta_1(P), \mathbf{V}_0(P))}{\partial \beta} = -\alpha_1 \mathbb{E}\left[\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}\right].$$

The corollary then follows from Theorem 6. □

Since $u_1(s)s$ is bounded, the influence function is uniformly bounded in $\mathbf{y}_0$, but not in $\mathbf{X}_0$. This illustrates the phenomenon in linear regression that leverage points can have a large effect on the regression estimator.

The expression found in Corollary 3 is the same as the one found for the regression S-functional in Lopuhaä et al. (2022) defined with the function $\rho_1$ (see their Corollary 5). For the multivariate location-scale model, for which $\mathbf{X} = \mathbf{I}_k$, Theorem 6 coincides with Theorem 4.2 in Lopuhaä (1992) and Corollary 3 matches with the results found in Salibián-Barrera et al. (2006). Furthermore, the expressions found in Theorem 6 and Corollary 3 are similar to the ones obtained for the regression MM-functionals in Yohai (1987) and Kudraszow and Maronna (2011), respectively. For the influence function of MM-functionals in linear mixed effects models, nothing seems to be available yet. For model (2), the expression for the influence function now follows from Theorem 6. For the special case of this model with multivariate normal errors, as considered in Copt and Heritier (2007), the expression for the influence function can be obtained from Corollary 3.

**Remark 8.1.** The multivariate linear regression model (3) is obtained from (1) by taking $\mathbf{X} = \mathbf{x}^T \otimes \mathbf{I}_k$ and $\beta = \text{vec}(\mathbf{B}^T)$. For this model, the expression in Corollary 3 for $\beta_1(P) = \mathbf{0}$ is similar, but slightly different from the one in Theorem 4 in Kudraszow and Maronna (2011). It seems that Theorem 4 in Kudraszow and Maronna (2011) contains some typos (as confirmed by personal communication). When $\mathbf{T}_1$ is the regression MM-functional considered in Kudraszow and Maronna (2011), then in our notation $\beta_1 = \text{vec}(\mathbf{T}_1^T)$. When $P$ is such that $\mathbf{y} \mid \mathbf{x}$ has an elliptically contoured density with parameters $\mu_0 = \mathbf{B}_0^T\mathbf{x}$ and $\mathbf{\Sigma}_0$, the correct expression for the influence function of $\mathbf{T}_1^T$ should be

$$\text{IF}(\mathbf{y}_0, \mathbf{x}_0; \mathbf{T}_1^T, P) = \frac{1}{\alpha_1}u_1\left(\sqrt{(\mathbf{y}_0 - \mathbf{B}_0^T\mathbf{x}_0)\mathbf{\Sigma}_0^{-1}(\mathbf{y}_0 - \mathbf{B}_0^T\mathbf{x}_0)}\right)(\mathbf{y}_0 - \mathbf{B}_0^T\mathbf{x}_0)\mathbf{x}_0^T\left(\mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right]\right)^{-1},$$

with $\alpha_1$ defined in (37) and $u_1(s) = \rho_1'(s)/s$.

## 9. Asymptotic Normality

Corollary 2 and Theorem 4 provide conditions under which $\beta_{1,n} \to \beta_1(P)$, with probability one, for $\rho_1$ satisfying either (R-BND) or (R-UNB). The next theorem establishes asymptotic normality for $\beta_{1,n}$ defined with either a bounded or an unbounded function $\rho_1$.

**Theorem 7.** *Suppose that $\rho_1 : \mathbb{R} \to [0, \infty)$ either satisfies (R-BND) and (R-CD1), or satisfies (R-UNB). Suppose that $u_1$ is of bounded variation and let $\mathbb{E}\|\mathbf{s}\|^2 < \infty$. Let $\beta_{1,n}$ and $\beta_1(P)$ be solutions to the minimization problems in Definitions 1 and 2, respectively. Suppose that $(\beta_{1,n}, \mathbf{V}_{0,n}) \to (\beta_1(P), \mathbf{V}_0(P))$, in probability, and suppose that $\beta_1(P)$ is a point of symmetry of P. Suppose that $\Lambda_1$, as defined in (33), has a partial derivative $\partial \Lambda_1 / \partial \beta$ that is continuous at $(\beta_1(P), \mathbf{V}_0(P))$ and that $\mathbf{D}_1 = (\partial \Lambda_1 / \partial \beta)(\beta_1(P), \mathbf{V}_0(P))$ is non-singular. Then $\sqrt{n}(\beta_{1,n} - \beta_1(P))$ is asymptotically normal with mean zero and covariance matrix $\mathbf{D}_1^{-1}\mathbb{E}\left[\Psi_1(\mathbf{s}, \beta_1(P), \mathbf{V}_0(P))\Psi_1(\mathbf{s}, \beta_1(P), \mathbf{V}_0(P))^T\right]\mathbf{D}_1^{-1}$.*

**Proof.** Recall that the estimator can be written as $\beta_{1,n} = \beta_1(\mathbb{P}_n)$. This means that it satisfies (31) for $P = \mathbb{P}_n$:

$$\int \Psi_1(\mathbf{s}, \beta_{1,n}, \mathbf{V}_{0,n})\, d\mathbb{P}_n(\mathbf{s}) = \mathbf{0}, \tag{40}$$

where $\Psi_1$ is defined in (32). Writing $\xi_n = (\beta_{1,n}, \mathbf{V}_{0,n})$ and $\xi_P = \xi(P)$, we decompose (40) as follows

$$\mathbf{0} = \int \Psi_1(\mathbf{s}, \xi_n)\, dP(\mathbf{s}) + \int \Psi_1(\mathbf{s}, \xi_P)\, d(\mathbb{P}_n - P)(\mathbf{s})$$
$$+ \int (\Psi_1(\mathbf{s}, \xi_n) - \Psi_1(\mathbf{s}, \xi_P))\, d(\mathbb{P}_n - P)(\mathbf{s}). \tag{41}$$

According to Lemma B.8 in Lopuhaä et al. (2022), the third term is of the order $o_P(1/\sqrt{n})$, whereas according to the central limit theorem the second term is of the order $O_P(1/\sqrt{n})$. This means we can write $\mathbf{0} = \Lambda_1(\xi_n) + O_P(1/\sqrt{n})$, where

$$\Lambda_1(\xi_n) = \Lambda_1(\beta_1(P), \mathbf{V}_{0,n}) + \frac{\partial \Lambda_1}{\partial \beta}(\beta_1(P), \mathbf{V}_{0,n})(\beta_{1,n} - \beta_1(P)) + o(\|\beta_{1,n} - \beta_1(P)\|)$$
$$= \Lambda_1(\beta_1(P), \mathbf{V}_{0,n}) + \left(\mathbf{D}_1 + o_P(1)\right)\left(\beta_{1,n} - \beta_1(P)\right) + o(\|\beta_{1,n} - \beta_1(P)\|).$$

Since $\beta_1(P)$ is a point of symmetry of $P$, according to (34) it holds that $\Lambda_1(\beta_1(P), \mathbf{V}_{0,n}) = \mathbf{0}$. It follows that

$$\mathbf{0} = \mathbf{D}_1(\beta_{1,n} - \beta_1(P)) + o(\|\beta_{1,n} - \beta_1(P)\|) + O_P(1/\sqrt{n}).$$

Because $\mathbf{D}_1$ is non-singular, this implies $\beta_{1,n} - \beta_1(P) = O(1/\sqrt{n})$. After inserting this in (41), it follows that

$$\mathbf{0} = \mathbf{D}_1\left(\beta_{1,n} - \beta_1(P)\right) + \int \Psi_1(\mathbf{s}, \xi_P)\, d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}).$$

We conclude

$$\sqrt{n}(\beta_{1,n} - \beta_1(P)) = -\mathbf{D}_1^{-1}\sqrt{n}\int \Psi_1(\mathbf{s}, \xi_P)\, d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1).$$

Since $\mathbb{E}[\Psi_1(\mathbf{s}, \xi_P)] = \Lambda_1(\xi_P) = \mathbf{0}$, it follows that

$$\sqrt{n}\int \Psi_1(\mathbf{s}, \xi_P)\, d(\mathbb{P}_n - P)(\mathbf{s}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Psi_1(\mathbf{s}_i, \xi_P),$$

converges in distribution to a multivariate normal random vector with mean zero and covariance $\mathbb{E}\left[\Psi_1(\mathbf{s}, \xi_P)\Psi_1(\mathbf{s}, \xi_P)^T\right]$. This finishes the proof. □

When $P$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density (9) we can obtain a more detailed expression for the asymptotic covariance.

**Corollary 4.** *Suppose that $P$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density $f_{\mu,\Sigma}$ from (9) with parameters $(\mu, \Sigma)$. Suppose that $\beta_1(P)$ is the unique minimizer of $R_P(\beta)$, such that $\mathbf{X}\beta_1(P) = \mu$ and suppose that $\mathbf{V}_0(P) = \Sigma$. Let $\beta_{1,n}$ and $\beta_1(P)$ be solutions to the minimization problems in Definitions 1 and 2, respectively, and suppose that $(\beta_{1,n}, \mathbf{V}_{0,n}) \to (\beta_1(P), \mathbf{V}_0(P))$, in probability. Suppose that $\rho_1 : \mathbb{R} \to [0, \infty)$ either satisfies (R-BND), (R-CD1) and (R-CD2) or satisfies (R-UNB) and (R-CD2), such that $u_1(s)$ is of bounded variation and $u_1'(s)s$ is bounded, and let $\mathbb{E}\|\mathbf{s}\|^2 < \infty$. Let $\alpha_1$ be defined in (37) and suppose that $\mathbb{E}_{0,\mathbf{I}_k}[\rho_1''(\|\mathbf{y}\|)] > 0$. If $\mathbf{X}$ has full rank with probability one, then $\sqrt{n}(\beta_{1,n} - \beta_1(P))$ is asymptotically normal with mean zero and covariance matrix*

$$\frac{\mathbb{E}_{0,\mathbf{I}_k}\left[\rho_1'(\|\mathbf{z}\|)^2\right]}{k\alpha_1^2}\left(\mathbb{E}\left[\mathbf{X}^T\Sigma^{-1}\mathbf{X}\right]\right)^{-1}. \tag{42}$$

**Proof.** Similar to the proof of Corollary 3 it can be shown that

$$\mathbf{D}_1 = \frac{\partial \Lambda_1(\beta_1(P), \mathbf{V}_0(P))}{\partial \beta} = -\alpha_1 \mathbb{E}\left[\mathbf{X}^T\Sigma^{-1}\mathbf{X}\right].$$

According to Theorem 7, it follows that $\sqrt{n}(\beta_n - \beta(P))$ is asymptotically normal with mean zero and covariance matrix

$$\frac{1}{\alpha_1^2}\left(\mathbb{E}\left[\mathbf{X}^T\Sigma^{-1}\mathbf{X}\right]\right)^{-1}\mathbb{E}\left[\Psi_1(\mathbf{s}, \beta_1(P), \mathbf{V}_0(P))\Psi_1(\mathbf{s}, \beta_1(P), \mathbf{V}_0(P))^T\right]\left(\mathbb{E}\left[\mathbf{X}^T\Sigma^{-1}\mathbf{X}\right]\right)^{-1},$$

where $\Psi_1$ is defined in (32). Similar to the proof of Corollary 6 in Lopuhaä et al. (2022), we find that

$$\mathbb{E}\left[\Psi_1(\mathbf{s}, \beta_1(P), \mathbf{V}_0(P))\Psi_1(\mathbf{s}, \beta_1(P), \mathbf{V}_0(P))^T\right] = \frac{\mathbb{E}_{0,\mathbf{I}_k}\left[\rho_1'(\|\mathbf{z}\|)^2\right]}{k\alpha_1^2}\mathbb{E}\left[\mathbf{X}^T\Sigma^{-1}\mathbf{X}\right].$$

This proves the corollary. □

The expression found in Corollary 4 coincides with the one for the regression S-estimator in Lopuhaä et al. (2022) defined with the function $\rho_1$ (see their Corollary 6). For the multivariate location-scale model, for which $\mathbf{X} = \mathbf{I}_k$, Theorem 7 for unbounded $\rho_1$, coincides with Theorem 3.2 in Lopuhaä (1992). The expression for the asymptotic variance in Corollary 4 for bounded $\rho_1$, coincides with the results mentioned at the beginning of Section 2.4 in Salibián-Barrera et al. (2006). Furthermore, the results in Theorem 7 and Corollary 4 are similar to the ones obtained in Yohai (1987) and Kudraszow and Maronna (2011) for regression MM-estimators in the multiple and multivariate linear regression model, respectively.

For the linear mixed effects model (2) with $\mathbf{X}_i = \mathbf{X}$ being the same for each subject and assuming multivariate normal measurement errors, Theorem 1 in Copt and Heritier (2007) provides asymptotic normality of the regression MM-estimator. Our Theorem 7 and Corollary 4 are extensions of this result to a larger class of linear mixed effects models also allowing error distributions much more general than the multivariate normal.

**Remark 9.1.** Similar to Remark 8.1, the expression in Corollary 4 for the multivariate linear regression model slightly differs from the one in Proposition 7 in Kudraszow and Maronna (2011). It seems that the expression in equation (6.4) in Kudraszow and Maronna (2011) contains a small typo (as confirmed by personal communication). When $\widehat{\mathbf{B}}_n$ is the regression MM-estimator considered in Kudraszow and Maronna (2011), then in our notation $\beta_{1,n} = \mathrm{vec}(\widehat{\mathbf{B}}_n^T)$. If $P$ is such that $\mathbf{y} \mid \mathbf{x}$ has an elliptically contoured density with parameters $\mu_0 = \mathbf{B}_0^T \mathbf{x}$ and $\mathbf{\Sigma}_0$, the correct expression for the asymptotic variance of $\beta_{1,n} = \mathrm{vec}(\widehat{\mathbf{B}}_n^T)$ should be

$$\frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}\left[\rho_1'(\|\mathbf{z}\|)^2\right]}{k\alpha_1^2}\left(\left(\mathbb{E}[\mathbf{x}\mathbf{x}^T]\right)^{-1} \otimes \mathbf{\Sigma}_0\right),$$

with $\alpha_1$ defined in (37).

Note that for $\rho_B(s; c)$ and $\rho_H(s; c)$, as defined in (5) and (6), respectively, it holds that they both converge to $s^2/2$, as $c \to \infty$. This means that the least squares estimators can be obtained as limiting case of the regression M-estimator defined with $\rho_1$ equal to either $\rho_B(s; c_1)$ or $\rho_H(s; c_1)$, for $c_1 \to \infty$. In both cases, the scalar

$$\lambda = \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}\left[\rho_1'(\|\mathbf{z}\|)^2\right]}{k\alpha_1^2} \to \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}\left[\|\mathbf{z}\|^2\right]}{k}, \quad \text{as } c_1 \to \infty.$$

For the multivariate normal $\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}\left[\|\mathbf{z}\|^2\right] = k$, so that the scalar $1/\lambda$ may serve as an index for the asymptotic efficiency relative to the least squares estimator in all models that are included in our setup.

When using the (bounded) biweight function from (5), Table 1 in Kudraszow and Maronna (2011) gives the cut-off values $c_0$ for which the initial estimators $(\beta_{0,n}, \mathbf{V}_{0,n})$ in Theorem 5 defined with $\rho_0(s) = \rho_B(s; c_0)$ have breakdown point 0.5. For the regression M-estimator $\beta_{1,n}$ defined with $\rho_1(s) = \rho_B(s; c_1)$, information on cut-off values $c_1$ and asymptotic relative efficiencies can be found at several places in the literature. Table 1 in Lopuhaä (1989) provides values of $\lambda$, that correspond to the location S-estimator defined with $\rho_B(s; c_1)$, for varying dimensions $k = 1, 2, 10$ and varying breakdown points $\epsilon^* = 10\%, 20\%, \ldots, 50\%$, from which $c_1$ can be determined from

$$\frac{6\mathbb{E}_{\Phi}[\rho_B(\|\mathbf{z}\|; c_1)]}{c_1^2} = \epsilon^*,$$

where the expectation is with respect to the standard multivariate normal distribution. Table 2 in Kudraszow and Maronna (2011) provides values of $c_1$ for the multivariate regression MM-estimator, for asymptotic efficiencies $1/\lambda = 80\%, 90\%, 0.95\%$ and varying dimensions $k = 1, \ldots, 5, 10$. Finally, Table 3.1 in Van Aelst and Willems (2005) gives asymptotic efficiencies that correspond to the multivariate regression S-estimator for varying dimensions $k = 1, 2, 3, 5, 10, 30, 50$ and breakdown points $\epsilon^* = 25\%, 50\%$.

When using the (unbounded) Huber function from (6), Proposition 1 shows that $\beta_{1,n}$ inherits the breakdown point of the initial covariance estimator $\mathbf{V}_{0,n}$, as long as all $\mathbf{X}_i = \mathbf{X}$ are the same and of full rank. For example, this applies to the linear mixed effects model considered in Copt and Heritier (2007). For the regression M-estimator $\beta_{1,n}$ defined with $\rho_1(s) = \rho_H(s; c_1)$, information on cut-off values $c_1$ and asymptotic relative efficiencies for the multivariate location M-estimator can be found in Table 1 in Maronna (1976) for varying dimensions $k = 2, 4, 6, 10$ and "winsorizing proportions" $w = 0\%, 10\%, 20\%, 30\%$, from which $c_1$ can be determined via

$$P_{\Phi}(\|\mathbf{z}\| > c_1) = w.$$

Table 1 in Lopuhaä (1989) provides values of $\lambda$, that correspond to the location S-estimator defined with $\rho_H(s; c_1)$, for varying dimensions $k = 1, 2, 10$ and the same values for $w$.

## 10. Simulation and data example

We illustrate the finite sample performance of the MM-estimator by means of a simulation. To this end we will study the behavior of the estimators for samples generated from a model that is similar to the one in Copt and Victoria-Feser (2006):

$$\mathbf{y}_i = \mathbf{X}\beta + \gamma_i \mathbf{Z} + \epsilon_i, \quad i = 1, \ldots, n, \tag{43}$$
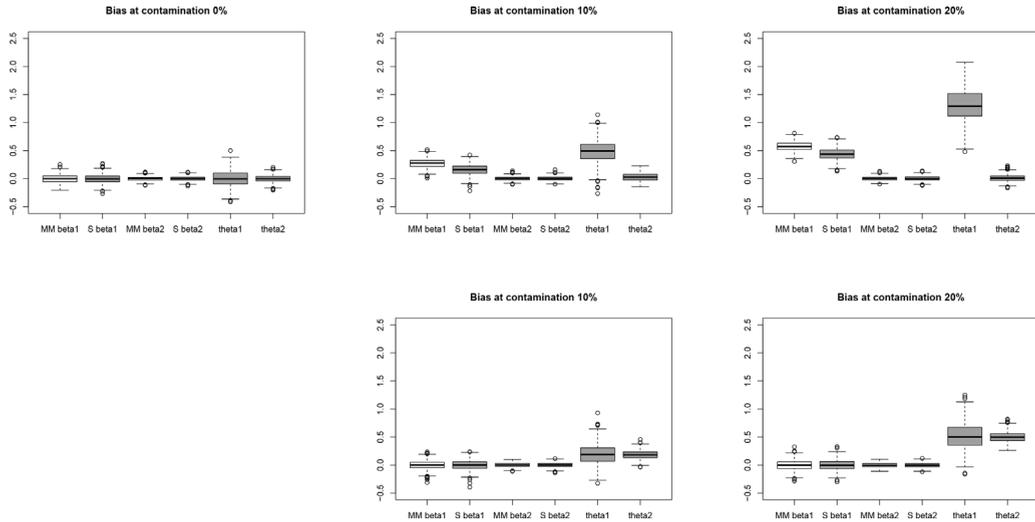
**Fig. 1.** MM-estimates and S-estimates at contaminations 0%, 10%, and 20%, with shift equal to 3 (top row) and 15 (bottom row).

a linear mixed effects model with $\mathbf{y}_i$ in dimension $k = 4$ and all subjects with the same design matrix $\mathbf{X}$ for the fixed effects $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$. Following the setup in Copt and Victoria-Feser (2006), the matrix $\mathbf{X}$ is built as follows. The first column of $\mathbf{X}$ is taken to be a vector $\mathbf{1}$ of length four consisting of ones. The four $x$-values in the second column are generated from a standard normal, and then $\mathbf{X}$ is rescaled to a new matrix $\mathbf{X} = [\mathbf{1} \quad \mathbf{x}]$, such that $\mathbf{X}^T\mathbf{X} = 4\mathbf{I}_2$. For our simulation we used

$$\mathbf{X} = \begin{pmatrix} 1 & -0.9504967 \\ 1 & -0.5428346 \\ 1 & 1.6650521 \\ 1 & -0.1717207 \end{pmatrix}.$$

The random effects $\gamma_i$ are independent $N(0, \sigma_\gamma^2)$ distributed random variables, which are independent from the measurement error $\epsilon_i \sim N(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}_4)$. Finally, $\mathbf{Z} = \mathbf{1}$, a vector of length four consisting of ones. This leads to a structured covariance $\boldsymbol{\Sigma} = \sigma_\gamma^2 \mathbf{1}\mathbf{1}^T + \sigma_\epsilon^2\mathbf{I}_4$, with covariance parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$, where $\theta_1 = \sigma_\gamma^2$ and $\theta_2 = \sigma_\epsilon^2$. Following the setup in Copt and Victoria-Feser (2006), we set $\beta_1 = \beta_2 = 1$ and $\theta_1 = \theta_2 = 1$.

We investigate the behavior of the MM-estimator and the S-estimator used in the first step. To this end we generate 500 samples of size $n = 200$ according to the model in (43) and compute the estimates for each sample. We also study the behavior under contamination, by replacing some of $\mathbf{y}_i$ by $\mathbf{y}_i^*$ obtained by shifting each coordinate of $\mathbf{y}_i$ over the same distance. We consider two distances: 3 and 15 and two contamination percentages: 10% and 20% contamination. We used the S-estimator corresponding to Tukey's bi-weight defined in (5). The tuning-constant was chosen to be $c_0 = 4.097$, which corresponds to asymptotic breakdown point 0.5. For the MM-estimator we used the bi-weight function with tuning constant $c_1 = 8.530$, which corresponds to 99% efficiency relative to the least squares estimator.

The biases of the estimators are displayed in Figure 1. In the case of no contamination, the performance of all estimators is similar; all the biases are very small. The fact that the MM-estimator improves the efficiency is illustrated by the fact that the sample variances of the MM-estimates are 83% and 82% of the sample variances of the S-estimates for $\beta_1$ and $\beta_2$, respectively. This is comparable to 81%, which is in both cases the ratio of the two asymptotic variances. With 10% contamination by shifting over distance 3, the S-estimator for $\theta_1$ is showing some bias, as well as the S-estimator for $\beta_1$ to a much lesser extent. The bias of the S-estimator for $\theta_1$, also seems to affect the MM-estimator for $\beta_1$. These effects are getting larger when we increase the contamination to 20%. When shifting over distance 15, the bias in the estimators for $\beta_1$ disappears, and is much less for the S-estimator for $\theta_1$, whereas the bias of the S-estimator for $\theta_2$ increases.

This behavior may be explained by the fact that shifting coordinates of contaminated points over distance 3 corresponds to having a difference of about two standard deviations between the means of the main model and the contaminated one. In this scenario, the S- and MM-estimators seem to have difficulty separating contaminated points from uncontaminated points. The S-estimator searches for the smallest cylinder that contains a sufficient number of points such that it satisfies the constraint in the corresponding minimization problem. By having contaminated points so close to the bulk of the original data, the S-estimator seems to adapt too much to the contamination by shifting the intercept of the central axis of the cylinder and stretching the size of the cylinder along the main principal component, both in the same direction of the shifted points. This seems to cause the biases in the S-estimators for $\beta_1$ and $\theta_1$. As a consequence, the MM-estimator is based on points that are standardized by a covariance matrix that is too large in the direction of the contaminated points. As such, it will shift even further in the direction of the contaminated points, causing an additional bias in the MM-estimator for $\beta_1$. Shifting over distance 15 corresponds to a difference of about ten standard deviations between the means of the

**Table 1**
LS estimates, S-estimates, and MM-estimates for $\beta_1, \ldots, \beta_6$ with their standard errors.

|  | LS |  | S |  | MM |  |
|---|---|---|---|---|---|---|
| $\beta_1$ | 23.973 | (0.899) | 23.188 | (0.962) | 23.341 | (0.883) |
| $\beta_2$ | 1.996 | (1.272) | 2.177 | (1.361) | 2.342 | (1.249) |
| $\beta_3$ | -7.541 | (0.591) | -6.298 | (0.548) | -7.047 | (0.503) |
| $\beta_4$ | 6.624 | (0.835) | 5.298 | (0.775) | 6.089 | (0.712) |
| $\beta_5$ | 1.196 | (0.097) | 0.985 | (0.085) | 1.105 | (0.078) |
| $\beta_6$ | -1.104 | (0.137) | -0.890 | (0.120) | -1.012 | (0.110) |
| $\sigma_{\gamma_1}^2$ | 23.253 |  | 25.756 |  |  |  |
| $\sigma_{\gamma_2}^2$ | -0.009 |  | 0.575 |  |  |  |
| $\sigma_{\gamma_3}^2$ | 0.004 |  | -0.020 |  |  |  |
| $\sigma_\epsilon^2$ | 21.785 |  | 14.293 |  |  |  |

main model and the contaminated one. In this scenario, the contaminated points are clearly separated from the bulk of the original data. This seems to make it easier for the S-estimator to find a smallest cylinder $\mathcal{C}(\beta, \theta)$ around the bulk of the original data. As such, the central axis of the cylinder remains unaffected, resulting in zero bias for the S-estimators for $\beta_1$ and $\beta_2$. In order to capture a sufficient number of points to satisfy the constraint of the S-minimization problem, the S-estimator for the covariance structure is inflated a bit in all directions. This causes the same amount of bias in the S-estimator for $\theta_1$ and $\theta_2$, but not as much as in the previous scenario. Moreover, the MM-estimator is now based on points that are standardized by a covariance matrix that no longer favors a particular direction. As such, it exhibits zero bias in both 10% and 20% contamination.

Finally, we illustrate the performance of the MM-estimator by an application to data from a trial on the treatment of lead-exposed children. This dataset is discussed in Fitzmaurice et al. (2011) and consists of four repeated measurements of blood lead levels obtained at baseline (or week 0), week 1, week 4, and week 6 on 100 children who were randomly assigned to chelation treatment with succimer (a chelation agent) or placebo. On the basis of a graphical display of the mean response over time, it is suggested in Fitzmaurice et al. (2011) that a quadratic trend over time seems suitable. We fitted the following model

$$y_{ij} = \beta_0 + \beta_1 \delta_i + (\beta_3 + \beta_4 \delta_i) t_j + (\beta_5 + \beta_6 \delta_i) t_j^2 + \gamma_{1i} + \gamma_{2i} t_j + \gamma_{3i} t_j^2 + \epsilon_{ij},$$

for $i = 1, \ldots, 100$ and $j = 1, \ldots, 4$, where $(t_1, \ldots, t_4) = (0, 1, 4, 6)$ refer to the different weeks, $y_{ij}$ is the blood lead level (mcg/dL) of subject $i$ obtained at time $t_j$, and $\delta_i = 0$ if the $i$-th subject is in the placebo group and $\delta_i = 1$, otherwise. The random effects $\gamma_i = (\gamma_{1i}, \gamma_{2i}, \gamma_{3i})$, $i = 1, \ldots, 100$, are assumed to be independent mean zero normal random vectors with a diagonal covariance matrix consisting of variances $\sigma_{\gamma_1}^2$, $\sigma_{\gamma_2}^2$ and $\sigma_{\gamma_3}^2$, respectively. The measurement errors $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{i4})$, $i = 1, \ldots, 100$, are assumed to be independent mean zero random vectors with covariance matrix $\sigma_\epsilon^2 \mathbf{I}_4$, also being independent of the random effects. In this way we are fitting a balanced linear mixed effects model with unknown parameters $\beta = (\beta_1, \ldots, \beta_6)$ and $\theta = (\sigma_{\gamma_1}^2, \sigma_{\gamma_2}^2, \sigma_{\gamma_3}^2, \sigma_\epsilon^2)$.

We estimated $(\beta, \theta)$ by means of least squares and by means of the S-estimator corresponding to Tukey's bi-weight defined in (5). The tuning-constant was chosen to be $c_0 = 4.097$, which corresponds to asymptotic breakdown point 0.5 and 80% efficiency relative to the least squares estimator. For the MM-estimator we used the bi-weight function with tuning constant $c_1 = 5.810$, which corresponds to 95% efficiency relative to the least squares estimator. The resulting estimates and their standard errors (between brackets) are given in Table 1.

The standard errors can be computed from (42), taking the tuning-constant $c_1$ equal to infinity, 4.097 and 5.810, for the LS, S-, and MM-estimator, respectively. The MM-estimates are more in line with the least squares estimates and have smaller standard errors than those of the S-estimates. At the same time, the robust standardized residuals computed from the MM-estimates $\beta_{1,n}$ and $\mathbf{V}(\theta_{0,n})$, are almost identical to ones computed from the S-estimates $\beta_{0,n}$ and $\mathbf{V}(\theta_{0,n})$. This means that the MM-estimator identifies the same observations 40 and 98 as outliers, whereas the least squares estimator only identifies observation 40 (see also Figure 2 in Lopuhaä et al. (2022)).

## 11. Discussion

We have provided a unified approach for constructing estimators of the regression parameter in balanced linear models with a structured covariance that combine good robustness properties, such as a high breakdown point and an influence function that is bounded in $\mathbf{y}$, with high asymptotic efficiency at models with multivariate normal errors. Our setup is sufficiently flexible to include several specific multivariate statistical models, including linear mixed effects models, multivariate and multiple linear regression, and multivariate location and scale. In this way, the theory for these multivariate models can be handled at the same time.

Combining high breakdown and high efficiency originated with the MM-estimators introduced by Yohai (1987) for the multiple regression model. For the models included in our setup, extensions have been developed in two ways. For example, one approach for multivariate location and scatter is to start by obtaining an initial high breakdown estimator of the covari-

ance matrix, and use this to standardize the observations and determine an location M-estimator from the rescaled observations in such a way that it has high efficiency and inherits the breakdown point of the covariance estimator, see Lopuhaä (1992). A similar approach was used in Copt and Heritier (2007) for the regression estimator in linear mixed effects models. Another approach for multivariate location and scatter is to start by obtaining initial high breakdown estimators of location and of the shape of the scatter matrix, and use these to determine an auxiliary univariate M-estimator of scale. The estimators of location and shape are then updated to improve the efficiency, after which the latter is combined with the M-estimator of scale to determine the final estimator of scatter, see Tatsuoka and Tyler (2000) and Salibián-Barrera et al. (2006). This approach was also used in Kudraszow and Maronna (2011) for multivariate linear regression. Since the main objective for our general setup was to construct an estimator for the regression parameter that combines high breakdown with high efficiency we extended the somewhat simpler approach used in Lopuhaä (1992); Copt and Heritier (2007). The more involved approach used in Tatsuoka and Tyler (2000); Salibián-Barrera et al. (2006); Kudraszow and Maronna (2011) does have the advantage that also the efficiency of the covariance shape estimator can be improved. This approach is beyond the scope of this paper and will be considered in a future manuscript.

Our main interest is to combine high breakdown with high efficiency for linear mixed effects models, for which the theory is far from complete. The results in the literature are limited to models in which the design matrix for the fixed effects is the same for all subjects, and only deal with the asymptotic behavior under the assumption of normally distributed errors. Moreover, no rigorous attention has been paid to whether the estimators actually exist and what their robustness properties are. Our general setup includes linear mixed effects models, which allow different subject specific design matrices and general multivariate distributions that go far beyond the multivariate normal or other elliptically contoured distributions. We have provided sufficient conditions under which the estimators and corresponding statistical functionals exist. Furthermore, we have given conditions under which the regression estimator inherits the (high) breakdown point of the covariance estimator in the initial step, and we have derived the expression of the influence function. Finally, we have established strong consistency and asymptotic normality of the regression estimator under general distributions, from which more detailed expressions can be determined at the multivariate normal or other elliptically contoured distributions.

Since our setup also includes the multivariate location-scale model, our results also include the ones obtained in Lopuhaä (1992) as a special case. In addition, our results have extended the ones on existence and breakdown point to location MM-estimators defined by means of a bounded $\rho_1$ in Definition 1. Although our simpler approach differs from the one in Salibián-Barrera et al. (2006), our results on the breakdown behavior of the location MM-estimator are similar. Furthermore, our results on the influence function and on the asymptotic behavior of the location MM-estimator are much more general, but are identical to the ones in Salibián-Barrera et al. (2006) at the special case of elliptically contoured distributions. Also the multivariate linear regression model is included in our setup, although our simpler approach differs from the one in Kudraszow and Maronna (2011). Again our results on the existence and the breakdown behavior of the regression MM-estimator are similar, and our results on the influence function and the asymptotic behavior of the regression MM-estimator are more general, but identical to the ones in Kudraszow and Maronna (2011) at the special case of elliptically contoured errors distributions.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecosta.2023.03.003

## References

Agostinelli, C., Yohai, V.J., 2016. Composite robust estimators for linear mixed models. Journal of the American Statistical Association 111 (516), 1764–1774.

Billingsley, P., 1968. Convergence of probability measures. John Wiley & Sons, Inc., New York-London-Sydney.

Cator, E.A., Lopuhaä, H.P., 2012. Central limit theorem and influence function for the MCD estimators at general multivariate distributions. Bernoulli 18 (2), 520–551.

Chervoneva, I., Vishnyakov, M., 2011. Constrained *S*-estimators for linear mixed effects models with covariance components. Stat. Med. 30 (14), 1735–1750.

Chervoneva, I., Vishnyakov, M., 2014. Generalized *s*-estimators for linear mixed effects models. Statistica Sinica 24 (3), 1257–1276.

Copt, S., Heritier, S., 2007. Robust alternatives to the f-test in mixed linear models based on mm-estimates. Biometrics 63 (4), 1045–1052.

Copt, S., Victoria-Feser, M.P., 2006. High-breakdown inference for mixed linear models. Journal of the American Statistical Association 101 (473), 292–300.

Davies, P.L., 1987. Asymptotic behaviour of *S*-estimates of multivariate location parameters and dispersion matrices. Ann. Statist. 15 (3), 1269–1292.

Demidenko, E., 2013. Mixed models. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ. Second edition. Theory and applications with R

Donoho, D., Huber, P.J., 1983. The notion of breakdown point. In: A Festschrift for Erich L. Lehmann. In: Wadsworth Statist./Probab. Ser., pages 157–184. Wadsworth, Belmont, CA.

Fishbone, J., 2021. Highly Robust and Efficient Estimators of Multivariate Location and Covariance with Applications to Array Processing and Financial Portfolio Optimization. PhD thesis. Virginia Polytechnic Institute and State University.

Fitzmaurice, G.M., Laird, N.M., Ware, J.H., 2011. Applied longitudinal analysis. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ. Second edition

Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 69, 383–393.

Hartley, H.O., Rao, J.N.K., 1967. Maximum-likelihood estimation for the mixed analysis of variance model. Biometrika 54, 93–108.

Heritier, S., Cantoni, E., Copt, S., Victoria-Feser, M.P., 2009. Robust methods in biostatistics. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Huber, P.J., 1984. Finite sample breakdown of $M$- and $P$-estimators. Ann. Statist. 12 (1), 119–126.

Jennrich, R.I., Schluchter, M.D., 1986. Unbalanced repeated-measures models with structured covariance matrices. Biometrics 42 (4), 805–820.

Koller, M., 2013. Robust estimation of linear mixed models. PhD thesis. ETH Zurich.

Kudraszow, N.L., Maronna, R.A., 2011. Estimates of MM type for the multivariate linear model. J. Multivariate Anal. 102 (9), 1280–1292.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38 (4), 963–974.

Lopuhaä, H.P., 1989. On the relation between $S$-estimators and $M$-estimators of multivariate location and covariance. Ann. Statist. 17 (4), 1662–1683.

Lopuhaä, H.P., 1992. Highly efficient estimators of multivariate location with high breakdown point. The Annals of Statistics 398–413.

Lopuhaä, H.P., Gares, V., Ruiz-Gazen, A., 2022. S-estimation in linear models with structured covariance matrices. Major revision under review. Available at arxive: 4361349.

Maronna, R.A., 1976. Robust $M$-estimators of multivariate location and scatter. Ann. Statist. 4 (1), 51–67.

Park, C.-H., Chang, J.H., 2021. Robust localization employing weighted least squares methid based on mm estimator and kalman filter with maximum versoria criterion. IEEE Signal Processing Letters 28, 1075–1079.

Pinheiro, J.C., Liu, C., Wu, Y.N., 2001. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate $t$ distribution. J. Comput. Graph. Statist. 10 (2), 249–276.

Pollard, D., 1984. Convergence of stochastic processes. Springer Series in Statistics. Springer-Verlag, New York.

Rao, C.R., 1972. Estimation of variance and covariance components in linear models. J. Amer. Statist. Assoc. 67, 112–115.

Rousseeuw, P., 1985. Multivariate estimation with high breakdown point. In: Mathematical statistics and applications, Vol. B (Bad Tatzmannsdorf, 1983), pages 283–297. Reidel, Dordrecht.

Rousseeuw, P., Yohai, V., 1984. Robust regression by means of S-estimators. In: Robust and nonlinear time series analysis (Heidelberg, 1983). Springer, New York. Volume 26 of Lect. Notes Stat., pages 256–272

Salibián-Barrera, M., Van Aelst, S., Willems, G., 2006. Principal components analysis based on multivariate MM estimators with fast and robust bootstrap. J. Amer. Statist. Assoc. 101 (475), 1198–1211.

Tatsuoka, K.S., Tyler, D.E., 2000. On the uniqueness of $S$-functionals and $M$-functionals under nonelliptical distributions. Ann. Statist. 28 (4), 1219–1243.

Van Aelst, S., Willems, G., 2005. Multivariate regression $S$-estimators for robust estimation and inference. Statist. Sinica 15 (4), 981–1001.

Yohai, V.J., 1987. High breakdown-point and high efficiency robust estimates for regression. Ann. Statist. 15 (2), 642–656.