

Multivariate MM-estimators with auxiliary Scale for Linear Models with Structured Covariance Matrices

Hendrik Paul Lopuhaä¹

¹*Delft University of Technology*

November 10, 2025

Abstract

We provide a unified approach to MM-estimation with auxiliary scale for balanced linear models with structured covariance matrices. This approach leads to estimators that are highly robust against outliers and highly efficient for normal data. These properties not only hold for estimators of the regression parameter, but also for estimators of scale invariant transformations of the variance parameters. Of main interest are MM-estimators for linear mixed effects models, but our approach also includes MM-estimators in several other standard multivariate models. We provide sufficient conditions for the existence of MM-functionals and MM-estimators, establish asymptotic properties such as consistency and asymptotic normality, and derive their robustness properties 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.

1 Introduction

Linear models with a structured covariance are a generalization of traditional linear models where the residuals are assumed to follow a specific covariance structure rather than being independent and identically distributed. This approach is useful when the residuals are correlated or exhibit some form of structure that can't be captured by simple uncorrelated noise. These models are often used in cases like repeated measures, longitudinal data, and hierarchical structures, where the observations within a group or over time might be more similar to each other than to observations from other groups or time points. An example are linear mixed effects models, which explicitly account for both fixed effects (predictors whose effects are the same across all units) and random effects (predictors whose effects vary across groups or subjects). In these models, the random effects together with the residuals yields a specific covariance structure depending on a vector of unknown covariance parameters.

Maximum likelihood estimation has been studied by Hartley and Rao [12], Rao [28], Laird and Ware [16], see also Fitzmaurice *et al* [10] and Demidenko [8]. To be resistant against outliers, robust methods have been investigated for linear mixed effects models by Pinheiro *et al* [26], Copt [5], Copt and Heritier [4], Heritier *et al* [13], Agostinelli and Yohai [1], and Chervoneva and Vishnyakov [3], or for more general linear models with a structured covariance by Lopuhaä *et al* [22]. This often concerns S-estimators originally proposed by Rousseeuw and Yohai [30] for the multiple linear regression model. These estimators have been extended to several multivariate statistical models and can be viewed as smooth versions of the minimum volume ellipsoid estimator, introduced by Rousseeuw [29], that are highly resistant against outliers. However, one drawback of S-estimators is that they suffer from a low efficiency.

Some extensions have been proposed that inherit the robustness of the S-estimator, but at the same time improve the efficiency. Among them are the MM-estimators introduced by Yohai [36]

for the multiple linear regression model. The idea is to estimate the scale by means of a robust M-estimator, and then estimate the regression parameter using a regression M-estimator with a different loss function that yields better efficiency. This idea has been extended in different ways to multivariate statistical models. Lopuhaä [18] proposed a version for multivariate location, Copt and Heritier [4] used the same approach to estimate the fixed effects in a linear mixed effects model, and a similar method has been studied in Lopuhaä [19] for more general linear models with a structured covariance. All these proposals use a robust estimator of the entire scatter matrix in the first step and only allow efficiency improvement of the location or regression estimator. Tatsuka and Tyler [32] introduced a more extensive version of multivariate MM-estimators for multivariate location and scatter, being members of a broad class of multivariate M-estimators with auxiliary scale. Their proposal only uses a robust M-estimator of the scale of the scatter matrix in the first step and allows efficiency improvement of both the location estimator as well as the estimator of the shape of the scatter matrix. For this reason, this version of multivariate MM-estimators with auxiliary scale is particularly useful for applications that require estimation of a covariance matrix.

The theory for these estimators is fairly limited. Kudraszow and Maronna [15] study MM-estimators with auxiliary scale for multivariate linear regression, but no rigorous results are derived for the covariance MM-estimator. Tatsuka and Tyler [32] study existence of the corresponding MM-functionals, but no attention is paid to the limiting behavior of the MM-estimators themselves. As a basis for a robust PCA method, Salibián-Barrera *et al* [31] use covariance MM-estimators and discusses their limiting behavior, but a rigorous derivation is missing.

In view of this, we provide a unified approach to MM-estimation with auxiliary scale in balanced linear models with structured covariance matrices. The balanced setup is already quite flexible and includes several specific multivariate statistical models. Of main interest are MM-estimators for linear mixed effects models, but our approach also includes MM-estimators in several other standard multivariate models, such as multivariate linear regression, and multivariate location and scatter. We provide sufficient conditions for the existence of MM-functionals and MM-estimators, 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.

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 MM-estimator and MM-functional and in Section 4 we give conditions under which they exist. In Section 5 we establish continuity of the MM-functional, which is then used to obtain consistency of the MM-estimator. Section 6 deals with the breakdown point. Section 7 provides the preparation for Sections 8 and 9 in which we obtain the influence function and establish asymptotic normality. Our results lead to single scalar indices for the asymptotic efficiency and the gross-error-sensitivity of standardized components of the MM-estimators of the variance parameters. In Section 10 we investigate the interplay between these two scalars at the multivariate normal and Student distributions. All proofs are available as supplemental material [21].

2 Balanced linear models with structured covariances

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

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i, \quad i = 1, \dots, n, \quad (2.1)$$

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 $\mathbf{V} \in \text{PDS}(k)$, 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 $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^l$. We first discuss some examples that are covered by this setup in the context of MM-estimators.

Example 1. An important case of interest is the (balanced) linear mixed effects model $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i$, for $i = 1, \dots, n$. This model arises from $\mathbf{u}_i = \mathbf{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i$, for $i = 1, \dots, n$, where $\mathbf{Z} \in \mathbb{R}^{k \times g}$ is known and $\boldsymbol{\gamma}_i \in \mathbb{R}^g$ and $\boldsymbol{\epsilon}_i \in \mathbb{R}^k$ are independent mean zero random variables, with unknown covariance matrices $\mathbf{G}$ and $\mathbf{R}$, respectively. In this case, $\mathbf{V}(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$ and $\boldsymbol{\theta} = (\text{vech}(\mathbf{G})^T, \text{vech}(\mathbf{R})^T)^T$, where

$$\text{vech}(\mathbf{A}) = (a_{11}, \dots, a_{k1}, a_{22}, \dots, a_{kk}) \quad (2.2)$$

is the unique $k(k+1)/2$ -vector that stacks the columns of the lower triangle elements of a symmetric matrix $\mathbf{A}$. In full generality, the model is usually overparametrized and one may run into identifiability problems. A more feasible example is obtained by taking $\mathbf{R} = \sigma_0^2 \mathbf{I}_k$, $\mathbf{Z} = [\mathbf{Z}_1 \cdots \mathbf{Z}_r]$ and $\boldsymbol{\gamma}_i = (\gamma_{i1}^T, \dots, \gamma_{ir}^T)^T$, 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 $j = 1, \dots, r$. This leads to

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{j=1}^r \mathbf{Z}_j\boldsymbol{\gamma}_{ij} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n, \quad (2.3)$$

which was considered in Copt and Heritier [4]. In this case, $\mathbf{V}(\boldsymbol{\theta}) = \sigma_0^2 \mathbf{I}_k + \sum_{j=1}^r \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j^T$ and $\boldsymbol{\theta} = (\sigma_0^2, \sigma_1^2, \dots, \sigma_r^2)$.

Example 2. Another example of (2.1) is the multivariate linear regression model

$$\mathbf{y}_i = \mathbf{B}^T \mathbf{x}_i + \mathbf{u}_i, \quad i = 1, \dots, n, \quad (2.4)$$

considered in Kudraszow and Maronna [15], 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 $i = 1, \dots, n$, are independent mean zero random variables with covariance matrix $\mathbf{V}(\boldsymbol{\theta}) = \boldsymbol{\Sigma} \in \text{PDS}(k)$. In this case, the vector of unknown covariance parameters is given by $\boldsymbol{\theta} = \text{vech}(\boldsymbol{\Sigma})$, where $\text{vech}(\cdot)$ is defined in (2.2). The model can be obtained as a special case of (2.1), 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 [36] is a special case of (2.4) with $k = 1$.

Example 3. Also the multivariate location-scatter model, as considered in Lopuhaä [18], Tatsuoaka and Tyler [32], and Salibián-Barrera et al [31], can be obtained as a special case of (2.1), 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 $\mathbf{V}(\boldsymbol{\theta}) = \boldsymbol{\Sigma} \in \text{PDS}(k)$, with $\boldsymbol{\theta} = \text{vech}(\boldsymbol{\Sigma})$. Note that this model can also be obtained as a special case of (2.4) by taking $\mathbf{x}_i = 1$ and $\mathbf{B}^T = \boldsymbol{\beta}$. This means that results in Kudraszow and Maronna [15] for model (2.4) also apply to the multivariate location-scatter model.

Example 4. Model (2.1) also includes examples, for which $\mathbf{u}_1, \dots, \mathbf{u}_n$ 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 $s, t = 1, \dots, n$. 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 $s, t = 1, \dots, n$, in which case $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^T \in \mathbb{R}^k$, 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 models in Examples 2, 3, and 4. This may not be true in general for the linear mixed effects model in Example 1 with unknown $\text{vech}(\mathbf{G})$ and $\text{vech}(\mathbf{R})$. For linear mixed effects models in (2.3), identifiability of $\boldsymbol{\theta} = (\sigma_0^2, \sigma_1^2, \dots, \sigma_r^2)$ holds for particular choices of the design matrices $\mathbf{Z}_1, \dots, \mathbf{Z}_r$.

3 Definitions

We start by representing our observations as points in $\mathbb{R}^k \times \mathbb{R}^{kq}$ in the following way. For $r = 1, \dots, k$, 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 $\mathbf{s}^T = (\mathbf{y}^T, \mathbf{x}_1^T, \dots, \mathbf{x}_k^T)$. In this way our observations can be represented as $\mathbf{s}_1, \dots, \mathbf{s}_n$, with $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i) \in \mathbb{R}^k \times \mathbb{R}^{kq}$.

Similar to MM-estimators for multiple linear regression introduced by Yohai [36], MM-estimators for $(\boldsymbol{\beta}, \boldsymbol{\theta})$ are based on two loss functions. We require the following conditions for a loss function ρ :

- (R1) ρ is symmetric around zero with $\rho(0) = 0$ and ρ is continuous at zero;
- (R2) There exists a finite constant $c > 0$, such that ρ is non-decreasing on $[0, c]$ and constant on $[c, \infty)$;
- (R3) ρ is continuous and strictly increasing on $[0, c]$.

In comparison with other proposals for MM-estimators, conditions (R1)-(R3) imply condition (A1) in Yohai [36] and Definition 2 in Kudraszow and Maronna [15]. The conditions are similar to the ones in Tatsuoka and Tyler [32] and the ones in Salibián-Barrera *et al* [31].

STAGE 1: Let $\boldsymbol{\beta}_{0,n}$ and $\boldsymbol{\theta}_{0,n}$ be initial (high breakdown) estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, and consider the shape estimator $\boldsymbol{\Gamma}(\boldsymbol{\theta}_{0,n})$, where for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \frac{\mathbf{V}(\boldsymbol{\theta})}{|\mathbf{V}(\boldsymbol{\theta})|^{1/k}}, \quad (3.1)$$

where $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$.

STAGE 2: Let ρ_0 satisfy (R1)-(R3) and determine σ_n by solving σ from

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left(\frac{\sqrt{(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{0,n})^T \boldsymbol{\Gamma}(\boldsymbol{\theta}_{0,n})^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_{0,n})}}{\sigma} \right) = b_0, \quad (3.2)$$

where $0 < b_0 < \sup \rho_0$.

STAGE 3: Let ρ_1 satisfy (R1)-(R3) and is such that

$$\frac{\rho_1}{\sup \rho_1} \leq \frac{\rho_0}{\sup \rho_0}. \quad (3.3)$$

For $(\boldsymbol{\beta}, \mathbf{C}) \in \mathbb{R}^q \times \text{PDS}(k)$, define

$$R_n(\boldsymbol{\beta}, \mathbf{C}) = \frac{1}{n} \sum_{i=1}^n \rho_1 \left(\frac{\sqrt{(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^T \mathbf{C}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})}}{\sigma_n} \right), \quad (3.4)$$

and let $\mathfrak{D} = \{(\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathbb{R}^q \times \boldsymbol{\Theta} : \mathbf{V}(\boldsymbol{\gamma}) \in \text{PDS}(k) \text{ with } |\mathbf{V}(\boldsymbol{\gamma})| = 1\}$. Let $(\boldsymbol{\beta}_{1,n}, \boldsymbol{\gamma}_n) \in \mathfrak{D}$ be any local minimum of $R_n(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$ that satisfies

$$R_n(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma})) \leq R_n(\boldsymbol{\beta}_{0,n}, \boldsymbol{\Gamma}(\boldsymbol{\theta}_{0,n})), \quad (3.5)$$

where $\boldsymbol{\Gamma}$ is defined in (3.1). Update the covariance estimator by $\mathbf{V}_{1,n} = \sigma_n^2 \mathbf{V}(\boldsymbol{\gamma}_n)$ and update the estimator $\boldsymbol{\theta}_{1,n}$ for the vector of covariance parameters as the solution of

$$\mathbf{V}(\boldsymbol{\theta}) = \sigma_n^2 \mathbf{V}(\boldsymbol{\gamma}_n). \quad (3.6)$$

The idea is to choose estimators $\beta_{0,n}$ and $\theta_{0,n}$ with high breakdown point and to choose loss function ρ_0 suitably, so that σ_n will also have high breakdown point. Estimators $\beta_{1,n}$ and $\theta_{1,n}$ will be shown to inherit this high breakdown point, but at the same time the regression estimator $\beta_{1,n}$ as well as the estimator of shape $\mathbf{V}_{1,n}/|\mathbf{V}_{1,n}|^{1/k}$ and the estimator of direction $\theta_{1,n}/\|\theta_{1,n}\|$ will also have high efficiency relative to the least squares estimators by suitable choice of loss function ρ_1 . We will show that the absolute minimum of $R_n(\beta, \mathbf{V}(\gamma))$ with $|\mathbf{V}(\gamma)| = 1$ exists. Clearly, this absolute minimum satisfies (3.5). However, any local minimum satisfying (3.5), will also be an MM-estimator with high breakdown point and high efficiency.

Examples of loss functions satisfying (3.3) can be constructed from Tukey's bi-weight, defined as

$$\rho_B(s; c) = \begin{cases} s^2/2 - s^4/(2c^2) + s^6/(6c^4), & |s| \leq c \\ c^2/6 & |s| > c. \end{cases} \quad (3.7)$$

The functions $\rho_0(s) = \rho_B(s; c_0)$ and $\rho_1(s) = \rho_B(s; c_1)$, for $0 < c_0 \leq c_1 < \infty$, satisfy (R1)-(R3) as well as (3.3). Examples of $\beta_{0,n}$ and $\theta_{0,n}$ with high breakdown point are the S-estimators discussed in Lopuhaä *et al* [22] defined with $\rho_0(s) = \rho_B(s; c_0)$. Small values of the cut-off constant c_0 will then correspond to a high breakdown point.

This definition of MM-estimators for the linear mixed effects model differs from the ones in Lopuhaä [19] and Copt and Heritier [4], where the entire initial covariance matrix is used as initial estimator, which then involves minimization of R_n over β only. The current definition only uses the univariate estimator σ_n for the scale parameter $|\mathbf{V}(\theta)|^{1/(2k)}$ as an auxiliary statistic. The advantage is that this version of the MM-estimator allows for improvement of the efficiency of both the regression estimator as well as the estimator of the shape component of $\mathbf{V}(\theta)$ and the estimator of the direction component of θ .

The corresponding MM-functionals are defined similarly.

STAGE 1: Let $\beta_0(P)$ and $\theta_0(P)$ be initial functionals and consider the shape functional $\Gamma(\theta_0(P))$, where Γ is defined in (3.1).

STAGE 2: Let ρ_0 satisfy (R1)-(R3) and determine $\sigma(P)$ by solving σ from

$$\int \rho_0 \left(\frac{\sqrt{(\mathbf{y} - \mathbf{X}\beta_0(P))^T \Gamma(\theta_0(P))^{-1} (\mathbf{y} - \mathbf{X}\beta_0(P))}}{\sigma} \right) dP(\mathbf{s}) = b_0, \quad (3.8)$$

where $0 < b_0 < \sup \rho_0$.

STAGE 3: Let ρ_1 satisfy (R1)-(R3) and is such that (3.3) holds. For $(\beta, \mathbf{C}) \in \mathbb{R}^q \times \text{PDS}(k)$, define

$$R_P(\beta, \mathbf{C}) = \int \rho_1 \left(\frac{\sqrt{(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{C}^{-1} (\mathbf{y} - \mathbf{X}\beta)}}{\sigma(P)} \right) dP(\mathbf{s}). \quad (3.9)$$

Let $(\beta_1(P), \gamma(P)) \in \mathfrak{D}$ be any local minimum of $R_P(\beta, \mathbf{V}(\gamma))$ that satisfies

$$R_P(\beta, \mathbf{V}(\gamma)) \leq R_P(\beta_0(P), \Gamma(\theta_0(P))), \quad (3.10)$$

where Γ is defined in (3.1). Update the covariance functional by $\mathbf{V}_1(P) = \sigma^2(P)\mathbf{V}(\gamma(P))$ and update the functional $\theta_1(P)$ for the vector of covariance parameters as the solution of

$$\mathbf{V}(\theta) = \sigma^2(P)\mathbf{V}(\gamma(P)). \quad (3.11)$$

Let $\mathbb{P}_n$ be the empirical measure corresponding to observations $(\mathbf{y}_1, \mathbf{X}_1), \dots, (\mathbf{y}_n, \mathbf{X}_n)$. We assume that the initial functionals $\beta_0(\cdot)$ and $\theta_0(\cdot)$ are such that

$$(\beta_0(\mathbb{P}_n), \theta_0(\mathbb{P}_n)) = (\beta_{0,n}, \theta_{0,n}), \quad (3.12)$$

where $(\beta_{0,n}, \theta_{0,n})$ are the initial estimators for β and θ . Examples for which (3.12) holds, are the S-functionals discussed in Lopuhaä *et al* [22] defined with loss function ρ_0 . If (3.12) holds, then $\sigma_n = \sigma(\mathbb{P}_n)$, $(\beta_{1,n}, \gamma_n) = (\beta_1(\mathbb{P}_n), \gamma(\mathbb{P}_n))$, and $\theta_{1,n} = \theta_1(\mathbb{P}_n)$.

The definition of MM-estimators and corresponding functionals in our current setup includes several special cases that are already available in the literature. For the multivariate location and scatter model of Example 3, our MM-functionals $\beta_1(P)$ and $\sigma^2(P)\mathbf{V}(\gamma(P))$ coincide with the multivariate location and scatter M-functionals with auxiliary scale $\sigma(P)$, as discussed in Tatsuoka and Tyler [32]. When, in addition, $\beta_{0,n}$ and $\mathbf{C}_{0,n} = \mathbf{V}(\theta_{0,n})$ are the S-estimators for location and scatter defined by means of ρ_0 , then our MM-estimators $\beta_{1,n}$ and $\sigma_n^2\mathbf{V}(\gamma_n)$ coincide with the MM-estimators for location and scatter considered in Salibián-Barrera *et al* [31]. For the multivariate linear regression model of Example 2, our MM-estimators $\beta_{1,n}$ and $\sigma_n^2\mathbf{V}(\gamma_n)$ coincide with the ones for multivariate linear regression in Kudraszow and Maronna [15]. If, in addition $k = 1$, our regression MM-estimator coincides with the one for multiple linear regression, as introduced in Yohai [36]. Our MM-functionals then coincide with the M-functional with general scale $\sigma(P)$, as treated in Martin *et al* [25].

4 Existence

We will first establish existence of the functionals $\sigma(P)$, $\beta_1(P)$, $\gamma(P)$, and $\theta_1(P)$, under particular conditions on the probability measure P . As a consequence, this will also yield the existence of the estimators σ_n , $\beta_{1,n}$, γ_n , and $\theta_{1,n}$. Recall that the observations are represented as points $(\mathbf{y}_i, \mathbf{X}_i)$ 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, such as in Copt and Heritier [4]. 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

$$P(\mathbf{X} \in \mathcal{X}) = 1. \quad (4.1)$$

Hence, P is then concentrated on the subset $\mathbb{R}^k \times \mathcal{X}$ of $\mathbb{R}^k \times \mathbb{R}^{kq}$, which is of dimension $k + p$, which may be smaller than $k + kq$.

The first condition that we require, expresses the fact that P cannot have too much mass at infinity, in relation to the ratio $r_0 = b_0 / \sup \rho_0$.

(C1 $_\epsilon$) There exists a compact set $K_\epsilon \subset \mathbb{R}^k \times \mathcal{X}$, such that $P(K_\epsilon) \geq r_0 + \epsilon$.

The second condition requires that P cannot have too much mass at arbitrarily thin strips in $\mathbb{R}^k \times \mathcal{X}$. For $\alpha \in \mathbb{R}^{k+kq}$, such that $\|\alpha\| = 1$, $\ell \in \mathbb{R}$, and $\delta \geq 0$, we define a strip $H(\alpha, \ell, \delta)$ as follows:

$$H(\alpha, \ell, \delta) = \{\mathbf{s} \in \mathbb{R}^k \times \mathbb{R}^{kq} : \ell - \delta/2 \leq \alpha^T \mathbf{s} \leq \ell + \delta/2\}. \quad (4.2)$$

Defined in this way, a strip is the area between two parallel hyperplanes which are symmetric around the hyperplane $H(\alpha, \ell, 0) = \{\mathbf{s} \in \mathbb{R}^k \times \mathbb{R}^{kq} : \alpha^T \mathbf{s} = \ell\}$. Since the distance between two parallel hyperplanes $\alpha^T \mathbf{s} = \ell_1$ and $\alpha^T \mathbf{s} = \ell_2$ is $|\ell_1 - \ell_2|$, the strip $H(\alpha, \ell, \delta)$ defined in (4.2) has width δ . We require the following condition.

(C2 $_\epsilon$) The value $\delta_\epsilon = \inf \{\delta : P(H(\alpha, \ell, \delta)) \geq \epsilon, \alpha \in \mathbb{R}^{k+kq}, \|\alpha\| = 1, \ell \in \mathbb{R}, \delta \geq 0\}$ is strictly positive.

According to (4.1), in (C2 $_\epsilon$) one only needs to consider strips in $\mathbb{R}^k \times \mathcal{X}$.

Both conditions are satisfied for any $0 < \epsilon \leq 1 - r_0$ by any probability measure P that is absolutely continuous. Clearly, condition (C1 $_\epsilon$) holds for any $0 \leq \epsilon \leq 1 - r_0$ for the empirical measure $\mathbb{P}_n$ corresponding to a collection of n points $\mathcal{S}_n = \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset \mathbb{R}^k \times \mathcal{X}$. Condition (C2 $_\epsilon$) with $\epsilon = (k + p + 1)/n$ is also satisfied by the empirical measure $\mathbb{P}_n$, when the collection $\mathcal{S}_n$ is in *general position*, i.e., no subset $J \subset \mathcal{S}_n$ of $k + p + 1$ points is contained in the same hyperplane in $\mathbb{R}^k \times \mathcal{X}$. Conditions (C1 $_\epsilon$) and (C2 $_\epsilon$) are the same as in Lopuhaä *et al* [22] and they are similar

to condition (C $_{\epsilon}$) in Lopuhaä [17]. The reason that (C1 $_{\epsilon}$) slightly deviates from Lopuhaä [17], is to handle the presence of $\mathbf{X}$ in minimizing (3.10).

To establish existence of $\sigma(P)$ we follow the reasoning in Yohai [36]. We require the following condition.

(C0) For $E_0 = \{(\mathbf{y}, \mathbf{X}) \in \mathbb{R}^k \times \mathbb{R}^{kq} : \|\mathbf{y} - \mathbf{X}\beta_0(P)\| = 0\}$, it holds $P(E_0) < 1 - b_0/\sup \rho_0$.

We then have the following lemma.

Lemma 1. *Let ρ_0 satisfy (R1)-(R3) and let $(\beta_0(P), \theta_0(P)) \in \mathbb{R}^q \times \Theta$ be the pair of initial functionals at P , such that (C0) holds. Then a solution $\sigma(P) > 0$ to (3.8) exists and is unique.*

To establish the existence of $(\beta_1(P), \gamma(P))$, we follow the reasoning in Lopuhaä *et al* [22]. The idea is to argue that one can restrict oneself to a compact set for finding solutions to minimizing $R_P(\beta, \mathbf{V}(\gamma))$ subject to $|\mathbf{V}(\gamma)| = 1$. When $R_P(\beta, \mathbf{V}(\gamma))$ is continuous, this immediately yields the existence of a minimum. To this end, we assume the following condition.

(V1) The mapping $\theta \mapsto \mathbf{V}(\theta)$ is continuous.

To restrict oneself to (β, γ) in a compact set, we make use of Lemma 4.1 in Lopuhaä *et al* [22]. It requires that the identity is an element of $\mathcal{V} = \{\mathbf{V}(\theta) \in \text{PDS}(k) : \theta \in \Theta \subset \mathbb{R}^l\}$ and that $\mathcal{V}$ is closed under multiplication with a positive scalar.

(V2) There exists a $\theta \in \Theta \subset \mathbb{R}^l$, such that $\mathbf{V}(\theta) = \mathbf{I}_k$. For any $\mathbf{V}(\theta) \in \mathcal{V}$ and any $\alpha > 0$, it holds that $\alpha \mathbf{V}(\theta) = \mathbf{V}(\theta')$, for some $\theta' \in \Theta \subset \mathbb{R}^l$.

Conditions (V1)-(V2) are not very restrictive. For example, all examples discussed in Section 2 satisfy these conditions. Also note that (V2) implies that (3.11) has a solution $\theta_1(P)$ and similarly for (3.6).

Lemma 4.1 in Lopuhaä *et al* [22] will ensure that there exists a compact set in $\mathbb{R}^q \times \text{PDS}(k)$ that contains all pairs $(\beta, \mathbf{V}(\gamma))$ that correspond to possible minima (β, γ) of $R_P(\beta, \mathbf{V}(\gamma))$. To establish that there also exists a compact set in $\mathfrak{D}$ that contains all possible minima (β, γ) of $R_P(\beta, \mathbf{V}(\gamma))$, we need that the pre-image $\{\theta \in \Theta : \mathbf{V}(\theta) \in K\}$ of a compact set $K \subset \mathbb{R}^{k \times k}$ is again compact. Recall that subsets of $\mathbb{R}^l$ are compact if and only if they are closed and bounded, and note that the pre-image of a continuous mapping of a closed set is closed. Hence, in view of condition (V1), it suffices to require the following condition.

(V3) The mapping $\theta \mapsto \mathbf{V}(\theta)$ is such that the pre-image of a bounded set is bounded.

We then have the following theorem.

Theorem 1. *Let ρ_0 and ρ_1 satisfy (R1)-(R2) and (3.3). Suppose ρ_1 is continuous and suppose that $\mathbf{V}$ satisfies (V1)-(V3). Suppose P satisfies (C1 $_{\epsilon}$) and (C2 $_{\epsilon}$), for some $0 < \epsilon \leq 1 - r_0$, where $r_0 = b_0/\sup \rho_0$. Let $(\beta_0(P), \theta_0(P)) \in \mathbb{R}^q \times \Theta$ be the pair of initial functionals at P and let $\sigma(P)$ be a solution to (3.8). Then there exists a pair $(\beta_1(P), \gamma(P)) \in \mathfrak{D}$ that minimizes $R_P(\beta, \mathbf{V}(\gamma))$ and a vector $\theta_1(P) \in \Theta$ that is the unique solution of (3.11).*

Theorem 1 has a direct corollary for the existence of the MM-estimators, when dealing with a collections of points. Let $\mathcal{S}_n = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, with $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$, be a collection of n points in $\mathbb{R}^k \times \mathcal{X}$. Define

$$\kappa(\mathcal{S}_n) = \text{maximal number of points of } \mathcal{S}_n \text{ lying on the same hyperplane in } \mathbb{R}^k \times \mathcal{X}. \quad (4.3)$$

For example, if the distribution P is absolutely continuous, then $\kappa(\mathcal{S}_n) \leq k + p$ with probability one. Existence of σ_n can be obtained from Lemma 1. Suppose that (3.12) holds and that $\#\{i : 1 \leq i \leq n, \|\mathbf{y}_i - \mathbf{X}_i\beta_0(P)\| = 0\} < n(1 - b_0/\sup \rho_0)$. Then $\mathbb{P}_n$ satisfies condition (C0), so that the solution σ_n of (3.2) exists and is unique, according to Lemma 1. We then have the following corollary.

Corollary 1. *Suppose that ρ_0 , ρ_1 , and $\mathbf{V}$ satisfy the conditions of Theorem 1. For a collection $\mathcal{S}_n = \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset \mathbb{R}^k \times \mathcal{X}$, with $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$, for $i = 1, \dots, n$, let $(\beta_{0,n}, \theta_{0,n}) \in \mathbb{R}^q \times \mathbb{R}^l$ be the pair of initial estimators satisfying (3.12) and let σ_n be a solution to (3.2). If $\kappa(\mathcal{S}_n) + 1 \leq n(1 - r_0)$, where $r_0 = b_0 / \sup \rho_0$, then there exists a pair $(\beta_{1,n}, \gamma_n) \in \mathfrak{D}$ that minimizes $R_n(\beta, \mathbf{V}(\gamma))$ and a vector $\theta_{1,n}$ that is the unique solution of (3.6).*

For the multivariate linear regression model of Example 2, Kudraszow and Maronna [15] prove existence of $\beta_{1,n} = \text{vec}(\mathbf{B}_{1,n}^T)$ and $\mathbf{V}(\gamma_n)$, assuming $\kappa(\mathcal{S}_n) < n/2$. Hence, their Theorem 1 follows from our Corollary 1, as long as $r_0 \leq 1/2 - 1/n$. This holds for example, when S-estimators with maximal breakdown point are used as initial estimators (see Theorem 6.1 in Lopuhaä *et al* [22]). Existence of the corresponding functionals is not discussed in Kudraszow and Maronna [15]. This now follows from our Theorem 1. For the multivariate location and scatter model in Example 3, the MM-functionals coincide with the multivariate location and scatter M-functionals defined with loss function ρ_1 and with auxiliary scale $\sigma(P)$, defined as the solution of (3.8). Tatsuoka and Tyler [32] establish existence for these functionals under the assumption

$$\inf_{(\beta, \gamma) \in \mathfrak{D}} R_P(\beta, \mathbf{V}(\gamma)) < (1 - P(B)) \sup \rho_1, \quad (4.4)$$

for all hyperplanes $B \subset \mathbb{R}^k$. It can be seen, using (3.2) and (3.3), that if our condition $(C2_\epsilon)$ holds for some $\epsilon < 1 - r_0$, then condition (4.4) is satisfied.

Existence of MM-estimators has been obtained from the existence of MM-functionals at the empirical measure $\mathbb{P}_n$, which converges to P , as n tends to infinity. The following corollary shows that existence can be established in general, for probability measures that are close to P . This will become useful when we want to establish existence at perturbed measures $(1 - h)P + h\delta_{\mathbf{s}}$, for h sufficiently small, in order to determine the influence function of the functionals at P (see Section 8). It requires the following condition on P .

(C3) Let $\mathfrak{C}$ be the class of all measurable convex subsets of $\mathbb{R}^k \times \mathbb{R}^{kq}$. Every $C \in \mathfrak{C}$ is a P -continuity set, i.e., $P(\partial C) = 0$, where ∂C denotes the boundary of C .

Condition (C3) is needed to apply Theorem 4.2 in Ranga Rao [27]. Clearly, this condition is satisfied if P is absolutely continuous.

Corollary 2. *Suppose that ρ_0 satisfies the conditions of Lemma 1. Let P satisfy (C0) and (C3), and let $(\beta_0(P), \theta_0(P)) \in \mathbb{R}^q \times \Theta$ be the pair of initial functionals at P . 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 $t \rightarrow \infty$. Suppose that $(\beta_0(P_t), \theta_0(P_t))$ exist, for t sufficiently large, such that $\beta_0(P_t) \rightarrow \beta_0(P)$. Then*

(i) *for t sufficiently large, equation (3.8) with $P = P_t$, has a unique solution $\sigma(P_t)$.*

In addition, suppose that ρ_0 , ρ_1 , and $\mathbf{V}$ satisfy the conditions of Theorem 1, and suppose that P satisfies $(C1_{\epsilon'})$ and $(C2_\epsilon)$, for some $0 < \epsilon < \epsilon' \leq 1 - r_0$, where $r_0 = b_0 / \sup \rho_0$. Then

(ii) *for t sufficiently large, there exists $(\beta_1(P_t), \gamma(P_t)) \in \mathfrak{D}$ that minimizes $R_{P_t}(\beta, \mathbf{V}(\gamma))$ and a vector $\theta_1(P_t) \in \Theta$ that is the unique solution of equation (3.11) with $P = P_t$.*

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 $t \rightarrow \infty$. By continuity of the MM-functional $(\beta_1(P), \theta_1(P))$ we mean that $(\beta_1(P_t), \theta_1(P_t)) \rightarrow (\beta_1(P), \theta_1(P))$, as $t \rightarrow \infty$. An example of such a sequence is the sequence of empirical measures $\mathbb{P}_n$, $n = 1, 2, \dots$, that converges weakly to P , almost surely. Continuity of the MM-functional for this sequence would then mean that the MM-estimator $(\beta_{1,n}, \theta_{1,n})$ is consistent, i.e., $(\beta_1(\mathbb{P}_n), \theta_1(\mathbb{P}_n)) \rightarrow (\beta_1(P), \theta_1(P))$, almost surely.

We have the following theorem establishing continuity of the MM-functionals.

Theorem 2. Let ρ_0 satisfy (R2)-(R3) and $\mathbf{V}$ satisfy (V1). Let $(\beta_0(P), \theta_0(P)) \in \mathbb{R}^q \times \Theta$ be the pair of initial functionals at P . 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 $t \rightarrow \infty$. Suppose that $(\beta_0(P_t), \theta_0(P_t))$ exist, for t sufficiently large, and suppose that $(\beta_0(P_t), \theta_0(P_t)) \rightarrow (\beta_0(P), \theta_0(P))$. Let $\sigma(P)$ be the unique solution of (3.8) and let $\sigma(P_t)$ be a solution of (3.8), with $P = P_t$. Then

(i) $\sigma(P_t) \rightarrow \sigma(P)$, as $t \rightarrow \infty$.

In addition, suppose that ρ_1 satisfies (3.3) and (R2)-(R3), and that $\mathbf{V}$ satisfies (V3). Suppose that P satisfies (C3), as well as $(C1_{\epsilon'})$ and $(C2_{\epsilon})$, for some $0 < \epsilon < \epsilon' \leq 1 - r_0$, where $r_0 = b_0 / \sup \rho_0$. For t sufficiently large, let $(\beta_1(P_t), \gamma(P_t)) \in \mathfrak{D}$ be a local minimum of $R_{P_t}(\beta, \mathbf{V}(\gamma))$ that satisfies (3.10) for $P = P_t$, and let $(\beta_1(P), \gamma(P)) \in \mathfrak{D}$ be the unique minimizer of $R_P(\beta, \mathbf{V}(\gamma))$. Then

(ii) $(\beta_1(P_t), \gamma(P_t)) \rightarrow (\beta_1(P), \gamma(P))$, as $t \rightarrow \infty$;

Let $\theta_1(P)$ and $\theta_1(P_t)$ be solutions of (3.11) and (3.11) with $P = P_t$, respectively. Then

(iii) $\theta_1(P_t) \rightarrow \theta_1(P)$, as $t \rightarrow \infty$.

Continuity of the MM-functionals will be used to derive the influence function of the MM-functionals in Section 8. Another convenient consequence of the continuity of the MM-functionals is that one can directly obtain consistency of the MM-estimators. Let $\mathcal{S}_n = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, with $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$, be a collection of n points in $\mathbb{R}^k \times \mathcal{X}$. We apply Theorem 2 to the sequence $\mathbb{P}_n$, $n = 1, 2, \dots$, of probability measures, where $\mathbb{P}_n$ is the empirical measure corresponding to $\mathcal{S}_n$.

Corollary 3. Let ρ_0 and $\mathbf{V}$ satisfy the conditions of Theorem 2(i). For a collection $\mathcal{S}_n = \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset \mathbb{R}^k \times \mathcal{X}$, with $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$, for $i = 1, \dots, n$, let $(\beta_{0,n}, \theta_{0,n}) \in \mathbb{R}^q \times \mathbb{R}^l$ be the pair of initial estimators satisfying (3.12) and suppose that $(\beta_{0,n}, \theta_{0,n}) \rightarrow (\beta_0(P), \theta_0(P))$, with probability one. Let $\sigma(P)$ be the unique solution of (3.8) and let σ_n be a solution of (3.2). Then

(i) $\sigma_n \rightarrow \sigma(P)$, with probability one.

In addition, suppose ρ_1 , $\mathbf{V}$, and P satisfy the conditions of Theorem 2(ii). Let $(\beta_{1,n}, \gamma_n) \in \mathfrak{D}$ be a local minimum of $R_n(\beta, \mathbf{V}(\gamma))$ that satisfies (3.5), and let $(\beta_1(P), \gamma(P)) \in \mathfrak{D}$ be the unique minimizer of $R_P(\beta, \mathbf{V}(\gamma))$. Then

(ii) $(\beta_{1,n}, \gamma_n) \rightarrow (\beta_1(P), \gamma(P))$, with probability one;

Let $\theta_1(P)$ and $\theta_{1,n}$ be solutions of (3.11) and (3.6), respectively. Then

(iii) $\theta_{1,n} \rightarrow \theta_1(P)$, with probability one.

When $\mathbf{V}$ also satisfies (V1), then for the covariance MM-estimator it follows from Corollary 3 that $\mathbf{V}(\theta_{1,n}) \rightarrow \mathbf{V}(\theta_1(P)) = \sigma^2(P)\mathbf{V}(\gamma(P))$, with probability one. This extends Theorem 5 in Kudraszow and Maronna [15]. Their result applies to MM-estimators for the multivariate models in Examples 2 and 3, but is obtained only for distributions with an elliptical contoured density.

Theorem 2 and Corollary 3 require that $(\beta_1(P), \gamma(P)) \in \mathfrak{D}$ uniquely minimizes $R_P(\beta, \mathbf{V}(\gamma))$. This situation is very similar to that of multivariate location-scatter M-estimators with auxiliary scale, considered by Tatsuoka and Tyler [32]. For the special case that $\mathbf{X} = \mathbf{I}_k$, their Theorem 4.2 shows that $R_P(\beta, \mathbf{C})$ has a unique minimum for a broad class of distributions, consisting of affine transformations of distributions on $\mathbb{R}^k$, which are invariant under permutations and sign changes of its components and which have densities g such that $g \circ \exp$ is Schur-concave (see [32] for details), i.e.,

$$f_{\mu, \Sigma}(\mathbf{y}) = |\Sigma|^{-1/2} g(\Sigma^{-1/2}(\mathbf{y} - \mu)). \quad (5.1)$$

The next theorem is a direct consequence of that result. Note that elliptically contoured densities are special cases of (5.1). Let $\mathbb{E}_{\mu, \Sigma}$ denote the expectation with respect to $f_{\mu, \Sigma}$.

Theorem 3. Let ρ_0 satisfy (R1)-(R3) and suppose ρ_1 is continuous and satisfies (R2) and (3.3). Suppose that P is absolutely continuous, such that for some $(\beta^*, \theta^*) \in \mathbb{R}^q \times \Theta$, for all $\mathbf{X}$, the distribution of $\mathbf{y} \mid \mathbf{X}$ has density $f_{\mu, \Sigma}$ from (5.1), with $\mu = \mathbf{X}\beta^*$ and $\Sigma = \mathbf{V}(\theta^*)$. Suppose that g in (5.1) is strictly M -concave (see [32, Definition 4.4]). Suppose $\mathbf{V}$ satisfies (V1)-(V3) and suppose $\mathbf{X}$ has full rank with probability one. Let $(\beta_0(P), \theta_0(P)) \in \mathbb{R}^q \times \Theta$ be the pair of initial functionals at P satisfying $(\beta_0(P), \theta_0(P)) = (\beta^*, \theta^*)$. Then, the following holds with probability one.

- (i) Equation (3.8) has a unique solution $\sigma(P)$ and the function $R_P(\beta, \mathbf{V}(\gamma))$ has a unique minimum $(\beta_1(P), \gamma(P)) \in \mathfrak{D}$, that satisfies $\beta_1(P) = \beta^*$ and $\mathbf{V}(\gamma(P)) = \Sigma/|\Sigma|^{1/k}$.
- (ii) When $\mathbf{V}(\alpha\theta) = \alpha\mathbf{V}(\theta)$, for all $\alpha > 0$, then $\theta_1(P) = \theta^* \sigma^2(P)/|\Sigma|^{1/k}$.
- (iii) When $b_0 = \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \rho_0(\|\mathbf{z}\|)$, then $\sigma(P) = |\Sigma|^{1/(2k)}$.

An example of initial functionals $(\beta_0(P), \theta_0(P))$ that satisfy the conditions of Theorem 3, are the S-functionals defined with loss function ρ_0 , see Theorem 5.3 in Lopuhaä *et al* [22] or Theorem 1 in Davies [7] for the multivariate location-scatter model.

The proof of Theorem 3 depends heavily on the application of Theorem 4.2 in Tatsuoka and Tyler [32] on the uniqueness of multivariate M-functionals with auxiliary scale. It considers strict M -concave densities g in (5.1), which is a broad class of densities that includes spherical symmetric densities, among others, see Tatsuoka and Tyler [32] for details. In this way, Theorem 3 can be seen as an extension of Theorem 1 in Davies [7] on the uniqueness of multivariate location-scatter S-functionals at distributions with an elliptically contoured density.

6 Global robustness: the breakdown point

Consider a collection of points $\mathcal{S}_n = \{\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i), i = 1, \dots, n\} \subset \mathbb{R}^k \times \mathcal{X}$. To investigate the global robustness of the estimators, we compute their finite-sample (replacement) breakdown point. For a given collection $\mathcal{S}_n$, the finite-sample breakdown point (see Donoho and Huber [9]) of an estimator is defined as the smallest proportion of points from $\mathcal{S}_n$ that one needs to replace in order to send the estimator to the boundary of its parameter space. To emphasize the dependence on the collection $\mathcal{S}_n$, denote an estimator for the regression parameter by $\beta_n(\mathcal{S}_n)$ and an estimator for the vector of covariance parameters by $\theta_n(\mathcal{S}_n)$. For a given collection $\mathcal{S}_n$, the finite-sample breakdown point of a regression estimator β_n is defined as

$$\epsilon_n^*(\beta_n, \mathcal{S}_n) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{\mathcal{S}'_m} \|\beta_n(\mathcal{S}_n) - \beta_n(\mathcal{S}'_m)\| = \infty \right\}, \quad (6.1)$$

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 $\mathbb{R}^k \times \mathcal{X}$.

An estimator θ_n for the vector of covariance parameters determines the covariance estimator $\mathbf{V}(\theta_n)$. For this reason it seems natural to let the breakdown point of θ_n correspond to the breakdown of a covariance estimator. For any $k \times k$ matrix $\mathbf{A}$, let $\lambda_k(\mathbf{A}) \leq \dots \leq \lambda_1(\mathbf{A})$ denote the eigenvalues of $\mathbf{A}$. We define the finite sample (replacement) breakdown point of an estimator θ_n at a collection $\mathcal{S}_n$, as

$$\epsilon_n^*(\theta_n, \mathcal{S}_n) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{\mathcal{S}'_m} \text{dist}(\mathbf{V}(\theta_n(\mathcal{S}_n)), \mathbf{V}(\theta_n(\mathcal{S}'_m))) = \infty \right\}, \quad (6.2)$$

with $\text{dist}(\cdot, \cdot)$ defined as $\text{dist}(\mathbf{A}, \mathbf{B}) = \max\{|\lambda_1(\mathbf{A}) - \lambda_1(\mathbf{B})|, |\lambda_k(\mathbf{A})^{-1} - \lambda_k(\mathbf{B})^{-1}|\}$, 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 $\mathbb{R}^k \times \mathcal{X}$. So the breakdown point of θ_n is the smallest proportion of points from $\mathcal{S}_n$ that one needs to replace in order to make the largest eigenvalue of $\mathbf{V}(\theta(\mathcal{S}'_m))$

arbitrarily large (explosion), or to make the smallest eigenvalue of $\mathbf{V}(\boldsymbol{\theta}(\mathcal{S}'_m))$ arbitrarily small (implosion).

Good global robustness is illustrated by a high breakdown point. The breakdown point of the MM-estimators is given in the theorem below.

Theorem 4. *Let ρ_0 satisfy (R1)-(R3). Let ρ_1 satisfy (R2) and (3.3) and suppose $\mathbf{V}$ satisfies (V1)-(V3). Let $\mathcal{S}_n \subset \mathbb{R}^k \times \mathcal{X}$ be a collection of n points $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$, $i = 1, \dots, n$. Let $r_0 = b_0 / \sup \rho_0$ and suppose that $0 < \lfloor nr_0 \rfloor < n - \kappa(\mathcal{S}_n)$, where $\kappa(\mathcal{S}_n)$ is defined by (4.3). Let $(\boldsymbol{\beta}_{0,n}, \boldsymbol{\theta}_{0,n})$ be initial estimators for $(\boldsymbol{\beta}, \boldsymbol{\theta})$. Let $(\boldsymbol{\beta}_{1,n}, \boldsymbol{\gamma}_n) \in \mathfrak{D}$ satisfy (3.5) and let $\boldsymbol{\theta}_{1,n}$ be a solution of (3.6). Then*

$$\epsilon_n^*(\boldsymbol{\beta}_{1,n}, \boldsymbol{\gamma}_n, \boldsymbol{\theta}_{1,n}, \mathcal{S}_n) \geq \min \left\{ \epsilon_n^*(\boldsymbol{\beta}_{0,n}, \boldsymbol{\theta}_{0,n}, \mathcal{S}_n), \frac{\lfloor nr_0 \rfloor}{n}, \frac{\lfloor n - nr_0 \rfloor - \kappa(\mathcal{S}_n)}{n} \right\}.$$

An example of initial estimators $(\boldsymbol{\beta}_{0,n}, \boldsymbol{\theta}_{0,n})$ with high breakdown point, are S-estimators defined with the function ρ_0 , as discussed in Lopuhaä *et al* [22]. According to their Theorem 6.1 and Remark 3, it holds that $\epsilon_n^*(\boldsymbol{\beta}_{0,n}, \boldsymbol{\theta}_{0,n}, \mathcal{S}_n) \geq \min\{\lfloor nr_0 \rfloor, \lfloor n - nr_0 \rfloor - \kappa(\mathcal{S}_n)\}/n$. In this case, the lower bound in Theorem 4 simplifies to $\min\{\lfloor nr_0 \rfloor, \lfloor n - nr_0 \rfloor - \kappa(\mathcal{S}_n)\}/n$. The largest possible value of this lower bound is attained when $r_0 = (n - \kappa(\mathcal{S}_n))/(2n)$. In this case $\lfloor nr_0 \rfloor = \lfloor n - nr_0 \rfloor - \kappa(\mathcal{S}_n) = \lceil (n - \kappa(\mathcal{S}_n))/2 \rceil = \lfloor (n - \kappa(\mathcal{S}_n) + 1)/2 \rfloor$. When the collection $\mathcal{S}_n$ is in general position, then $\kappa(\mathcal{S}_n) = k + p$. In that case the breakdown point of the MM-estimators is at least equal to $\lfloor (n - k - p + 1)/2 \rfloor / n$. When all $\mathbf{X}_i$ are equal to the same $\mathbf{X}$, as in the multivariate location-scatter model, but also in the linear mixed effects models considered in Copt and Victoria-Feser [5] and Copt and Heritier [4], one has $p = 0$ and $\kappa(\mathcal{S}_n) = k$. In that case, the lower bound of the breakdown point is equal to $\lfloor (n - k + 1)/2 \rfloor / n$. This value coincides with the maximal breakdown point for affine equivariant estimators for $k \times k$ covariance matrices (see Davies [7, Theorem 6]).

The breakdown point for (a simpler version of) regression MM-estimators for the linear mixed effects model (2.3) has only been discussed in Copt and Heritier [4]. They conjecture that the exact value can be derived using the technique in Van Aelst and Willems [35], but do not pursue a rigorous derivation. The result in Theorem 4 applies to the current more extensive version of MM-estimators for the linear mixed effects model (2.3). Furthermore, for $0 < r_0 \leq (n - \kappa(\mathcal{S}_n))/(2n)$, it holds that $\lfloor nr_0 \rfloor \leq \lfloor n - nr_0 \rfloor - \kappa(\mathcal{S}_n)$. In this case, the lower bound for the breakdown point in Theorem 4 coincides with that of the regression MM-estimator considered in Lopuhaä [19].

For the multivariate linear regression model, Kudraszow and Maronna [15] take $r_0 = 1/2$ and consider the case $\kappa(\mathcal{S}_n) < n/2$. For this situation $\lfloor nr_0 \rfloor > \lfloor n - nr_0 \rfloor - \kappa(\mathcal{S}_n)$. Hence, their Theorem 3 follows from our Theorem 4 for the case $r_0 = 1/2$. For the multivariate location-scatter model, Salibián-Barrera *et al* [31] consider MM-estimators with S-estimators as initial estimators. Our Theorem 4 then coincides with their Theorem 1. For the MM-estimators in this model, Tyler [34] considers the gross error breakdown point, which for finite collections is related to the finite sample contamination breakdown point.

7 Score equations

Up to this point, properties of MM-functionals and MM-estimators have been derived from minimizing $R_n(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$ and $R_P(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$, as defined in (3.4) and (3.9), respectively. To obtain the influence function and to establish the limiting distribution of MM-estimators, we use the score equations that can be found by differentiation of the Lagrangian corresponding to the constrained minimization problem. To this end, we require the following additional condition on the function ρ_1 ,

(R4) ρ_1 is continuously differentiable and $u_1(s) = \rho'_1(s)/s$ is continuous,

and the following condition on the mapping $\boldsymbol{\theta} \mapsto \mathbf{V}(\boldsymbol{\theta})$,

(V4) $\mathbf{V}(\boldsymbol{\theta})$ is continuously differentiable.

Obviously, condition (V4) implies the former condition (V1). For $\mathbf{y} \in \mathbb{R}^k$, $\mathbf{t} \in \mathbb{R}^k$, and $\mathbf{C} \in \text{PDS}(k)$, define the Mahalanobis distances by

$$d^2(\mathbf{y}, \mathbf{t}, \mathbf{C}) = (\mathbf{y} - \mathbf{t})^T \mathbf{C}^{-1} (\mathbf{y} - \mathbf{t}). \quad (7.1)$$

We then have the following proposition.

Proposition 1. *Let ρ_1 satisfy (R2) and (R4), and $\mathbf{V}$ satisfy (V4), and suppose that $\mathbb{E}\|\mathbf{X}\| < \infty$. Let $(\beta_0(P), \theta_0(P))$ be the pair of initial functionals and let $\sigma(P)$ be a solution of (3.8). Then any local minimum $\xi(P) = (\beta_1(P), \gamma(P)) \in \mathfrak{D}$ of $R_P(\beta, \mathbf{V}(\gamma))$ satisfies*

$$\int \Psi(\mathbf{s}, \xi, \sigma(P)) dP(\mathbf{s}) = \mathbf{0}, \quad (7.2)$$

where $\Psi = (\Psi_\beta, \Psi_\gamma)$, with Ψ_β and $\Psi_\gamma = (\Psi_{\gamma,1}, \dots, \Psi_{\gamma,l})$ given by

$$\begin{aligned} \Psi_\beta(\mathbf{s}, \xi, \sigma) &= u_1 \left(\frac{d}{\sigma} \right) \mathbf{X}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) \\ \Psi_{\gamma,j}(\mathbf{s}, \xi, \sigma) &= u_1 \left(\frac{d}{\sigma} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \mathbf{H}_{1,j} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) \\ &\quad - \text{tr} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) \log |\mathbf{V}|, \end{aligned} \quad (7.3)$$

with $d = d(\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\gamma))$, as defined in (7.1), and $\mathbf{V}(\gamma)$ is abbreviated by $\mathbf{V}$, and where

$$\mathbf{H}_{1,j} = \text{tr} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) \left(\sum_{t=1}^l \gamma_t \frac{\partial \mathbf{V}}{\partial \gamma_t} \right) - \text{tr} \left(\mathbf{V}^{-1} \sum_{t=1}^l \gamma_t \frac{\partial \mathbf{V}}{\partial \gamma_t} \right) \frac{\partial \mathbf{V}}{\partial \gamma_j}, \quad (7.4)$$

for $j = 1, \dots, l$.

Since $d(\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\gamma))/\sigma = d(\mathbf{y}, \mathbf{X}\beta, \sigma^2 \mathbf{V}(\gamma))$, the regression score equation for Ψ_β is similar to the one for the regression MM-estimator considered in Lopuhaä [19], defined with initial covariance functional $\mathbf{V}_0(P)$, and both score equations coincide when $\mathbf{V}_0(P) = \sigma^2(P) \mathbf{V}(\gamma(P))$. Similarly, the regression score equation for Ψ_β with the empirical measure $\mathbb{P}_n$ for P in (7.2) is similar to equation (8) for the regression MM-estimator in the linear mixed effects model considered in Copt and Heritier [4], defined with initial covariance estimator $\widehat{\Sigma}_S$, and both equations coincide when $\widehat{\Sigma}_S = \sigma_n^2 \mathbf{V}(\gamma_n)$. For the multivariate linear regression model, the empirical score equation for Ψ_β coincides with equation (2.10) for the regression MM-estimator discussed in Kudraszow and Maronna [15]. When the initial estimators $(\beta_{0,n}, \theta_{0,n})$ are S-estimators, then $\sigma_n = |\mathbf{V}(\theta_{0,n})|^{1/(2k)}$. In that case, for the multivariate location-scatter model, the empirical score equation for Ψ_β coincides with fixed point equation (16) for the location MM-estimator considered in Salibián-Barrera *et al* [31].

The function Ψ_γ simplifies in situations where $\mathbf{V}(\gamma)$ has a linear structure, that is

$$\mathbf{V}(\gamma) = \sum_{j=1}^l \gamma_j \mathbf{L}_j, \quad (7.5)$$

The covariance structures in Examples 1, 2, and 3, satisfy this property.

Proposition 2. *Suppose the conditions of Proposition 1 hold and that $\mathbf{V}$ has a linear structure (7.5). Let $v_1(s) = \rho'_1(s)s$ and let $\mathbf{L}$ be the $k^2 \times l$ matrix*

$$\mathbf{L} = \begin{bmatrix} \text{vec}(\mathbf{L}_1) & \cdots & \text{vec}(\mathbf{L}_l) \end{bmatrix}. \quad (7.6)$$

Then any local minimum $\xi(P) = (\beta_1(P), \gamma(P)) \in \mathfrak{D}$ of $R_P(\beta, \mathbf{V}(\gamma))$ satisfies (7.2), with $\Psi = (\Psi_\beta, \Psi_\gamma)$, where

$$\begin{aligned}\Psi_\beta(\mathbf{s}, \xi, \sigma) &= u_1 \left(\frac{d}{\sigma} \right) \mathbf{X}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) \\ \Psi_\gamma(\mathbf{s}, \xi, \sigma) &= -\mathbf{L}^T (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \text{vec}(\Psi_{\mathbf{V}}(\mathbf{s}, \xi, \sigma)),\end{aligned}\tag{7.7}$$

where

$$\Psi_{\mathbf{V}}(\mathbf{s}, \xi, \sigma) = ku_1 \left(\frac{d}{\sigma} \right) (\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)^T - v_1 \left(\frac{d}{\sigma} \right) \sigma^2 \mathbf{V} - \mathbf{V} \log |\mathbf{V}|,\tag{7.8}$$

with $d = d(\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\gamma))$, as defined in (7.1), and $\mathbf{V}(\gamma)$ is abbreviated by $\mathbf{V}$.

For the multivariate linear regression model in Example 2, one has $\mathbf{V}(\gamma) = \mathbf{\Gamma} = \Sigma/|\Sigma|^{1/k}$, with $\gamma = \text{vech}(\mathbf{\Gamma})$. The matrix $\mathbf{L} = \partial \text{vec}(\mathbf{V}) / \partial \gamma^T$ is then equal to the so-called duplication matrix $\mathcal{D}_k$, which is the unique $k^2 \times k(k+1)/2$ matrix, with the properties $\mathcal{D}_k \text{vech}(\mathbf{C}) = \text{vec}(\mathbf{C})$ and $(\mathcal{D}_k^T \mathcal{D}_k)^{-1} \mathcal{D}_k^T \text{vec}(\mathbf{C}) = \text{vech}(\mathbf{C})$ (e.g., see [24, Ch. 3, Sec. 8]). Because $\mathbf{V}(\gamma)$ has full rank, it follows that equation (7.2) holds for $\Psi = (\Psi_\beta, \Psi_{\mathbf{V}})$. When we insert $\sigma_n^2 \mathbf{V}(\gamma_n) = \sigma_n^2 \mathbf{\Gamma}_n$, the resulting score equations for the empirical measure $\mathbb{P}_n$ corresponding to observations $(\mathbf{y}_i, \mathbf{X}_i)$, for $i = 1, \dots, n$, are then equivalent to equations (2.10) and (2.11) found in Kudraszow and Maronna [15] for the regression MM-estimators. For the multivariate location-scatter model in Example 3, one also has $\mathbf{V}(\gamma) = \mathbf{\Gamma}$, with $\gamma = \text{vech}(\mathbf{\Gamma})$, so that again equation (7.2) holds for $\Psi = (\Psi_\beta, \Psi_{\mathbf{V}})$. When the initial estimators $(\beta_{0,n}, \theta_{0,n})$ are S-estimators, it can be shown that the empirical score equation for $\Psi_{\mathbf{V}}$ coincides with fixed point equation (17) for the covariance shape MM-estimator considered in Salibián-Barrera *et al* [31].

8 Local robustness: the influence function

For $0 < h < 1$ and $\mathbf{s} = (\mathbf{y}, \mathbf{X}) \in \mathbb{R}^k \times \mathbb{R}^{kq}$ fixed, define the perturbed probability measure $P_{h,\mathbf{s}} = (1-h)P + h\delta_{\mathbf{s}}$, where $\delta_{\mathbf{s}}$ denotes the Dirac measure at $\mathbf{s} \in \mathbb{R}^k \times \mathbb{R}^{kq}$. The *influence function* of a functional $T(\cdot)$ at probability measure P , is defined as

$$\text{IF}(\mathbf{s}; T, P) = \lim_{h \downarrow 0} \frac{T((1-h)P + h\delta_{\mathbf{s}}) - T(P)}{h},\tag{8.1}$$

if this limit exists. In contrast to the global robustness measured by the breakdown point, the influence function measures the local robustness. It describes the effect of an infinitesimal contamination at a single point $\mathbf{s}$ on the functional (see Hampel [11]). Good local robustness is therefore illustrated by a bounded influence function.

8.1 The general case

We will investigate when the limit in (8.1) exists for the functionals σ , $\xi = (\beta_1, \gamma)$, and θ_1 and derive their expression at general P . Since the value of θ_1 determines the covariance matrix $\mathbf{V}(\theta_1)$, we also include the influence function of the covariance functional. Consider the MM-functional at $P_{h,\mathbf{s}}$. From the Portmanteau theorem [2, Theorem 2.1] it can easily be seen that $P_{h,\mathbf{s}} \rightarrow P$, weakly, as $h \downarrow 0$. Therefore, under the conditions of Corollary 2 and Theorem 2, it follows that there exist a unique solution $\sigma(P_{h,\mathbf{s}})$ of equation (3.8) with $P = P_{h,\mathbf{s}}$, a pair $\xi(P_{h,\mathbf{s}}) = (\beta_1(P_{h,\mathbf{s}}), \gamma(P_{h,\mathbf{s}}))$ that minimizes $R_{P_{h,\mathbf{s}}}(\beta, \mathbf{V}(\gamma))$ over $(\beta, \gamma) \in \mathfrak{D}$, and a vector $\theta_1(P_{h,\mathbf{s}}) \in \Theta$ that uniquely solves (3.10), for $P = P_{h,\mathbf{s}}$. Moreover, under these conditions $(\sigma(P_{h,\mathbf{s}}), \xi(P_{h,\mathbf{s}}), \theta_1(P_{h,\mathbf{s}})) \rightarrow (\sigma(P), \xi(P), \theta_1(P))$, as $h \downarrow 0$.

For $\xi = (\beta, \gamma) \in \mathfrak{D}$ and $\sigma > 0$, define

$$\Lambda(\xi, \sigma) = \int \Psi(\mathbf{s}, \xi, \sigma) dP(\mathbf{s}),\tag{8.2}$$

where $\Psi = (\Psi_\beta, \Psi_\gamma)$ is defined in (7.3), and write $\Lambda = (\Lambda_\beta, \Lambda_\gamma)$, where

$$\Lambda_\beta(\xi, \sigma) = \int \Psi_\beta(\mathbf{s}, \xi, \sigma) dP(\mathbf{s}), \quad \Lambda_\gamma(\xi, \sigma) = \int \Psi_\gamma(\mathbf{s}, \xi, \sigma) dP(\mathbf{s}). \quad (8.3)$$

By definition of the variance components MM-functional θ_1 , it can be expected that for general P , the influence function of θ_1 depends on the influence functions of σ and γ . Furthermore, because the function Ψ depends on σ , it can be expected that for general P , the influence function of $\xi = (\beta_1, \gamma)$ depends on the influence function of σ , see Lemma 2 in the supplemental material [21]. In turn, since the functional σ is defined as a solution of (3.8), it can be expected that for general P , its influence function will depend on the initial functionals $\zeta_0 = (\beta_0, \theta_0)$, see Lemma 3 in the supplemental material [21].

The situation becomes a bit simpler if we assume some kind of symmetry of the distribution P . 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 $P(\mathbf{X}\mathbf{b} + A \mid \mathbf{X}) = P(\mathbf{X}\mathbf{b} - A \mid \mathbf{X})$, for all measurable sets $A \subset \mathbb{R}^k$, where for $\lambda \in \mathbb{R}$ and $\mathbf{b} \in \mathbb{R}^q$, $\mathbf{X}\mathbf{b} + \lambda A$ denotes the set $\{\mathbf{X}\mathbf{b} + \lambda \mathbf{y} : \mathbf{y} \in A\}$. If $\mathbf{b}$ is a point of symmetry of P , it has the property that $\mathbb{E}[G(\mathbf{y} - \mathbf{X}\mathbf{b})] = \mathbf{0}$, for any $G(\mathbf{z})$, which is an odd function of $\mathbf{z} \in \mathbb{R}^k$. Furthermore, in order to obtain simpler expressions we also require the following condition on the function ρ_1 .

(R5) ρ_1 is twice continuously differentiable.

Condition (R5) is needed to ensure that $\partial\Lambda_\beta/\partial\beta$ is continuous at $(\xi(P), \sigma(P))$. The expressions for $\text{IF}(\mathbf{s}; \gamma, P)$ and $\text{IF}(\mathbf{s}; \sigma, P)$ given in Lemmas 2 and 3 in [21] do simplify, but without further knowledge on the influence function of ζ_0 we can still not provide an explicit expression. The situation is different for β_1 , for which the influence function is given in the next theorem.

Theorem 5. *Suppose ρ_1 satisfies (R4) and $\mathbf{V}$ satisfies (V1). Let $\sigma(P)$ be a solution of (3.8) and let $\xi(P) \in \mathfrak{D}$ be a local minimum of $R_P(\beta, \mathbf{V}(\gamma))$. Let $\sigma(P_{h,\mathbf{s}})$ be a solution of (3.8) with $P = P_{h,\mathbf{s}}$ and let $\xi(P_{h,\mathbf{s}}) \in \mathfrak{D}$ be a local minimum of $R_P(\beta, \mathbf{V}(\gamma))$ with $P = P_{h,\mathbf{s}}$. Suppose that $(\xi(P_{h,\mathbf{s}}), \sigma(P_{h,\mathbf{s}})) \rightarrow (\xi(P), \sigma(P))$, as $h \downarrow 0$. Let Λ_β be defined by (8.3) with Ψ_β from (7.3) and suppose Λ_β is continuously differentiable with a non-singular derivative $\mathbf{D}_\beta = \partial\Lambda_\beta/\partial\beta$ at $(\xi(P), \sigma(P))$. Moreover, suppose that $\beta_1(P)$ is a point of symmetry of P . Then, for $\mathbf{s} \in \mathbb{R}^k \times \mathbb{R}^{kq}$, we have $\text{IF}(\mathbf{s}; \beta_1, P) = -\mathbf{D}_\beta^{-1} \Psi_\beta(\mathbf{s}, \xi(P), \sigma(P))$.*

The expression of $\text{IF}(\mathbf{s}; \beta_1, P)$ in Theorem 5 is similar to that of the influence function of the regression MM-functional considered in Lopuhaä [19] defined with ρ_1 and initial covariance functional $\mathbf{V}_0$, and both expressions coincide when $\mathbf{V}_0(P_{h,\mathbf{s}}) \rightarrow \sigma^2(P)\mathbf{V}(\gamma(P))$.

If one is allowed to interchange integration and differentiation in $\mathbf{D}_\beta$, then the expression for $\text{IF}(\mathbf{s}_0; \beta_1, P)$ in Corollary 5 coincides with that of the regression MM-functional in the multiple linear regression model considered in Yohai [36]. For linear mixed effects models, multivariate linear regression models, or multivariate location-scatter model, expressions for the influence function of MM-functionals are either not available or are restricted to distributions with an elliptically contoured density. This situation is discussed in the next section for model (2.1).

8.2 Elliptically contoured densities

We can obtain an even more detailed expression for the influence functions, when $\mathbf{V}$ has a linear structure and P has the following property.

(E) There exists $(\beta^*, \theta^*) \in \mathbb{R}^q \times \Theta$, such that for all $\mathbf{X}$, the distribution of $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density

$$f_{\mu, \Sigma}(\mathbf{y}) = |\Sigma|^{-1/2} m((\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu)), \quad (8.4)$$

with $\mu = \mathbf{X}\beta^*$ and $\Sigma = \mathbf{V}(\theta^*)$ and $m : [0, \infty) \rightarrow [0, \infty)$.

We require the following condition on the mapping $\theta \mapsto \mathbf{V}(\theta)$.

(V5) $\mathbf{V}(\boldsymbol{\theta})$ is twice continuously differentiable.

Condition (V5) is needed to interchange the order of integration and differentiation in $\partial\Lambda/\partial\xi$, where Λ is defined in (8.2). Clearly, condition (V5) implies former conditions (V4) and (V1).

When the MM-functionals are affine equivariant, then it suffices to determine the influence function for the case $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{0}, \mathbf{I}_k)$. However, this does not hold in general for the MM-functionals in our setting. Nevertheless, for the general case with $\boldsymbol{\mu} \in \mathbb{R}^k$ and $\boldsymbol{\Sigma} \in \text{PDS}(k)$, we can still use the fact that, conditionally on $\mathbf{X}$, the distribution of $\mathbf{y}$ is the same as that of $\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}$, where $\mathbf{z}$ has a spherical density $f_{\mathbf{0}, \mathbf{I}_k}$. Let

$$\begin{aligned}\alpha_1 &= \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\left(1 - \frac{1}{k}\right) \frac{\rho'_1(c_\sigma \|\mathbf{z}\|)}{c_\sigma \|\mathbf{z}\|} + \frac{1}{k} \rho''_1(c_\sigma \|\mathbf{z}\|) \right], \\ \gamma_1 &= \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho''_1(c_\sigma \|\mathbf{z}\|)(c_\sigma \|\mathbf{z}\|)^2 + (k+1)\rho'_1(c_\sigma \|\mathbf{z}\|)c_\sigma \|\mathbf{z}\|]}{k+2},\end{aligned}\tag{8.5}$$

where $c_\sigma = |\boldsymbol{\Sigma}|^{1/(2k)}/\sigma(P)$. The influence functions of the MM-functionals β_1 and γ at distributions P , such that $\mathbf{y} \mid \mathbf{X}$ has an elliptical contoured density, are now given by the following theorem.

Theorem 6. *Suppose P satisfies (E) for some $(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*) \in \mathbb{R}^q \times \boldsymbol{\Theta}$ and $\mathbb{E}\|\mathbf{X}\|^2 < \infty$. Let ρ_1 satisfy (R2), (R4)-(R5) and suppose $\mathbf{V}$ satisfies (V5) and has a linear structure (7.5). Let $\sigma(P)$ be a solution of (3.8) and let $\boldsymbol{\xi}(P) \in \mathfrak{D}$ be a local minimum of $R_P(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$ that satisfies $\beta_1(P) = \boldsymbol{\beta}^*$ and $\mathbf{V}(\boldsymbol{\gamma}(P)) = \boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/k}$. Let $\sigma(P_{h,s})$ be a solution of (3.8) with $P = P_{h,s}$ and let $\boldsymbol{\xi}(P_{h,s}) \in \mathfrak{D}$ be a local minimum of $R_P(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$ with $P = P_{h,s}$. Suppose that $(\boldsymbol{\xi}(P_{h,s}), \sigma(P_{h,s})) \rightarrow (\boldsymbol{\xi}(P), \sigma(P))$, as $h \downarrow 0$. Let α_1 be defined in (8.5) and suppose that $\alpha_1 \neq 0$. Suppose that $\mathbf{X}$ has full rank with probability one. Then, for $\mathbf{s}_0 \in \mathbb{R}^k \times \mathbb{R}^{kq}$,*

$$\text{IF}(\mathbf{s}_0; \beta_1, P) = \frac{u_1(c_\sigma \|\mathbf{z}_0\|)}{\alpha_1} (\mathbb{E}[\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}])^{-1} \mathbf{X}_0^T \boldsymbol{\Sigma}^{-1/2} \mathbf{z}_0,$$

where $u_1(s) = \rho'_1(s)/s$, $\mathbf{z}_0 = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_0 - \mathbf{X}_0 \boldsymbol{\beta}^*)$, and $c_\sigma = |\boldsymbol{\Sigma}|^{1/(2k)}/\sigma(P)$. In addition, suppose that $\text{IF}(\mathbf{s}; \sigma, P)$ exists. Let γ_1 be defined in (8.5) and suppose that $\gamma_1 > 0$. Suppose the $k^2 \times l$ matrix $\mathbf{L}$, as defined in (7.6), has full rank. Then, for $\mathbf{s}_0 \in \mathbb{R}^k \times \mathbb{R}^{kq}$,

$$\begin{aligned}\text{IF}(\mathbf{s}_0; \gamma, P) &= \frac{ku_1(c_\sigma \|\mathbf{z}_0\|)}{\sigma^2(P)\gamma_1} (\mathbf{L}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{L})^{-1} \mathbf{L}^T (\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) \\ &\quad - \frac{v_1(c_\sigma \|\mathbf{z}_0\|)}{|\boldsymbol{\Sigma}|^{1/k}\gamma_1} \boldsymbol{\theta}^*,\end{aligned}$$

where $v_1(s) = \rho'_1(s)s$.

When $c_\sigma = 1$ (e.g., see Theorem 3(iii)), then $\text{IF}(\mathbf{s}_0; \beta_1, P)$ coincides with the influence function of the regression MM-functional considered in Lopuhaä [19] defined with ρ_1 and an initial Fisher consistent covariance functional $\mathbf{V}_0$. Moreover, it also coincides with the influence function of the regression S-functional defined with ρ_1 for model (2.1), see Corollary 8.4 in Lopuhaä *et al* [22]. This confirms the claim made by Salibián-Barrera *et al* [31] about the influence function of the location MM-functional in the model of Example 3.

For the linear mixed effects model (2.3), Copt and Heritier [4] discuss the influence function, but an expression is not provided. The expression for the influence function of the regression MM-functional now follows from Theorem 6. For the multivariate linear regression model of Example 2, the expression for $\text{IF}(\mathbf{s}_0; \beta_1, P)$ in Theorem 6 coincides with one found for the regression MM-functional in Kudraszow and Maronna [15]. Since the multivariate location-scatter model of Example 3 is a special case of the multivariate linear regression model by taking $\mathbf{x} = 1$ and $\mathbf{B}^T = \boldsymbol{\mu}$, this also yields the expression for the influence function of the location MM-functional. Finally, there is an interesting connection with the CM-functionals considered in Kent and Tyler [14], whose

influence function depends on a parameter λ_0 . The expression for $\text{IF}(\mathbf{s}_0; \boldsymbol{\beta}_1, P)$ in Theorem 6 is similar to the one for the location CM-functional and they both coincide when $c_\sigma = 1/\sqrt{\lambda_0}$.

It remains to determine the influence function of the variance component MM-functional $\boldsymbol{\theta}_1$. This is done in the next corollary.

Corollary 4. *Suppose that the conditions hold of Theorem 6 and suppose that ρ_0 satisfies (R2) and (R4). Let $\boldsymbol{\zeta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ be the pair of initial functionals satisfying $(\boldsymbol{\beta}_0(P), \boldsymbol{\theta}_0(P)) = (\boldsymbol{\beta}^*, \boldsymbol{\theta}^*)$, and suppose that $\text{IF}(\mathbf{s}, \boldsymbol{\zeta}_0, P)$ exists. Let $\boldsymbol{\theta}_1(P)$ and $\boldsymbol{\theta}_1(P_{h, \mathbf{s}_0})$ be solutions of equation (3.11) and equation (3.11) with $P = P_{h, \mathbf{s}_0}$, respectively. Let $u_1(s) = \rho'_1(s)/s$ and $v_1(s) = \rho'_1(s)s$, and let $\mathbf{L}$ be defined in (7.6). Let $c_\sigma = |\boldsymbol{\Sigma}|^{1/(2k)}/\sigma(P)$ and suppose that $\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|] > 0$. Then, for $\mathbf{s}_0 \in \mathbb{R}^k \times \mathbb{R}^{kq}$,*

$$\begin{aligned} \text{IF}(\mathbf{s}_0, \boldsymbol{\theta}_1, P) = & \frac{ku_1(c_\sigma \|\mathbf{z}_0\|)}{\gamma_1} \left(\mathbf{L}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{L} \right)^{-1} \mathbf{L}^T \left(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2} \right) \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) \\ & - \frac{1}{c_\sigma^2} \left(\frac{v_1(c_\sigma \|\mathbf{z}_0\|)}{\gamma_1} - \frac{2(\rho_0(c_\sigma \|\mathbf{z}_0\|) - b_0)}{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|]} \right) \boldsymbol{\theta}^*, \end{aligned}$$

where $\mathbf{z}_0 = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_0 - \mathbf{X}_0 \boldsymbol{\beta}^*)$.

Note that for $\mathbf{V}$ with a linear structure, one has $\text{vec}(\mathbf{V}(\boldsymbol{\theta}_1(P))) = \mathbf{L} \boldsymbol{\theta}_1(P)$. Hence, the influence function for the covariance MM-functional follows immediately from Corollary 4:

$$\text{IF}(\mathbf{s}_0, \text{vec}(\mathbf{V}(\boldsymbol{\theta}_1)), P) = \mathbf{L} \text{IF}(\mathbf{s}_0, \boldsymbol{\theta}_1, P). \quad (8.6)$$

Since $\rho'_1(s)s$ and $\rho_0(s)$ are bounded, it follows that the influence functions of γ , $\boldsymbol{\theta}_1$, and $\mathbf{V}(\boldsymbol{\theta}_1)$ are bounded uniformly in both $\mathbf{y}_0$ and $\mathbf{X}_0$, whereas $\text{IF}(\mathbf{s}_0, \boldsymbol{\beta}_1, P)$ is bounded uniformly in $\mathbf{y}_0$, but not in $\mathbf{X}_0$.

The expressions for the influence function of the covariance MM-functionals in Corollary 4 and (8.6) are characterized by two real-valued functions,

$$\alpha_C(s) = \frac{k\rho'_1(s)}{\gamma_1 s}, \quad \beta_C(s) = \frac{1}{c_\sigma^2} \left(\frac{\rho'_1(s)s}{\gamma_1} - \frac{2(\rho_0(s) - b_0)}{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|]} \right). \quad (8.7)$$

When $c_\sigma = 1$, using that for $\mathbf{V}$ with a linear structure, it holds $\mathbf{L} \boldsymbol{\theta}^* = \text{vec}(\boldsymbol{\Sigma})$, this matches with the characterization of general structured covariance functionals obtained in Lopuhaä [20]. Such a characterization was already observed by Croux and Haesbroeck [6] for affine equivariant covariance functionals. The function α_C in (8.7) coincides with the one for the covariance S-functional defined with ρ_1 , see Corollary 8.4 in Lopuhaä *et al* [22]. The function β_C in (8.7) (with $c_\sigma = 1$) has the same structure as the one for covariance S-functionals, but the first term is built from ρ_1 , whereas the second term is built from ρ_0 . As expected, when $\rho_0 = \rho_1 = \rho$, the above characterization coincides with the influence function of the covariance S-functional defined with loss function ρ .

Furthermore, for covariance functionals $\mathbf{C}$, it holds that the influence function of a scale invariant mapping $H(\mathbf{C})$, i.e., $H(\lambda \mathbf{C}) = H(\mathbf{C})$, for $\lambda > 0$, only depends on the function α_C , see (8.3) in Kent and Tyler [14] for covariance CM-functionals or see Lemma 2 in Lopuhaä [20] for linearly structured covariance functionals. Because the characterizations of the influence functions of covariance MM- and S-functionals have the same function α_C , it follows that the influence functions of any scale invariant mapping of covariance MM- and S-functionals are the same. A typical example is the shape component $\boldsymbol{\Gamma}(\boldsymbol{\theta}_1)$ of the covariance MM-functional, where $\boldsymbol{\Gamma}$ is defined in (3.1). Lemma 2 in Lopuhaä [20], together with Corollary 4 and (8.6), yields that $\boldsymbol{\Gamma}(\boldsymbol{\theta}_1)$ has influence function

$$\begin{aligned} \text{IF}(\mathbf{s}_0; \boldsymbol{\Gamma}(\boldsymbol{\theta}_1), P) = & \frac{ku_1(c_\sigma \|\mathbf{z}_0\|)}{\sigma^2(P) \gamma_1} \left\{ \mathbf{L} \left(\mathbf{L}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{L} \right)^{-1} \mathbf{L}^T \left(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2} \right) \right. \\ & \left. \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) - \frac{\|\mathbf{z}_0\|^2}{k} \text{vec}(\boldsymbol{\Sigma}) \right\}, \end{aligned} \quad (8.8)$$

where $\mathbf{z}_0 = \Sigma^{-1/2}(\mathbf{y}_0 - \mathbf{X}_0\beta^*)$ and γ_1 is defined in (8.5). When $c_\sigma = 1$, then this is the same as the influence function of the shape S-functional defined with ρ_1 , see Example 6 in Lopuhaä [20]. For the location-scatter model of Example 3, this confirms the claim made in Salibián-Barrera *et al* [31] for the shape MM-functional.

Similarly, the influence function of a scale invariant mapping of variance component functionals only depends on the function α_C , see Lemma 2 in Lopuhaä [20]. Since the characterizations of the variance components MM- and S-functional share the same function α_C , see Corollary 4 and Corollary 8.4 in Lopuhaä *et al* [22], it follows that the influence functions of any scale invariant mapping of variance component MM- and S-functionals are the same. Examples are $\boldsymbol{\theta}/\|\boldsymbol{\theta}\|$ or $\boldsymbol{\theta}/|\mathbf{V}(\boldsymbol{\theta})|^{1/k}$, for linear covariance structures, which represent a direction component of $\boldsymbol{\theta}$.

Remark 8.1. From (3.11) and the fact that $|\mathbf{V}(\gamma)| = 1$, it follows that $\boldsymbol{\Gamma}(\boldsymbol{\theta}_1) = \mathbf{V}(\gamma)$. This means that $\mathbf{V}(\gamma)$ represents the shape component of $\mathbf{V}(\boldsymbol{\theta}_1)$, so that the expression in (8.8) also coincides with the influence function of $\mathbf{V}(\gamma)$. Similarly, the expression for $\text{IF}(\mathbf{s}_0; \gamma, P)$ in Theorem 6 coincides with the influence function of the direction component $\boldsymbol{\theta}_1/|\mathbf{V}(\boldsymbol{\theta}_1)|^{1/k}$, corresponding to the variance components MM-functional $\boldsymbol{\theta}_1$. The influence function of the direction component $\boldsymbol{\theta}_1/\|\boldsymbol{\theta}_1\|$ can be found in Example 7 in Lopuhaä [20].

The results in Corollary 4, and in (8.6) and (8.8) can be applied to derive influence functions for the covariance functionals in the multivariate statistical models of Examples 1, 2, and 3. Details are given in the supplemental material [21].

9 Asymptotic normality

To establish asymptotic normality of the MM-estimators, we use the score equations obtained in Proposition 1. We will use score equation (7.2), with P replaced by the empirical measure $\mathbb{P}_n$ corresponding to observations $\mathbf{s}_1, \dots, \mathbf{s}_n$, with $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i) \in \mathbb{R}^k \times \mathbb{R}^{kq}$. From Proposition 1, we see that any local minimum $\boldsymbol{\xi}_n = (\boldsymbol{\beta}_{1,n}, \gamma_n) \in \mathfrak{D}$ of $R_n(\boldsymbol{\beta}, \mathbf{V}(\gamma))$ with $|\mathbf{V}(\gamma)| = 1$, must satisfy

$$\int \Psi(\mathbf{s}, \boldsymbol{\xi}_n, \sigma_n) d\mathbb{P}_n(\mathbf{s}) = \mathbf{0}, \quad (9.1)$$

where $\Psi = (\Psi_\beta, \Psi_\gamma)$ is defined in (7.3) and σ_n is a solution of (3.2).

9.1 The general case

Since Ψ also depends on σ it can be expected that for general P , the limiting behavior of $\boldsymbol{\xi}_n$ will depend on that of σ_n , see Lemma 13 in the supplemental material [21]. In turn, since σ_n is a solution of (3.2), it can be expected that in general its limiting behavior depends on that of the initial estimators $\boldsymbol{\zeta}_{0,n} = (\boldsymbol{\beta}_{0,n}, \boldsymbol{\theta}_{0,n})$, see Lemma 14 in the supplemental material [21].

Similar to Section 8, the situation becomes somewhat simpler if the distribution P has a point of symmetry. The asymptotic expansion for $\sigma_n - \sigma(P)$ obtained in Lemma 14 in [21] does simplify, but details on the limiting distribution of $\boldsymbol{\xi}_n - \boldsymbol{\xi}(P)$ can still not be provided without further information on the limiting behavior of $\boldsymbol{\zeta}_{0,n} - \boldsymbol{\zeta}_0(P)$. The situation differs for $\boldsymbol{\beta}_{1,n} - \boldsymbol{\beta}_1(P)$, for which the limiting distribution is given by the following corollary.

Theorem 7. Suppose ρ_1 satisfies (R2) and (R4), such that $u_1(s)$ is of bounded variation. Suppose $\mathbf{V}$ satisfies (V1) and $\mathbb{E}\|\mathbf{s}\|^2 < \infty$. Let σ_n and $\sigma(P)$ be solutions of (3.2) and (3.8), respectively, and let $\boldsymbol{\xi}_n$ and $\boldsymbol{\xi}(P)$ be local minima of $R_n(\boldsymbol{\beta}, \mathbf{V}(\gamma))$ and $R_P(\boldsymbol{\beta}, \mathbf{V}(\gamma))$, respectively. Suppose $(\boldsymbol{\xi}_n, \sigma_n) \rightarrow (\boldsymbol{\xi}(P), \sigma(P))$, in probability, and that $\boldsymbol{\beta}_1(P)$ is a point of symmetry of P . Let Λ_β be defined by (8.3) with Ψ_β from (7.3) and suppose that Λ_β is continuously differentiable with a non-singular derivative $\mathbf{D}_\beta = \partial \Lambda_\beta / \partial \boldsymbol{\beta}$ at $(\boldsymbol{\xi}(P), \sigma(P))$. Then $\sqrt{n}(\boldsymbol{\beta}_{1,n} - \boldsymbol{\beta}_1(P))$ is asymptotically normal with mean zero and variance

$$\mathbf{D}_\beta^{-1} \mathbb{E} [\Psi_\beta(\mathbf{s}, \boldsymbol{\xi}(P), \sigma(P)) \Psi_\beta(\mathbf{s}, \boldsymbol{\xi}(P), \sigma(P))^T] \mathbf{D}_\beta^{-1}.$$

The limiting distribution of $\beta_{1,n}$ given in Theorem 7 is similar to that of the regression MM-estimator β_1 in Lopuhaä [19] defined with loss function ρ_1 and initial covariance estimator $\mathbf{V}_{0,n}$, and they both coincide when $\mathbf{V}_{0,n} \rightarrow \sigma^2(P)\mathbf{V}(\gamma(P))$, in probability.

If one is allowed to interchange integration and differentiation in $\mathbf{D}_\beta$, then for the multiple linear regression model, the limiting distribution of $\beta_{1,n}$ established in Theorem 7 coincides with that of the regression MM-estimator considered in Yohai [36]. For linear mixed effects models, multivariate linear regression models, or multivariate location-scatter models, the limiting distribution of MM-estimators is only available at distributions with an elliptically contoured density. This situation is discussed in the next section for model (2.1).

9.2 Elliptical contoured densities

Consider the special case that P satisfies (E). As before, when determining the limiting normal distribution of the MM-estimators, we cannot use affine equivariance and restrict ourselves to the case $(\mathbf{0}, \mathbf{I}_k)$. Instead, we use some of the results obtained in Section 8.2 to establish the limiting normal distributions of the MM-estimators $\xi_n = (\beta_{1,n}, \gamma_n)$, $\theta_{1,n}$, and $\mathbf{V}(\theta_{1,n})$. Let

$$\sigma_1 = \frac{k\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho_1'(c_\sigma \|\mathbf{z}\|)^2 (c_\sigma \|\mathbf{z}\|)^2]}{(k+2)\gamma_1^2}, \quad (9.2)$$

where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma(P)$ and γ_1 is defined in (8.5). The limiting distribution of MM-estimators $\beta_{1,n}$ and γ_n at distributions P , such that $\mathbf{y} \mid \mathbf{X}$ has an elliptical contoured density, are now given by the following theorem.

Theorem 8. *Suppose P satisfies (E) for some $(\beta^*, \theta^*) \in \mathbb{R}^q \times \Theta$ and $\mathbb{E}\|\mathbf{s}\|^2 < \infty$. Suppose ρ_1 satisfies (R1)-(R5), such that $u_1(s)$ is of bounded variation, and suppose $\mathbf{V}$ satisfies (V5) and has a linear structure (7.5). Let σ_n and $\sigma(P)$ be solutions of (3.2) and (3.8), respectively, and suppose that $\sigma_n - \sigma(P) = O_P(1/\sqrt{n})$. Let $\xi_n = (\beta_{1,n}, \gamma_n)$ and $\xi(P) = (\beta_1(P), \gamma(P))$ be local minima of $R_n(\beta, \mathbf{V}(\gamma))$ and $R_P(\beta, \mathbf{V}(\gamma))$, respectively, and suppose that $\xi_n \rightarrow \xi(P)$, in probability. Suppose that $\beta_1(P) = \beta^*$ and that $\mathbf{V}(\gamma(P)) = \Sigma/|\Sigma|^{1/k}$. Let α_1 and γ_1 be defined in (8.5) and suppose that $\alpha_1 \neq 0$ and $\gamma_1 > 0$. Suppose $\mathbf{X}$ has full rank with probability one and $\mathbf{L}$, as defined in (7.6), has full rank. Then $\sqrt{n}(\beta_{1,n} - \beta^*)$ and $\sqrt{n}(\gamma_n - \gamma(P))$ are asymptotically independent.*

Furthermore, $\sqrt{n}(\beta_{1,n} - \beta^)$ is asymptotically normal with mean zero and variance*

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

where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma(P)$, and $\sqrt{n}(\gamma_n - \gamma(P))$ is asymptotically normal with mean zero and variance

$$\frac{2\sigma_1}{|\Sigma|^{2/k}} \left\{ \left(\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} \right)^{-1} - \frac{1}{k} \theta^* (\theta^*)^T \right\},$$

where σ_1 is defined in (9.2).

When $c_\sigma = 1$, then similar to Theorem 7, we find that the limiting distribution of $\beta_{1,n}$ coincides with that of the regression MM-estimator considered in Lopuhaä [19], defined with loss function ρ_1 and an initial covariance estimator $\mathbf{V}_{0,n}$ that is consistent for Σ . Moreover, it also coincides with the limiting distribution of the regression S-estimator defined with loss function ρ_1 , see Corollary 9.2 in Lopuhaä *et al* [22]. This confirms the claim made by Salibián-Barrera *et al* [31] about the location MM-estimator in the model of Example 3.

For the linear mixed effects model, the limiting distribution of $\beta_{1,n}$ obtained in Theorem 8 extends Theorem 1 in Copt and Heritier [4], which is restricted to $\mathbf{X}_i = \mathbf{X}$. For the multivariate linear regression model, the limiting distribution of $\beta_{1,n}$ in Theorem 8 coincides with the one found for the regression MM-estimator in Kudraszow and Maronna [15]. This also applies to the location MM-estimator in the multivariate location-scatter model, since this model is a special case of the multivariate linear regression model. Furthermore, there is a connection with CM-estimators

considered in Kent and Tyler [14], whose limiting distribution depends on a parameter λ_0 . The limiting distribution of $\sqrt{n}(\beta_{1,n} - \beta^*)$ obtained in Theorem 8 is similar to that of the location CM-estimator, see (7.9) in Kent and Tyler [14], and they both coincide when $c_\sigma = 1/\sqrt{\lambda_0}$.

Let

$$\sigma_3 = \frac{4\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[(\rho_0(c_\sigma \|\mathbf{z}\|) - b_0)^2 \right]}{\left(\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|] \right)^2}, \quad (9.3)$$

where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma(P)$. It remains to determine the limiting distribution of the variance components MM-estimator $\theta_{1,n}$. This is given in the next corollary.

Corollary 5. *Suppose that the conditions hold of Theorem 8 and suppose that ρ_0 satisfies (R1), (R2) and (R4). Let $\zeta_{0,n} = (\beta_{0,n}, \theta_{0,n})$ be the pair of initial estimators and let $\zeta_0 = (\beta_0, \theta_0)$ be the corresponding functional. Suppose that $(\beta_0(P), \theta_0(P)) = (\beta^*, \theta^*)$ and that $\zeta_{0,n} - \zeta_0(P) = O_P(1/\sqrt{n})$. Let σ_n and $\sigma(P)$ be solutions of (3.2) and (3.8), respectively, and suppose that $\sigma_n \rightarrow \sigma(P)$, in probability. Let $\theta_{1,n}$ and $\theta_1(P)$ be solutions of (3.6) and (3.11), respectively, and suppose that $\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|] > 0$, where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma(P)$. Then $\sqrt{n}(\theta_{1,n} - \theta_1(P))$ is asymptotically normal with mean zero and variance*

$$\frac{2\sigma_1}{c_\sigma^2} \left(\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} \right)^{-1} + \left(-\frac{2\sigma_1}{kc_\sigma^2} + \sigma_3 \right) \theta^* (\theta^*)^T,$$

where σ_1 and σ_3 are defined in (9.2) and (9.3).

For linearly structured $\mathbf{V}$, one has $\text{vec}(\mathbf{V}(\theta_1(P))) = \mathbf{L}\theta_1(P)$ and $\text{vec}(\Sigma) = \mathbf{L}\theta^*$. Hence, application of the delta-method yields that $\sqrt{n}(\text{vec}(\mathbf{V}(\theta_{1,n})) - \text{vec}(\Sigma))$ is asymptotically normal with mean zero and variance

$$\frac{2\sigma_1}{c_\sigma^2} \mathbf{L} \left(\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} \right)^{-1} \mathbf{L}^T + \left(-\frac{2\sigma_1}{kc_\sigma^2} + \sigma_3 \right) \text{vec}(\Sigma) \text{vec}(\Sigma)^T. \quad (9.4)$$

The expressions for the limiting variances of the covariance MM-estimators in Corollary 5 and (9.4) are characterized by two scalars σ_1/c_σ^2 and $\sigma_2 = -2\sigma_1/(kc_\sigma^2) + \sigma_3$. When $c_\sigma = 1$, this matches with the characterization of general structured covariance estimators, see Corollary 2 in Lopuhaä [20]. Such a characterization was already observed by Tyler [33] for affine equivariant covariance estimators in the multivariate location-scatter model. The constant σ_1 (with $c_\sigma = 1$) coincides with the one for the covariance S-estimator defined with loss function ρ_1 . The constant σ_2 (with $c_\sigma = 1$) has the same structure as the one for covariance S-estimators, but the first term $-2\sigma_1/k$ is built from ρ_1 , whereas the second term σ_3 is built from ρ_0 . As expected, when $\rho_0 = \rho_1 = \rho$, the above characterization coincides with the one for the covariance S-estimator defined with loss function ρ .

Note that the limiting variance of scale invariant mappings $H(\mathbf{C}_n)$ of a covariance estimator $\mathbf{C}_n$, only depends on the scalar σ_1 , see (8.2) in Kent and Tyler [14] for affine equivariant covariance estimators or Theorem 3 in Lopuhaä [20] for estimators of a linearly structured covariance. Because the characterizations of the limiting variances of covariance MM- and S-estimators have the same scalar σ_1 , it follows that the limiting distributions of any scale invariant mapping of covariance MM- and S-estimators are the same. A typical example is the shape component $\Gamma(\theta_{1,n})$ of the covariance MM-estimator, where Γ is defined in (3.1). Theorem 2 and Example 4 in Lopuhaä [20], together with (9.4), yield that $\sqrt{n}(\text{vec}(\Gamma(\theta_{1,n})) - \text{vec}(\Gamma(\theta_1(P))))$ is asymptotically normal with mean zero and variance

$$\frac{2\sigma_1}{c_\sigma^2 |\Sigma|^{2/k}} \left\{ \mathbf{L} \left(\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} \right)^{-1} \mathbf{L}^T - \frac{1}{k} \text{vec}(\Sigma) \text{vec}(\Sigma)^T \right\}. \quad (9.5)$$

When $c_\sigma = 1$, this coincides with the limiting distribution of the shape S-estimator defined with ρ_1 , see Examples 3 and 4 in Lopuhaä [20]. For the location-scatter model in Example 3, this confirms the claim made in Salibián-Barrera *et al* [31] for the shape MM-estimator.

Similarly, the limiting distribution of a scale invariant mapping of variance component estimators only depends on the scalar σ_1 , see Theorem 2 in Lopuhaä [20]. Since the characterizations

of the limiting distribution of the variance components MM- and S-estimators share the same scalar σ_1 , see Corollary 5 and Corollary 9.2 in Lopuhaä *et al* [22], it follows that the limiting distributions of any scale invariant mapping of variance component MM- and S-estimators are the same. Examples are direction components, such as $\boldsymbol{\theta}/\|\boldsymbol{\theta}\|$ or $\boldsymbol{\theta}/|\mathbf{V}(\boldsymbol{\theta})|^{1/k}$, for linear covariance structures.

Remark 9.1. From (3.6) and the fact that $|\mathbf{V}(\gamma_n)| = 1$, it follows that $\boldsymbol{\Gamma}(\boldsymbol{\theta}_{1,n}) = \mathbf{V}(\gamma_n)$. This means that $\mathbf{V}(\gamma_n)$ represents the shape component of $\mathbf{V}(\boldsymbol{\theta}_{1,n})$, so that the limiting distribution of the shape component of $\mathbf{V}(\boldsymbol{\theta}_{1,n})$ is the same as that of $\mathbf{V}(\gamma_n)$. Similarly, the limiting distribution of $\sqrt{n}(\gamma_n - \gamma(P))$ established in Theorem 8, coincides with that of the direction component $\boldsymbol{\theta}/|\mathbf{V}(\boldsymbol{\theta})|^{1/k}$, corresponding to the variance components MM-estimator $\boldsymbol{\theta}_{1,n}$. The limiting distribution of the direction component $\boldsymbol{\theta}_{1,n}/\|\boldsymbol{\theta}_{1,n}\|$ can be found in Example 5 in Lopuhaä [20].

The results in Corollary 5, and in (9.4) and (9.5) can be applied to derive the limiting distributions for the covariance estimators in the multivariate statistical models of Examples 1, 2, and 3. Details are given in the supplemental material [21].

10 Application

We apply our results to MM-estimators and MM-functionals to linear model (2.1). Consider a distribution P that satisfies (E), where $\mathbf{V}$ has linear structure (7.5). The loss functions ρ_0 and ρ_1 are constructed from Tukey's biweight, as defined in (3.7), by taking $\rho_j(s) = \rho_B(s; c_j)$, for $j = 0, 1$, such that $0 < c_0 \leq c_1 < \infty$. As initial estimators we use the S-estimators $(\boldsymbol{\beta}_{0,n}, \boldsymbol{\theta}_{0,n})$, defined by minimizing $|\mathbf{V}(\boldsymbol{\theta})|$, subject to

$$\frac{1}{n} \sum_{i=1}^n \rho_0(d(\mathbf{y}_i, \mathbf{X}_i \boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\theta}))) = b_0,$$

where d is defined in (7.1) and $b_0 = \mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho_0(\|\mathbf{z}\|)]$, and where the minimum is taken over all $\boldsymbol{\beta} \in \mathbb{R}^q$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^l$, such that $\mathbf{V}(\boldsymbol{\theta}) \in \text{PDS}(k)$. The cut-off c_0 is chosen such that $b_0/(c_0^2/6) = 0.5$, so that the initial S-estimator has (asymptotic) breakdown point 0.50. This means that according to Theorem 8, the scalar $\lambda = \mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_1(\|\mathbf{z}\|)^2]/(k\alpha_1^2)$, where α_1 is defined in (8.5) with $c_\sigma = 1$, represents the asymptotic efficiency of the regression MM-estimator relative to the least squares estimator (for which $\lambda = 1$). Similarly, according to (part two of) Theorem 8 and (9.5), the scalar $\sigma_1 = k\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_1(\|\mathbf{z}\|)^2(\|\mathbf{z}\|)^2]/((k+2)\gamma_1^2)$, where γ_1 is defined in (8.5) with $c_\sigma = 1$, represents the asymptotic relative efficiency of both the MM-estimator of shape as well as the MM-estimator for the direction of the variance components, relative to the least squares estimators of shape and direction, respectively (for which $\sigma_1 = 1$). Hence, the scalars λ and σ_1 only depend on the function ρ_1 . By keeping c_0 fixed, the breakdown point of the MM-estimators remains unaffected, and by varying $c_1 \geq c_0$ we will investigate how the scalars λ and σ_1 for the asymptotic relative efficiency will vary.

We further investigate how at the same time the gross error sensitivity (GES) of the corresponding MM-functionals will vary. For simplicity we only consider perturbations in $\mathbf{y}$ and leave $\mathbf{X}$ unchanged. According to Theorem 6, it can be seen that for any norm, $\|\text{IF}(\mathbf{s}; \boldsymbol{\beta}_1, P)\|$ is proportional to $\alpha_1^{-1} |\rho'_1(d(\mathbf{y}))|$, where α_1 is defined in (8.5) with $c_\sigma = 1$ and $d(\mathbf{y})^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)$. Therefore, we propose the scalar

$$G_1 = \frac{1}{\alpha_1} \sup_{s>0} |\rho'_1(s)|,$$

as an index for the GES of regression MM-functionals. This coincides with the GES index for location CM-functionals in Kent and Tyler [14]. Similarly, from (part two of) Theorem 6 and (8.8) it follows that $\|\text{IF}(\mathbf{y}; \boldsymbol{\gamma}, P)\|$ and $\|\text{IF}(\mathbf{y}; \mathbf{V}(\boldsymbol{\gamma}), P)\|$ are proportional to $\gamma_1^{-1} |\rho'_1(d(\mathbf{y}))d(\mathbf{y})|$, where γ_1 is defined in (8.5) with $c_\sigma = 1$. We propose scalar

$$G_2 = \frac{k}{(k+2)\gamma_1} \sup_{s>0} |\rho'_1(s)s|,$$

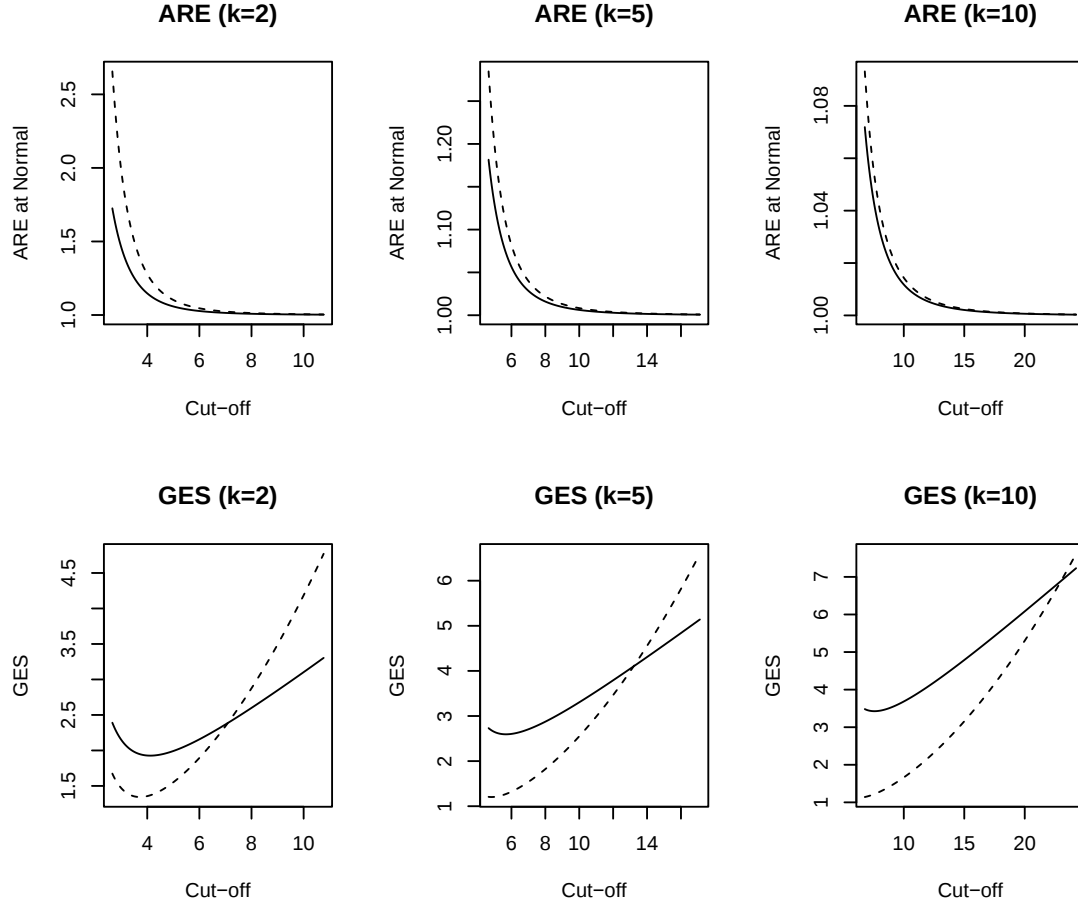


Figure 1: Asymptotic efficiencies at the multivariate normal distribution (top row) for the MM-estimators of regression (solid) and for shape and direction (dashed) and the GES (bottom row) of the corresponding MM-functionals for dimensions $k = 2, 5, 10$.

as an index for the GES of shape and direction functionals. In this way, G_2 coincides with the GES index for CM-functionals of shape in Kent and Tyler [14].

We investigate how the asymptotic efficiency at the multivariate normal of MM-estimators, and the GES of the corresponding MM-functionals behave as we vary the cut-off constant $c_1 \geq c_0$. In Figure 1, on the top row we plotted the indices λ (solid lines) and σ_1 (dashed lines) together as a function of $c_1 \geq c_0$ in dimensions $k = 2, 5$ and 10 . In dimension $k = 2$, the asymptotic efficiencies $\lambda = 1.725$ and $\sigma_1 = 2.656$ of the MM-estimators at cut-off $c_1 = c_0 = 2.661$ are the same as that of the initial 50% breakdown S-estimators. When increasing the cut-off c_1 , one can both gain efficiency and lower the GES. For example, the GES index for the shape and direction MM-functional attains its minimal value $G_2 = 1.344$ at $c_1 = 3.724$. For this cut-off value the GES index for the regression MM-functional is $G_1 = 1.947$ and the asymptotic efficiencies are $\lambda = 1.197$ and $\sigma_1 = 1.383$. Similarly, the GES index for the regression MM-functional attains its minimal value $G_1 = 1.927$ at $c_1 = 4.113$. This would yield $G_2 = 1.368$, $\lambda = 1.131$ and $\sigma_1 = 1.246$.

In dimension $k = 5$, the asymptotic efficiencies at cut-off $c_1 = c_0 = 4.652$ are $\lambda = 1.182$ and $\sigma_1 = 1.285$. The GES index for the regression MM-functional attains its minimal value $G_1 = 2.595$ at $c_1 = 5.675$. For this cut-off value the GES index $G_2 = 1.270$ and the asymptotic efficiencies are $\lambda = 1.073$ and $\sigma_1 = 1.107$. The index for the shape and direction MM-functional attains its minimal value $G_2 = 1.204$ at $c_1 = c_0 = 4.652$.

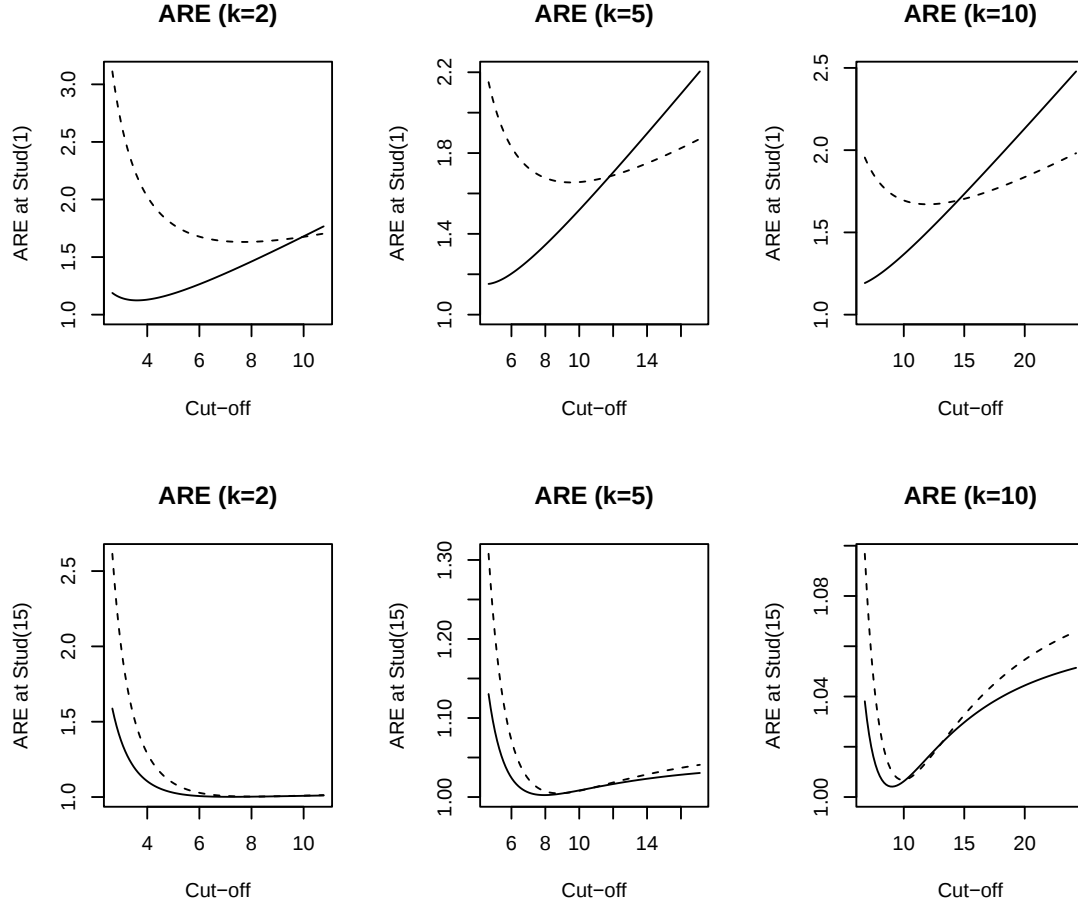


Figure 2: Asymptotic efficiencies at the multivariate Student distribution with degrees of freedom $\nu = 1$ (top row) and $\nu = 15$ (bottom row) for the MM-estimators of regression (solid) and for shape and direction (dashed) for dimensions $k = 2, 5, 10$.

In dimension $k = 10$, the asymptotic efficiencies at cut-off $c_1 = c_0 = 6.776$ are $\lambda = 1.072$ and $\sigma_1 = 1.093$. The GES index for the regression MM-functional attains its minimal value $G_1 = 3.426$ at $c_1 = 7.580$. For this cut-off value the GES index $G_2 = 1.270$ and the asymptotic efficiencies are $\lambda = 1.042$ and $\sigma_1 = 1.053$. The GES index for the shape and direction MM-functional attains its minimal value $G_2 = 1.142$ at $c_1 = c_0 = 6.776$.

In the top row of Figure 1 it can be seen that the asymptotic efficiencies become closer to one when the dimension is large. This is a well known phenomenon observed when the efficiency is computed under a multivariate normal setting. As a comparison, we have investigated whether this behavior is observed in a neighborhood of the multivariate normal. We have computed asymptotic efficiencies relative to the maximum likelihood estimator at the k -variate Student distribution with degrees of freedom $\nu = 1$ and $\nu = 15$. The scalars λ and σ_1 for the ML estimator at the Student(ν) distribution are given by

$$\lambda^{\text{ML}} = \frac{k\mathbb{E}_\nu[w_1(\|\mathbf{z}\|)^2\|\mathbf{z}\|^2]}{\left(k\mathbb{E}_\nu[w_1(\|\mathbf{z}\|)] + \mathbb{E}_\nu[w'_1(\|\mathbf{z}\|)\|\mathbf{z}\|]\right)^2}; \quad \sigma_1^{\text{ML}} = \frac{k(k+2)\mathbb{E}_\nu[w_1(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4]}{\left(\mathbb{E}_\nu[w'_1(\|\mathbf{z}\|)\|\mathbf{z}\|^3] + k(k+2)\right)^2},$$

where $w_1(s) = (\nu + k)/(k + s^2)$.

The asymptotic efficiencies relative to the ML estimator at the k -variate Student(ν) distribution

with $\nu \in \{1, 15\}$ are visible in Figure 2. The graphs in the top row correspond to $\nu = 1$ and are quite different from the ones in the top row in Figure 1. Moreover, the behavior of the MM-regression estimator (solid lines) differs from that of the MM-estimators of shape and direction (dashed lines). The best efficiencies for the regression MM-estimator $\lambda = 1.124, 1.152, 1.192$, for $k = 2, 5, 10$, are obtained for values of c_1 very close or equal to c_0 , whereas the efficiency for the MM-estimators of shape and direction can be improved for larger values of c_1 and are equal to $\sigma_1 = 1.631, 1.655, 1.671$ at $c_1 = 7.656, 9.561, 11.893$, for $k = 2, 5, 10$. As expected both MM-estimators with large values for c_1 have poor efficiencies, because they tend to behave similar to the least squares estimators.

For the Student distribution with $\nu = 15$ degrees of freedom, the behavior of the efficiency is more or less in between the ones at the multivariate normal and the Student distribution with $\nu = 1$ degrees of freedom. The graphs in the bottom row of Figure 2 are more similar to the ones in the top row of Figure 1, although in higher dimensions the efficiencies get worse. The best efficiencies for the regression MM-estimator $\lambda = 1.001, 1.002, 1.004$, for $k = 2, 5, 10$, are obtained at $c_1 = 7.246, 7.925, 9.070$, and the best efficiencies for the MM-estimators of shape and direction are equal to $\sigma_1 = 1.003, 1.005, 1.008$, for $k = 2, 5, 10$, obtained at $c_1 = 8.065, 8.806, 9.952$.

References

- [1] C. Agostinelli and V. J. Yohai. Composite robust estimators for linear mixed models. *Journal of the American Statistical Association*, 111(516):1764–1774, 2016.
- [2] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.
- [3] I. Chervoneva and M. Vishnyakov. Generalized S-estimators for linear mixed effects models. *Statistica Sinica*, 24(3):1257–1276, 2014.
- [4] S. Copt and S. Heritier. Robust alternatives to the F-test in mixed linear models based on mm-estimates. *Biometrics*, 63(4):1045–1052, 2007.
- [5] S. Copt and M. P. Victoria-Feser. High-breakdown inference for mixed linear models. *Journal of the American Statistical Association*, 101(473):292–300, 2006.
- [6] C. Croux and G. Haesbroeck. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, 87(3):603–618, 2000.
- [7] P. L. Davies. Asymptotic behaviour of S -estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.*, 15(3):1269–1292, 1987.
- [8] E. Demidenko. *Mixed models*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2013. Theory and applications with R.
- [9] D. Donoho and P. J. Huber. The notion of breakdown point. In *A Festschrift for Erich L. Lehmann*, Wadsworth Statist./Probab. Ser., pages 157–184. Wadsworth, Belmont, CA, 1983.
- [10] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware. *Applied longitudinal analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2011.
- [11] F. R. Hampel. The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.*, 69:383–393, 1974.
- [12] H. O. Hartley and J. N. K. Rao. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54:93–108, 1967.
- [13] S. Heritier, E. Cantoni, S. Copt, and M.-P. Victoria-Feser. *Robust methods in biostatistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2009.

- [14] J. T. Kent and D. E. Tyler. Constrained M -estimation for multivariate location and scatter. *Ann. Statist.*, 24(3):1346–1370, 1996.
- [15] N. L. Kudraszow and R. A. Maronna. Estimates of MM type for the multivariate linear model. *J. Multivariate Anal.*, 102(9):1280–1292, 2011.
- [16] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.
- [17] H. P. Lopuhaä. On the relation between S -estimators and M -estimators of multivariate location and covariance. *Ann. Statist.*, 17(4):1662–1683, 1989.
- [18] H. P. Lopuhaä. Highly efficient estimators of multivariate location with high breakdown point. *Ann. Statist.*, pages 398–413, 1992.
- [19] H. P. Lopuhaä. Highly efficient estimators with high breakdown point for linear models with structured covariance matrices. *Econometrics and Statistics*, 2023.
- [20] H. P. Lopuhaä. Asymptotics of estimators for structured covariance matrices. *J. Multivariate Anal.*, 208:Paper No. 105443, 20, 2025.
- [21] H. P. Lopuhaä. Supplement to “multivariate MM-estimators with auxiliary scale for linear models with structured covariance matrices”. 2025.
- [22] H. P. Lopuhaä, V. Garés, and A. Ruiz-Gazen. S -estimation in linear models with structured covariance matrices. *Ann. Statist.*, 51(6):2415–2439, 2023.
- [23] H. P. Lopuhaä, V. Gares, and A. Ruiz-Gazen. Supplement to “ S -estimation in linear models with structured covariance matrices”. 2023.
- [24] J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1988.
- [25] R. D. Martin, V. J. Yohai, and R. H. Zamar. Min-max bias robust regression. *Ann. Statist.*, 17(4):1608–1630, 1989.
- [26] J. C. Pinheiro, C. Liu, and Y. N. Wu. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *J. Comput. Graph. Statist.*, 10(2):249–276, 2001.
- [27] R. Ranga Rao. Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.*, 33:659–680, 1962.
- [28] C. R. Rao. Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.*, 67:112–115, 1972.
- [29] P. Rousseeuw. Multivariate estimation with high breakdown point. In *Mathematical statistics and applications, Vol. B (Bad Tatzmannsdorf, 1983)*, pages 283–297. Reidel, Dordrecht, 1985.
- [30] P. Rousseeuw and V. Yohai. Robust regression by means of S -estimators. In *Robust and nonlinear time series analysis (Heidelberg, 1983)*, volume 26 of *Lect. Notes Stat.*, pages 256–272. Springer, New York, 1984.
- [31] M. Salibián-Barrera, S. Van Aelst, and G. Willems. Principal components analysis based on multivariate MM estimators with fast and robust bootstrap. *J. Amer. Statist. Assoc.*, 101(475):1198–1211, 2006.
- [32] K. S. Tatsuoaka and D. E. Tyler. On the uniqueness of S -functionals and M -functionals under nonelliptical distributions. *Ann. Statist.*, 28(4):1219–1243, 2000.

- [33] D. E. Tyler. Radial estimates and the test for sphericity. *Biometrika*, 69(2):429–436, 1982.
- [34] D. E. Tyler. High breakdown point multivariate M -estimation. volume 54, pages 213–247 (2003). 2002. Special issue on robust statistics.
- [35] S. Van Aelst and G. Willems. Multivariate regression S -estimators for robust estimation and inference. *Statist. Sinica*, 15(4):981–1001, 2005.
- [36] V. J. Yohai. High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.*, 15(2):642–656, 1987.

11 Supplemental Material

11.1 Proofs for Section 4

For any $k \times k$ matrix $\mathbf{A}$, let $\lambda_k(\mathbf{A}) \leq \dots \leq \lambda_1(\mathbf{A})$ denote the eigenvalues of $\mathbf{A}$.

Proof of Lemma 1

Proof. From (R1)-(R2) we have that ρ_0 is bounded and continuous at zero. Hence, by dominated convergence, it follows that

$$\lim_{\sigma \rightarrow \infty} \int \rho_0 \left(\frac{\sqrt{(\mathbf{y} - \mathbf{X}\beta_0(P))^T \Gamma(\theta_0(P))^{-1} (\mathbf{y} - \mathbf{X}\beta_0(P))}}{\sigma} \right) dP(\mathbf{s}) = 0.$$

Similarly, together with (C0) and (R3), we find that

$$\begin{aligned} \lim_{\sigma \downarrow 0} \int \rho_0 \left(\frac{\sqrt{(\mathbf{y} - \mathbf{X}\beta_0(P))^T \Gamma(\theta_0(P))^{-1} (\mathbf{y} - \mathbf{X}\beta_0(P))}}{\sigma} \right) dP(\mathbf{s}) \\ = (\sup \rho_0)(1 - P(E_0)) > b_0. \end{aligned}$$

Since ρ_0 is continuous and $0 < b_0 < \sup \rho_0$, we conclude that there exists a solution $\sigma(P) > 0$ to (3.8). Because ρ_0 is strictly increasing on $[0, c_0]$, together with (C0), it follows that $\sigma(P)$ is unique. $\square$

Proof of Theorem 1

Proof. For $(\beta, \gamma) \in \mathbb{R}^k \times \mathbb{R}^l$, define cylinder

$$\mathcal{C}(\beta, \gamma, c) = \{(\mathbf{y}, \mathbf{X}) \in \mathbb{R}^k \times \mathbb{R}^{kq} : (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}(\gamma)^{-1} (\mathbf{y} - \mathbf{X}\beta) \leq c^2\}. \quad (11.1)$$

According to (V2), there exists $\gamma_0 \in \Theta$, such that

$$\mathbf{V}(\gamma_0) = \frac{\mathbf{V}(\theta_0(P))}{|\mathbf{V}(\theta_0(P))|^{1/k}} = \Gamma(\theta_0(P)),$$

and clearly $|\mathbf{V}(\gamma_0)| = 1$. If $(\beta, \gamma) \in \mathfrak{D}$ minimizes $R_P(\beta, \mathbf{V}(\gamma))$, then together with (3.3) and (3.8), it must satisfy

$$\begin{aligned} & P(\mathcal{C}(\beta, \gamma, c_1 \sigma(P))) \\ & \geq 1 - \frac{1}{\sup \rho_1} \int \rho_1 \left(\frac{\sqrt{(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}(\gamma)^{-1} (\mathbf{y} - \mathbf{X}\beta)}}{\sigma(P)} \right) dP(\mathbf{s}) \\ & \geq 1 - \frac{1}{\sup \rho_1} \int \rho_1 \left(\frac{\sqrt{(\mathbf{y} - \mathbf{X}\beta_0(P))^T \mathbf{V}(\gamma_0)^{-1} (\mathbf{y} - \mathbf{X}\beta_0(P))}}{\sigma(P)} \right) dP(\mathbf{s}) \\ & \geq 1 - \frac{1}{\sup \rho_0} \int \rho_0 \left(\frac{\sqrt{(\mathbf{y} - \mathbf{X}\beta_0(P))^T \Gamma_0(P)^{-1} (\mathbf{y} - \mathbf{X}\beta_0(P))}}{\sigma(P)} \right) dP(\mathbf{s}) \\ & = 1 - \frac{b_0}{\sup \rho_0} = 1 - r_0 \geq \epsilon. \end{aligned} \quad (11.2)$$

Since P satisfies $(C2_\epsilon)$ for $1 - r_0 \geq \epsilon$, from Lemma 4.1(i) in Lopuhaä *et al* [22], it follows that there exist $a_1 > 0$, only depending on c_1 and $(C2_\epsilon)$, such that $\lambda_k(\sigma^2(P)\mathbf{V}(\gamma)) \geq a_1$, so that $\lambda_k(\mathbf{V}(\gamma)) \geq a_1/\sigma^2(P) > 0$. Since $|\mathbf{V}(\gamma)| = 1$, it immediately follows that

$$\lambda_1(\mathbf{V}(\gamma)) \leq a_2 := \left(\frac{\sigma^2(P)}{a_1} \right)^{k-1} < \infty.$$

Let $K \subset \mathbb{R}^k \times \mathcal{X}$ be a compact set, such that $P(K) \geq r_0 + \epsilon$, which exists according to condition $(C1_\epsilon)$. From Lemma 4.1(iii) in Lopuhaä *et al* [22], it follows that $\|\beta\| \leq M < \infty$, for some $M > 0$ that only depends on c_1 , a_2 , $\sigma(P)$, K and $(C2_\epsilon)$. We conclude that β is in a compact subset of $\mathbb{R}^q$ and $\mathbf{V}(\gamma)$ is in a compact set $B \subset \mathbb{R}^{k \times k}$. By identifiability, the mapping $\gamma \mapsto \mathbf{V}(\gamma)$ is one-to-one, so we can restrict γ to the pre-image $\mathbf{V}^{-1}(B)$. Then with conditions (V1) and (V3), it follows that also $\mathbf{V}^{-1}(B)$ is a compact set in Θ . We conclude that for minimizing $R_P(\beta, \mathbf{V}(\gamma))$ we can restrict ourselves to a compact set $B' \subset \mathfrak{D}$.

Because ρ_1 is continuous, together with condition (V1) and dominated convergence, it follows that $R_P(\beta, \mathbf{V}(\gamma))$ is a continuous function of (β, γ) , so that it must attain a minimum on B' . Hence, there exists a pair $(\beta_1(P), \gamma(P)) \in \mathfrak{D}$ that minimizes $R_P(\beta, \mathbf{V}(\gamma))$. Finally, condition (V2) immediately yields that there exists a $\theta_1(P)$ that solves (3.11) and by identifiability it follows that $\theta_1(P)$ is unique. $\square$

Proof of Corollary 1

Proof. Let $\mathbb{P}_n$ be the empirical measure corresponding to the collection $\mathcal{S}_n$. Then $\mathbb{P}_n$ satisfies $(C1_\epsilon)$ for any $0 < \epsilon \leq 1 - r_0$ and satisfies $(C2_\epsilon)$, for $\epsilon = (\kappa(\mathcal{S}_n) + 1)/n$. Clearly, $0 < (\kappa(\mathcal{S}_n) + 1)/n \leq 1 - r_0$, where $r_0 = b_0/\sup \rho_0$. Furthermore, since $(\beta_{0,n}, \theta_{0,n})$ satisfies (3.12), it follows that $\sigma_n = \sigma(\mathbb{P}_n)$ is a solution of (3.8), with $P = \mathbb{P}_n$. Hence, according to Theorem 1 there exists a pair $(\beta_1(\mathbb{P}_n), \gamma(\mathbb{P}_n))$ that minimizes $R_P(\beta, \mathbf{V}(\gamma))$, with $P = \mathbb{P}_n$, and a vector $\theta_1(\mathbb{P}_n)$ that is the unique solution of (3.11), with $P = \mathbb{P}_n$. But this is equivalent with saying that there exists a pair $(\beta_{1,n}, \gamma_n) \in \mathfrak{D}$ that minimizes $R_n(\beta, \mathbf{V}(\gamma))$ and a vector $\theta_{1,n}$ that is the unique solution of (3.6). $\square$

Proof of Corollary 2

Proof. Because P satisfies condition (C3), according to Ranga Rao [27, Theorem 4.2] we have

$$\sup_{C \in \mathfrak{C}} |P_t(C) - P(C)| \rightarrow 0, \quad \text{as } t \rightarrow \infty. \quad (11.3)$$

Consider the set $E_{t,0} = \{(\mathbf{y}, \mathbf{X}) \in \mathbb{R}^k \times \mathbb{R}^{kq} : \|\mathbf{y} - \mathbf{X}\beta_0(P_t)\| = 0\}$. Then $E_{t,0} \in \mathfrak{C}$, so that $P_t(E_{t,0}) - P(E_{t,0}) \rightarrow 0$, as $t \rightarrow \infty$. Since $\beta_0(P_t) \rightarrow \beta_0(P)$, as $t \rightarrow \infty$, it follows that $P(E_{t,0}) \rightarrow P(E_0)$, which implies that

$$P_t(E_{t,0}) = P_t(E_{t,0}) - P(E_{t,0}) + P(E_{t,0}) \rightarrow P(E_0) < 1 - \frac{b_0}{\sup \rho_0}.$$

Therefore, P_t satisfies (C0), for t sufficiently large. According to Lemma 1, a solution $\sigma(P_t)$ of (3.8) with $P = P_t$ exists and is unique. This proves part(i).

The argument that minimizing $R_{P_t}(\beta, \mathbf{V}(\gamma))$ over $(\beta, \gamma) \in \mathfrak{D}$ has at least one solution, is similar to the proof of Corollary 4.4 in Lopuhaä *et al* [22]. First note there exists $0 < \eta < \epsilon' - \epsilon$. Because strips $H(\alpha, \ell, \delta) \in \mathfrak{C}$, property (11.3) implies that every strip with $P_t(H(\alpha, \ell, \delta)) \geq \epsilon + \eta$ must also satisfy $P(H(\alpha, \ell, \delta)) \geq \epsilon$, for t sufficiently large. Together with the fact that P satisfies $(C2_\epsilon)$, this means that, for t sufficiently large,

$$\inf \{\delta : P_t(H(\alpha, \ell, \delta)) \geq \epsilon + \eta\} \geq \inf \{\delta : P(H(\alpha, \ell, \delta)) \geq \epsilon\} > 0.$$

It follows that, for t sufficiently large, P_t satisfies condition $(C2_{\epsilon+\eta})$. Next, consider the compact set K from $(C1_{\epsilon'})$. Without loss of generality we may assume that it belongs to $\mathfrak{C}$. Therefore, as

$P(K) \geq r_0 + \epsilon'$, for t sufficiently large $P_t(K) \geq r_0 + \epsilon + \eta$. It follows that, for t sufficiently large, P_t satisfies condition $(C1_{\epsilon+\eta})$. Since $\epsilon + \eta < 1 - r_0$, according to Theorem 1, for t sufficiently large, there exists a pair $(\beta_1(P_t), \gamma(P_t)) \in \mathfrak{D}$ that minimizes $R_{P_t}(\beta, \mathbf{V}(\gamma))$ and a vector $\theta_1(P_t)$ that is the unique solution of (3.11) with $P = P_t$. This proves part(ii). $\square$

11.2 Proofs for Section 5

Proof of Theorem 2

Proof. Let $\beta_{0,t} = \beta_0(P_t)$ and $\beta_{0,P} = \beta_0(P)$. Let Γ be the functional defined in (3.1), and define $\Gamma_{0,t} = \Gamma(\theta_0(P_t))$ and $\Gamma_{0,P} = \Gamma(\theta_0(P))$. Since ρ_0 satisfies (R2)-(R3) and $\mathbf{V}$ satisfies (V1), we can apply Lemma 3.2 from Lopuhaä [17]. As $(\beta_{0,t}, \Gamma_{0,t}) \rightarrow (\beta_{0,P}, \Gamma_{0,P})$, it follows that for s fixed,

$$\int \rho_0 \left(\frac{d(\mathbf{y}, \mathbf{X}\beta_{0,t}, \Gamma_{0,t})}{s} \right) dP_t(\mathbf{s}) \rightarrow \int \rho_0 \left(\frac{d(\mathbf{y}, \mathbf{X}\beta_{0,P}, \Gamma_{0,P})}{s} \right) dP(\mathbf{s}), \quad (11.4)$$

where d is defined in (7.1). Let $\sigma(P)$ be the unique solution of (3.8). Let $\delta > 0$ and suppose that $\liminf_{t \rightarrow \infty} \sigma(P_t) > \sigma(P) + \delta$. Since ρ_0 is strictly increasing on $[0, c_0]$, together with (11.4), it follows that

$$\begin{aligned} \int \rho_0 \left(\frac{d(\mathbf{y}, \mathbf{X}\beta_{0,t}, \Gamma_{0,t})}{\sigma(P_t)} \right) dP_t(\mathbf{s}) &\leq \int \rho_0 \left(\frac{d(\mathbf{y}, \mathbf{X}\beta_{0,t}, \Gamma_{0,t})}{\sigma(P) + \delta} \right) dP_t(\mathbf{s}) \\ &\rightarrow \int \rho_0 \left(\frac{d(\mathbf{y}, \mathbf{X}\beta_{0,P}, \Gamma_{0,P})}{\sigma(P) + \delta} \right) dP(\mathbf{s}) \\ &< \int \rho_0 \left(\frac{d(\mathbf{y}, \mathbf{X}\beta_{0,P}, \Gamma_{0,P})}{\sigma(P)} \right) dP(\mathbf{s}) = b_0, \end{aligned}$$

which is in contradiction with the definition of $\sigma(P_t)$. The argument is similar for $\sigma(P_t) < \sigma(P) - \delta$. We conclude that $|\sigma(P_t) - \sigma(P)| < \delta$, for t sufficiently large. Since $\delta > 0$ is arbitrary, this means that $\sigma(P_t) \rightarrow \sigma(P)$. This proves part (i).

To prove part (ii), first note that there exists $0 < \eta < \epsilon' - \epsilon$. Because $(\beta_{1,t}, \gamma_t) = (\beta_1(P_t), \gamma(P_t))$ is a local minimum of $R_{P_t}(\beta, \mathbf{V}(\gamma))$ that satisfies (3.10), we have

$$R_{P_t}(\beta_{1,t}, \mathbf{V}(\gamma_t)) \leq R_{P_t}(\beta_{0,t}, \Gamma_{0,t}). \quad (11.5)$$

Then, together with (3.3), similar to (11.2) we find that $P_t(\mathcal{C}(\beta_{1,t}, \gamma_t, c_1\sigma(P_t))) \geq 1 - r_0$. Therefore, since P satisfies (C3) and $\mathcal{C}(\beta_{1,t}, \gamma_t, c_1\sigma(P_t)) \in \mathfrak{C}$, and $1 - r_0 > 1 - r_0 - \eta$, it follows from (11.3) that

$$\begin{aligned} P(\mathcal{C}(\beta_{1,t}, \gamma_t, c_1\sigma(P_t))) &\geq P_t(\mathcal{C}(\beta_{1,t}, \gamma_t, c_1\sigma(P_t))) - \sup_{C \in \mathfrak{C}} |P_t(C) - P(C)| \\ &\geq 1 - r_0 - \eta, \end{aligned} \quad (11.6)$$

for t sufficiently large. Since $1 - r_0 - \eta > \epsilon$, according to Lemma 4.1(i) in Lopuhaä *et al* [22], there exists $a_1 > 0$ only depending on c_1 and $(C2_\epsilon)$, such that $\lambda_k(\sigma^2(P_t)\mathbf{V}(\gamma_t)) \geq a_1$. Hence,

$$\lambda_k(\mathbf{V}(\gamma_t)) \geq \frac{a_1}{\sigma^2(P_t)},$$

for t sufficiently large. Since $|\mathbf{V}(\gamma_t)| = 1$, it immediately follows that

$$\lambda_1(\mathbf{V}(\gamma_t)) \leq \left(\frac{\sigma^2(P_t)}{a_1} \right)^{k-1}.$$

Because $\sigma(P_t) \rightarrow \sigma(P)$ according to part (i), there exists $0 < L_1 < L_2 < \infty$, such that

$$L_1 \leq \lambda_k(\mathbf{V}(\gamma_t)) \leq \lambda_1(\mathbf{V}(\gamma_t)) \leq L_2, \quad (11.7)$$

for t sufficiently large. Let $K \subset \mathbb{R}^k \times \mathcal{X}$ be a compact set, such that $P(K) \geq r_0 + \epsilon' \geq r_0 + \epsilon + \eta$, which exists according to condition (C1 $_{\epsilon'}$). Then according to (11.6), it follows from Lemma 4.1(iii) in Lopuhaä *et al* [22] with $a = 1 - r_0 - \eta$, that $\|\beta_{1,t}\| \leq M < \infty$, for some $M > 0$ that only depends on $c_1, a_2, \sigma(P), K$ and (C2 $_{\epsilon}$). We conclude that for t sufficiently large the sequence $\{\beta_{1,t}\}$ lies in a compact subset of $\mathbb{R}^q$ and the sequence $\{\mathbf{V}(\gamma_t)\}$ lies in a compact set $B \subset \mathbb{R}^{k \times k}$. Then similar to the second part of the proof of Theorem 1, conditions (V1) and (V3) yield there exists a compact set $B' \subset \mathbb{R}^{q+l}$, such that for t sufficiently large the sequence $\{(\beta_{1,t}, \gamma_t)\} \subset B'$.

Since ρ_1 satisfies (R2)-(R3), together with of part (i), similar to (11.4), for fixed $(\beta, \gamma) \in \mathbb{R}^k \times \Theta \subset \mathbb{R}^k \times \mathbb{R}^l$,

$$R_{P_t}(\beta, \mathbf{V}(\gamma)) \rightarrow R_P(\beta, \mathbf{V}(\gamma)), \quad (11.8)$$

where R_P is defined in (3.9). For the sake of brevity, let us write $R_t = R_{P_t}$. Since the sequence $\{(\beta_{1,t}, \gamma_t)\}$ lies in a compact set, it has a convergent subsequence $(\beta_{1,t_j}, \gamma_{t_j}) \rightarrow (\beta_{1,L}, \gamma_L)$. Since ρ_1 satisfies (R2)-(R3) and $\mathbf{V}$ satisfies (V1), similar to (11.4), it follows that

$$\lim_{j \rightarrow \infty} R_{t_j}(\beta_{1,t_j}, \mathbf{V}(\gamma_{t_j})) = R_P(\beta_{1,L}, \mathbf{V}(\gamma_L)).$$

Now, suppose that $(\beta_{1,L}, \gamma_L) \neq (\beta_1(P), \gamma(P))$. Then, since $R_P(\beta, \mathbf{V}(\gamma))$ is uniquely minimized at $(\beta_1(P), \gamma(P))$, this would mean that there exists $\epsilon > 0$, such that together with (11.8),

$$\begin{aligned} R_{t_j}(\beta_{1,t_j}, \mathbf{V}(\gamma_{t_j})) &> R_P(\beta_{1,L}, \mathbf{V}(\gamma_L)) - \epsilon \geq R_P(\beta_1(P), \mathbf{V}(\gamma(P))) + 2\epsilon \\ &\geq R_{t_j}(\beta_1(P), \mathbf{V}(\gamma(P))) + \epsilon > R_{t_j}(\beta_1(P), \mathbf{V}(\gamma(P))), \end{aligned}$$

for t_j sufficiently large. This would mean that $(\beta_{1,t_j}, \gamma_{t_j})$ is not the minimizer of $R_{t_j}(\beta, \mathbf{V}(\gamma))$. We conclude that $(\beta_{1,L}, \gamma_L) = (\beta_1(P), \gamma(P))$, which proves part (ii).

Finally, from part (i) and (V1), we have that

$$\mathbf{V}(\theta_1(P_t)) = \sigma^2(P_t)\mathbf{V}(\gamma(P_t)) \rightarrow \sigma^2(P)\mathbf{V}(\gamma(P)) = \mathbf{V}(\theta_1(P)).$$

Because $\mathbf{V}$ is continuous and one-to-one, part (iii) follows. $\square$

Proof of Corollary 3

Proof. Let $\mathbb{P}_n$ be the empirical measure corresponding to the collection $\mathcal{S}_n$. According to the Portmanteau Theorem (e.g., see Theorem 2.1 in [2]), $\mathbb{P}_n$ converges weakly to P , with probability one. Because $(\beta_{0,n}, \theta_{0,n})$ satisfies (3.12), it follows that $(\beta_0(\mathbb{P}_n), \theta_0(\mathbb{P}_n)) = (\beta_{0,n}, \theta_{0,n}) \rightarrow (\beta_0(P), \theta_0(P))$ and that $\sigma_n = \sigma(\mathbb{P}_n)$ the unique solution of (3.8), with $P = \mathbb{P}_n$. Hence, part (i) follows immediately from Theorem 2(i). Furthermore, $(\beta_{1,n}, \gamma_n) = (\beta_1(\mathbb{P}_n), \gamma(\mathbb{P}_n))$ is a local minimum of $R_P(\beta, \mathbf{V}(\gamma))$, with $P = \mathbb{P}_n$, that satisfies (3.10), for $P = \mathbb{P}_n$. Therefore, part (ii) follows immediately from Theorem 2(ii). Finally, since $\theta_{1,n} = \theta_1(\mathbb{P}_n)$, part (iii) follows immediately from Theorem 2(iii). $\square$

Proof of Theorem 3

Proof. First consider the multivariate location-scatter M-functional with auxiliary scale at the distribution F of $\mathbf{y} \mid \mathbf{X}$, for some $\mathbf{X}$ fixed. So F has density $f_{\mu, \Sigma}$ from (5.1). Tatsuoka and Tyler [32] define location-scale M-functionals with auxiliary scale by means of a function $\tilde{\rho} : [0, \infty) \rightarrow [0, 1]$. It relates to our ρ_1 -function as $\rho_1(d) = \tilde{\rho}(d^2)$. The M-functionals with auxiliary scale $\sigma(F)$ are defined to be $\alpha(F)$ and $\mathbf{A}(F) = \sigma^2(F)\mathbf{G}(F)$, where $(\alpha(F), \mathbf{G}(F))$ minimizes

$$\int \rho_1 \left(\frac{\sqrt{(\mathbf{y} - \alpha)^T \mathbf{G}^{-1}(\mathbf{y} - \alpha)}}{\sigma(F)} \right) f_{\mu, \Sigma}(\mathbf{y}) d\mathbf{y}, \quad (11.9)$$

over all $\alpha \in \mathbb{R}^q$ and $\mathbf{G} \in \text{PDS}(k)$ with $|\mathbf{G}| = 1$. Our conditions (R1)-(R2) on ρ_1 imply Condition 2.1 on $\tilde{\rho}$ imposed in Tatsuoka and Tyler [32]. It then follows from Theorem 4.2 in Tatsuoka

and Tyler [32], together with the affine equivariance of the M-functional, that for any minimizer $(\boldsymbol{\alpha}(F), \mathbf{G}(F))$, the multivariate M-functional $(\boldsymbol{\alpha}(F), \mathbf{A}(F)) = (\boldsymbol{\alpha}(F), \sigma^2(F)\mathbf{G}(F))$ has a unique solution

$$(\boldsymbol{\alpha}(F), \mathbf{A}(F)) = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{X}\boldsymbol{\beta}^*, \mathbf{V}(\boldsymbol{\theta}^*)).$$

Since this solution is unique, any candidate minimizer of (11.9) must be of the form

$$(\boldsymbol{\alpha}, \mathbf{G}) = \left(\mathbf{X}\boldsymbol{\beta}, \frac{\mathbf{V}(\boldsymbol{\theta})}{\sigma^2(F)} \right) = (\mathbf{X}\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma})),$$

for some $\boldsymbol{\beta} \in \mathbb{R}^q$ and $\boldsymbol{\gamma} \in \mathbb{R}^\ell$, where we use that $\mathbf{V}$ satisfies (V2). Furthermore,

$$|\boldsymbol{\Sigma}| = |\mathbf{A}(F)| = \sigma^{2k}(F)|\mathbf{G}(F)| = \sigma^{2k}(F),$$

so that $\sigma(F) = |\boldsymbol{\Sigma}|^{1/(2k)}$. Then, for $\mathbf{X}$ fixed, minimizing (11.9) over all $\boldsymbol{\alpha} \in \mathbb{R}^q$ and $\mathbf{G} \in \text{PDS}(k)$ with $|\mathbf{G}| = 1$ is equivalent to minimizing

$$\int \rho_1 \left(\frac{d(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))}{|\boldsymbol{\Sigma}|^{1/(2k)}} \right) f_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{y}) d\mathbf{y}, \quad (11.10)$$

over $\boldsymbol{\beta} \in \mathbb{R}^q$ and $\boldsymbol{\gamma} \in \mathbb{R}^\ell$, such that $|\mathbf{V}(\boldsymbol{\gamma})| = 1$, where d is defined in (7.1). As a consequence, for $\mathbf{X}$ fixed, this minimization problem has a unique solution $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})$, which satisfies $\mathbf{X}\tilde{\boldsymbol{\beta}} = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}^*$ and

$$\mathbf{V}(\tilde{\boldsymbol{\gamma}}) = \frac{\mathbf{V}(\boldsymbol{\theta}^*)}{\sigma^2(F)} = \frac{\boldsymbol{\Sigma}}{|\boldsymbol{\Sigma}|^{1/k}}. \quad (11.11)$$

Since $\mathbf{X}$ has full rank, with probability one, it follows that $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}^*$.

We can transfer this result to the minimization of $R_P(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$ as follows. Because P is absolutely continuous, it satisfies condition (C0). Hence, according to Lemma 1 the solution $\sigma(P)$ of (3.8) is unique. Because $\boldsymbol{\beta}_0(P) = \boldsymbol{\beta}^*$ and $\boldsymbol{\Gamma}(\boldsymbol{\theta}_0(P)) = \boldsymbol{\Gamma}(\boldsymbol{\theta}^*) = \mathbf{V}(\boldsymbol{\theta}^*)/|\mathbf{V}(\boldsymbol{\theta}^*)|^{1/k} = \boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/k}$, in (3.8) we have that

$$\int \rho_0 \left(\frac{d(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}_0(P), \boldsymbol{\Gamma}(\boldsymbol{\theta}_0(P)))}{\sigma} \right) dP(\mathbf{s}) = \mathbb{E} \left[\mathbb{E} \left[\rho_0 \left(\frac{d(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}^*, \boldsymbol{\Sigma})}{\sigma/|\boldsymbol{\Sigma}|^{1/(2k)}} \right) \middle| \mathbf{X} \right] \right].$$

The inner expectation is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$, which has the same distribution as $\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \mathbf{X}\boldsymbol{\beta}^*$, where $\mathbf{z}$ has density $f_{\mathbf{0}, \mathbf{I}_k}$. This means that we must solve $\sigma(P)$ from

$$\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \rho_0 \left(\frac{\|\mathbf{z}\|}{\sigma/|\boldsymbol{\Sigma}|^{1/(2k)}} \right) = b_0. \quad (11.12)$$

Furthermore, since P is absolutely continuous, it satisfies (C1 $_\epsilon$) and (C2 $_\epsilon$), for some $0 < \epsilon \leq 1 - r_0$, where $r_0 = b_0/\sup \rho_0$. Hence, since $\mathbf{V}$ satisfies (V1)-(V3), according to Theorem 1 there exists $(\boldsymbol{\beta}_1(P), \boldsymbol{\gamma}(P)) \in \mathfrak{D}$ that minimizes $R_P(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$ and $\boldsymbol{\theta}_1(P) \in \boldsymbol{\Theta}$ that is the unique solution of (3.11). But then, there must be an $\mathbf{X}$ such that

$$\int \rho_1 \left(\frac{d(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}_1(P), \mathbf{V}(\boldsymbol{\gamma}(P)))}{\sigma(P)} \right) f_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{y}) d\mathbf{y} \leq \int \rho_1 \left(\frac{d(\mathbf{y}, \mathbf{X}\tilde{\boldsymbol{\beta}}, \mathbf{V}(\tilde{\boldsymbol{\gamma}}))}{\sigma(P)} \right) f_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{y}) d\mathbf{y}.$$

But since for any $\mathbf{X}$ fixed, $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}})$ is the unique minimizer of (11.10), with probability one, we must have $\boldsymbol{\beta}_1(P) = \tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}(P) = \tilde{\boldsymbol{\gamma}}$, with probability one. Together with (11.11) this proves part(i).

To prove part (ii), note that $\boldsymbol{\theta}_1(P)$ satisfies

$$\mathbf{V}(\boldsymbol{\theta}_1(P)) = \sigma^2(P)\mathbf{V}(\boldsymbol{\gamma}(P)) = \frac{\sigma^2(P)}{|\boldsymbol{\Sigma}|^{1/k}}\boldsymbol{\Sigma} = \frac{\sigma^2(P)}{|\boldsymbol{\Sigma}|^{1/k}}\mathbf{V}(\boldsymbol{\theta}^*) = \mathbf{V} \left(\frac{\boldsymbol{\theta}^*\sigma^2(P)}{|\boldsymbol{\Sigma}|^{1/k}} \right),$$

where we use that $\mathbf{V}$ has a linear structure. By identifiability this means that $\boldsymbol{\theta}_1(P) = \boldsymbol{\theta}^*\sigma^2(P)/|\boldsymbol{\Sigma}|^{1/k}$.

Finally, for part (iii) suppose that $b_0 = \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \rho_0(\|\mathbf{z}\|)$. It follows immediately from solving (11.12) that $\sigma(P) = |\boldsymbol{\Sigma}|^{1/(2k)}$. This finishes the proof. $\square$

11.3 Proofs for Section 6

Proof of Theorem 4

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

$$m \leq \min \left(\lceil nr_0 \rceil, \lceil n - nr_0 \rceil - \kappa(\mathcal{S}_n), n\epsilon_n^*(\beta_{0,n}, \theta_{0,n}, \mathcal{S}_n) \right) - 1$$

Let $\mathcal{S}'_m$ be the corrupted collection of points. Write $(\beta_{0,m}, \theta_{0,m}) = (\beta_{0,n}(\mathcal{S}'_m), \theta_{0,n}(\mathcal{S}'_m))$, $\mathbf{V}_{0,m} = \mathbf{V}(\theta_{0,n}(\mathcal{S}'_m))$, $\Gamma_{0,m} = \Gamma(\theta_{0,n}(\mathcal{S}'_m))$. Let $\mathbb{P}'_m$ be the empirical measure corresponding to the corrupted collection $\mathcal{S}'_m$ of n points. We must show that there exists a pair $(\beta_{1,m}, \gamma_m) = (\beta_{1,n}(\mathcal{S}'_m), \gamma(\mathcal{S}'_m)) \in \mathfrak{D}$ that satisfies (3.5), for the corrupted collection $\mathcal{S}'_m$, and $\theta_{1,m} = \theta_{1,n}(\mathcal{S}'_m) \in \Theta$ that satisfies $\mathbf{V}(\theta_{1,m}) = \sigma_m^2 \mathbf{V}(\gamma_m)$, and that all pairs $(\beta_{1,m}, \gamma_m)$ satisfying (3.5) and $\theta_{1,m}$ do not break down.

We first show that a solution $\sigma_m = \sigma_n(\mathcal{S}'_m)$ of equation (3.2) exists, for the collection $\mathcal{S}'_m$. Note that the maximum number of points of $\mathcal{S}'_m$ that lie in the same hyperplane is $m + \kappa(\mathcal{S}_n)$. Because $m \leq \lceil n - nr_0 \rceil - \kappa(\mathcal{S}_n) - 1$, it follows that

$$m + \kappa(\mathcal{S}_n) \leq \lceil n - nr_0 \rceil - 1 < n(1 - r_0).$$

This means that $\mathbb{P}'_m$ satisfies condition (C0). Since ρ_0 satisfies (R1)-(R3), according to Lemma 1, there exists a unique solution $\sigma_m > 0$ of (3.2). Furthermore, from (R1)-(R3) we have that $\rho_0(s)$, for $s \in [0, c_0]$, varies continuously between zero and $\sup \rho_0$. Since $m \leq \lceil nr_0 \rceil - 1 < nr_0$, there exists $\eta > 0$, such that

$$\frac{m}{n} \sup \rho_0 + \eta < b_0.$$

Because $\beta_{0,m}$ and $\theta_{0,m}$ do not break down, there exist $M > 0$ and $0 < L_1 \leq L_2 < \infty$, not depending on $\mathcal{S}'_m$, such that $\|\beta_{0,m}\| \leq M$ and $L_1 \leq \lambda_k(\mathbf{V}_{0,m}) \leq \lambda_1(\mathbf{V}_{0,m}) \leq L_2$. This means that for all $\mathbf{s}_i \in \mathcal{S}_n$:

$$\sqrt{(\mathbf{y}_i - \mathbf{X}_i \beta_{0,m})^T \Gamma_{0,m}^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta_{0,m})} \leq \frac{\|\mathbf{y}_i - \mathbf{X}_i \beta_{0,m}\|}{\sqrt{\lambda_k(\Gamma_{0,m})}} \leq \frac{\|\mathbf{y}_i\| + \|\mathbf{X}_i\| \|\beta_{0,m}\|}{\sqrt{\lambda_k(\Gamma_{0,m})}},$$

where

$$\lambda_k(\Gamma_{0,m}) = \frac{\lambda_k(\mathbf{V}_{0,m})}{|\mathbf{V}_{0,m}|^{1/k}} \geq \frac{L_1}{(L_2^k)^{1/k}} = \frac{L_1}{L_2}.$$

Hence, for all $\mathbf{s}_i \in \mathcal{S}_n$, we obtain

$$d_{i,0} = \sqrt{(\mathbf{y}_i - \mathbf{X}_i \beta_{0,m})^T \Gamma_{0,m}^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta_{0,m})} \leq \left(\max_{\mathbf{s}_i \in \mathcal{S}_n} \|\mathbf{y}_i\| + M \max_{\mathbf{s}_i \in \mathcal{S}_n} \|\mathbf{X}_i\| \right) \sqrt{L_2/L_1}.$$

This means, there exists $0 < K < \infty$, only depending on $\mathcal{S}_n$, such that $\max_{\mathbf{s}_i \in \mathcal{S}_n} d_{i,0} \leq K$. Since ρ_0 is continuous and $0 < \eta < b_0$, we can define $\delta > 0$, such that $\rho_0(\delta) = \eta$. Let $s_0 = K/\delta$. Then,

$$\begin{aligned} \frac{1}{n} \sum_{(\mathbf{y}_i, \mathbf{X}_i) \in \mathcal{S}'_m} \rho_0 \left(\frac{d_{i,0}}{s_0} \right) &= \frac{1}{n} \sum_{(\mathbf{y}_i, \mathbf{X}_i) \in \mathcal{S}'_m \cap \mathcal{S}_n} \rho_0 \left(\frac{d_{i,0}}{s_0} \right) + \frac{1}{n} \sum_{(\mathbf{y}_i, \mathbf{X}_i) \in \mathcal{S}'_m \setminus \mathcal{S}_n} \rho_0 \left(\frac{d_{i,0}}{s_0} \right) \\ &\leq \frac{n-m}{n} \rho_0 \left(\frac{K}{s_0} \right) + \frac{m}{n} \sup \rho_0 \\ &\leq \rho_0(\delta) + \frac{m}{n} \sup \rho_0 \\ &= \eta + \frac{m}{n} \sup \rho_0 < b_0. \end{aligned}$$

As σ_m is the solution of (3.2), we must have that $\sigma_m \leq s_0$.

Next, we show that there exists a pair $(\beta_{1,m}, \gamma_m) = (\beta_{1,n}(\mathcal{S}'_m), \gamma(\mathcal{S}'_m)) \in \mathfrak{D}$ that minimizes

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

Any minimizer $(\beta_{1,m}, \gamma_m)$ of $R_m(\beta, \mathbf{V}(\gamma))$ must satisfy (3.5). Together with (3.3) similar to (11.2), this implies

$$\mathbb{P}'_m(\mathcal{C}(\beta_{1,m}, \gamma_m, c_1 \sigma_m)) \geq 1 - r_0.$$

It then follows that the cylinder $\mathcal{C}(\beta_{1,m}, \gamma_m, c_1 \sigma_m)$ contains at least $\lceil n - nr_0 \rceil$ number of points from the corrupted collection $\mathcal{S}'_m$. Furthermore, any corrupted collection $\mathcal{S}'_m$ contains

$$\lceil n - nr_0 \rceil - m \geq \kappa(\mathcal{S}_n) + 1 \quad (11.13)$$

points of the original collection $\mathcal{S}_n$. This means that the cylinder $\mathcal{C}(\beta_{1,m}, \gamma_m, c_1 \sigma_m)$ must contain a non-empty simplex only depending on the original collection $\mathcal{S}_n$. This implies that

$$\lambda_k(\sigma_m^2 \mathbf{V}(\gamma_m)) \geq a_1 > 0, \quad (11.14)$$

where a_1 only depends on the original collection $\mathcal{S}_n$. Hence, $\lambda_k(\mathbf{V}(\gamma_m)) \geq a_1/s_0^2 > 0$. Since $|\mathbf{V}(\gamma_m)| = 1$, it immediately follows that $\lambda_1(\mathbf{V}(\gamma_m)) \leq (s_0^2/a_1)^{k-1} < \infty$. Furthermore, recall that $\mathcal{C}(\beta_{1,m}, \gamma_m, c_1 \sigma_m)$ contains a subset J_0 of $\kappa(\mathcal{S}_n) + 1$ points from the original collection $\mathcal{S}_n$, according to (11.13). By definition, $\kappa(\mathcal{S}_n) + 1$ original points cannot be on the same hyperplane, so that

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

where the first infimum runs over all subsets $J \subset \mathcal{S}_n$ of $\kappa(\mathcal{S}_n) + 1$ points. By definition of α_n , there exists an original point $\mathbf{s}_0 \in J_0 \subset \mathcal{S}_n \cap \mathcal{C}(\beta_{1,m}, \gamma_m, c_1 \sigma_m)$, such that

$$\|\beta_{1,m}\| = \|\mathbf{X}_0 \beta_{1,m}\| \times \frac{\|\beta_{1,m}\|}{\|\mathbf{X}_0 \beta_{1,m}\|} \leq \frac{1}{\alpha_n} \|\mathbf{X}_0 \beta_{1,m}\|.$$

Because $\mathbf{s}_0 \in \mathcal{C}(\beta_{1,m}, \gamma_m, c_1 \sigma_m)$, it follows that

$$\begin{aligned} \|\mathbf{y}_0 - \mathbf{X}_0 \beta_{1,m}\|^2 &\leq (\mathbf{y}_0 - \mathbf{X}_0 \beta_{1,m})^T (\sigma_m^2 \mathbf{V}(\gamma_m))^{-1} (\mathbf{y}_0 - \mathbf{X}_0 \beta_{1,m}) \\ &\leq c_1^2 \lambda_1(\sigma_m^2 \mathbf{V}(\gamma_m)) \leq a_2, \end{aligned}$$

where $a_2 = c_1^2 s_0^2 (s_0^2/a_1)^{k-1}$ only depends on the original collection $\mathcal{S}_n$. Because $\mathbf{s}_0 \in \mathcal{S}_n$, we have that

$$\|\mathbf{X}_0 \beta_{1,m}\| \leq \sqrt{a_2} + \max_{(\mathbf{y}_i, \mathbf{X}_i) \in \mathcal{S}_n} \|\mathbf{y}_i\| < \infty.$$

We conclude that there exists a compact set of $\mathbb{R}^q$ that contains $\beta_{1,m}$ and a compact set $B \subset \mathbb{R}^{k \times k}$ that contains $\mathbf{V}(\gamma_m)$. This means that γ_m is in the pre-image $\mathbf{V}^{-1}(B)$, and with conditions (V1) and (V3), it follows that $\mathbf{V}^{-1}(B)$ is a compact set in Θ . We conclude that for minimizing $R_m(\beta, \mathbf{V}(\gamma))$ we can restrict ourselves to a compact set $B' \subset \mathfrak{D}$, only depending on the original collection $\mathcal{S}_n$. Because ρ_1 and $\mathbf{V}$ are continuous, it follows that $R_m(\beta, \mathbf{V}(\gamma))$ is a continuous function of (β, γ) , so it attains a minimum on B' . This also means that there exists a pair $(\beta_{1,m}, \gamma_m) \in \mathfrak{D}$ that satisfies (3.5), for the corrupted collection $\mathcal{S}'_m$. Now, consider any pair $(\beta_{1,m}, \gamma_m)$ that satisfies (3.5). Then the reasoning above yields that there is a compact set B' , only depending on the original collection $\mathcal{S}_n$, that contains $(\beta_{1,m}, \gamma_m)$. Hence, $\beta_{1,m}$ and γ_m do not break down.

Finally, because $\mathbf{V}$ satisfies (V2), there exists $\theta_{1,m} = \theta_{1,n}(\mathcal{S}'_m) \in \Theta$ that satisfies $\mathbf{V}(\theta_{1,m}) = \sigma_m^2 \mathbf{V}(\gamma_m)$. From (11.14), we have that $\lambda_k(\mathbf{V}(\theta_{1,m})) = \lambda_k(\sigma_m^2 \mathbf{V}(\gamma_m)) \geq a_1 > 0$, where a_1 only depends on the original collection $\mathcal{S}_n$. Furthermore,

$$\lambda_1(\mathbf{V}(\theta_{1,m})) = \sigma_m^2 \lambda_1(\mathbf{V}(\gamma_m)) \leq s_0^2 \left(\frac{s_0^2}{a_1} \right)^{k-1} < \infty.$$

This means that $\theta_{1,m}$ does not break down. □

11.4 Proofs for Section 7

Proof of Proposition 1

Proof. In STAGE 3 for the MM-functional, we are considering local minima of $R_P(\beta, \mathbf{V}(\gamma))$ that satisfy $|\mathbf{V}(\gamma)| = 1$, or equivalently $\log |\mathbf{V}(\gamma)| = 0$. The Lagrangian corresponding to this constrained minimization problem is given by

$$L_P(\xi, \lambda) = R_P(\beta, \mathbf{V}(\gamma)) - \lambda \log |\mathbf{V}(\gamma)|.$$

Suppose $\xi(P) = (\beta_1(P), \gamma(P))$ is a local minimum of $R_P(\beta, \mathbf{V}(\gamma))$. Then this is also a zero of the partial derivatives $\partial L_P / \partial \beta$, $\partial L_P / \partial \gamma$, and $\partial L_P / \partial \lambda$. Let us write $\sigma_P = \sigma(P)$ and consider

$$R_P(\beta, \mathbf{V}(\gamma)) = \int \rho_1 \left(\frac{d}{\sigma_P} \right) dP(\mathbf{s})$$

where $d = d(\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\gamma))$, as defined in (7.1). We find that

$$\begin{aligned} \frac{\partial \rho_1(d/\sigma_P)}{\partial \beta} &= \frac{1}{2\sigma_P} u_1 \left(\frac{d}{\sigma_P} \right) \mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) \\ \frac{\partial \rho_1(d/\sigma_P)}{\partial \gamma_j} &= \frac{1}{2\sigma_P} u_1 \left(\frac{d}{\sigma_P} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta), \end{aligned}$$

for $j = 1, \dots, l$. Similar to the proof of Lemma 11.2 in Lopuhaä *et al* [22], using that ρ_1 and $\mathbf{V}$ satisfy (R2), (R4) and (V4), respectively, we find that at $\xi(P)$ it holds that $\|\partial \rho_1(d_P/\sigma_P) / \partial \beta\| \leq C_1 \|\mathbf{X}\|$ and $\|\partial \rho_1(d_P/\sigma_P) / \partial \gamma_j\| \leq C_2$, for universal constants $0 < C_1, C_2 < \infty$ only depending on P and σ_P , where $d_P = d(\mathbf{y}, \mathbf{X}\beta_1(P), \mathbf{V}(\gamma(P)))$. Since $\mathbb{E}_P \|\mathbf{X}\| < \infty$, this implies that under conditions (R4) and (V4), we may interchange the order of integration and differentiation in $\partial L_P / \partial \beta$ and $\partial L_P / \partial \gamma$, on a neighborhood of $\xi(P)$. Similar to the derivation of equations (21) in Lopuhaä *et al* [22], it follows that besides the constraint $\log |\mathbf{V}(\gamma)| = 0$, the pair $(\xi(P), \lambda_P)$ satisfies

$$\begin{aligned} \int u_1 \left(\frac{d}{\sigma_P} \right) \mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) dP(\mathbf{s}) &= \mathbf{0} \\ \frac{1}{2\sigma_P} \int u_1 \left(\frac{d}{\sigma_P} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) dP(\mathbf{s}) & \\ + \lambda \operatorname{tr} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) &= 0, \end{aligned} \tag{11.15}$$

for $j = 1, \dots, l$, where $u_1(s) = \rho'_1(s)/s$ and $d = d(\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\gamma))$, as defined by (7.1), and where we abbreviate $\mathbf{V}(\gamma)$ by $\mathbf{V}$. To solve λ_P from the second set of equations, we multiply the j -th equation by γ_j and then sum over $j = 1, \dots, l$. This leads to

$$\begin{aligned} \frac{1}{2\sigma_P} \int u_1 \left(\frac{d}{\sigma_P} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \left(\sum_{j=1}^l \gamma_j \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) dP(\mathbf{s}) & \\ + \lambda \operatorname{tr} \left(\mathbf{V}^{-1} \sum_{j=1}^l \gamma_j \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) &= 0, \end{aligned}$$

which is solved by

$$\lambda_P = \frac{- \int u_1(d/\sigma_P) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \left(\sum_{t=1}^l \gamma_t (\partial \mathbf{V} / \partial \gamma_t) \right) \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) dP(\mathbf{s})}{2\sigma_P \operatorname{tr} \left(\mathbf{V}^{-1} \sum_{t=1}^l \gamma_t (\partial \mathbf{V} / \partial \gamma_t) \right)}.$$

When we insert this back into the second equation in (11.15), we find

$$\begin{aligned} & -\operatorname{tr} \left(\mathbf{V}^{-1} \sum_{t=1}^l \gamma_t \frac{\partial \mathbf{V}}{\partial \gamma_t} \right) \int u_1 \left(\frac{d}{\sigma_P} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) dP(\mathbf{s}) \\ & + \operatorname{tr} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) \int u_1 \left(\frac{d}{\sigma_P} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \left(\sum_{t=1}^l \gamma_t \frac{\partial \mathbf{V}}{\partial \gamma_t} \right) \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) dP(\mathbf{s}) = 0, \end{aligned}$$

or briefly

$$\int u_1 \left(\frac{d}{\sigma_P} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \mathbf{H}_{1,j} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta) dP(\mathbf{s}) = 0, \quad j = 1, \dots, l, \quad (11.16)$$

where

$$\mathbf{H}_{1,j} = \operatorname{tr} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) \left(\sum_{t=1}^l \gamma_t \frac{\partial \mathbf{V}}{\partial \gamma_t} \right) - \operatorname{tr} \left(\mathbf{V}^{-1} \sum_{t=1}^l \gamma_t \frac{\partial \mathbf{V}}{\partial \gamma_t} \right) \frac{\partial \mathbf{V}}{\partial \gamma_j}.$$

Because $\sum_{j=1}^l \gamma_j \mathbf{H}_{1,j} = \mathbf{0}$, the system of equations (11.16) is linearly dependent. Similar to Lopuhaä *et al* [22], we subtract the constraint from each equation. Here, for each $j = 1, \dots, l$, we subtract the term

$$\operatorname{tr} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) \log |\mathbf{V}|$$

from the left hand side of equation (11.16). This finishes the proof. $\square$

Proof of Proposition 2

Proof. When $\mathbf{V}$ is of the form (7.5), then $\partial \mathbf{V} / \partial \gamma_j = \mathbf{L}_j$ and $\sum_{j=1}^l \gamma_j \partial \mathbf{V} / \partial \gamma_j = \mathbf{V}$. In this case, $\mathbf{H}_{1,j} = \operatorname{tr} (\mathbf{V}^{-1} \mathbf{L}_j) \mathbf{V} - k \mathbf{L}_j$, and $\Psi_{\gamma,j}$ in Proposition 1 becomes

$$\begin{aligned} \Psi_{\gamma,j}(\mathbf{s}, \xi, \sigma) &= \operatorname{tr} (\mathbf{V}^{-1} \mathbf{L}_j) \left\{ u_1 \left(\frac{d}{\sigma} \right) d^2 - \log |\mathbf{V}| \right\} \\ &\quad - k u_1 \left(\frac{d}{\sigma} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \mathbf{L}_j \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\beta). \end{aligned}$$

Using that

$$\operatorname{tr}(\mathbf{A}^T \mathbf{B}) = \operatorname{vec}(\mathbf{A})^T \operatorname{vec}(\mathbf{B}) \quad (11.17)$$

the right hand side can be written as

$$-\operatorname{vec} \left(k u_1 \left(\frac{d}{\sigma} \right) (\mathbf{y} - \mathbf{X}\beta)(\mathbf{y} - \mathbf{X}\beta)^T - v_1 \left(\frac{d}{\sigma} \right) \sigma^2 \mathbf{V} - \mathbf{V} \log |\mathbf{V}| \right)^T \operatorname{vec} (\mathbf{V}^{-1} \mathbf{L}_j \mathbf{V}^{-1}),$$

where $u_1(s) = \rho'_1(s)/s$ and $v_1(s) = u_1(s)s^2 = \rho'_1(s)s$. The functions $\Psi_{\gamma,j}$, for $j = 1, \dots, l$, can be combined in one expression for the vector valued function Ψ_{γ} as follows. First note that

$$\operatorname{vec} (\mathbf{V}^{-1} \mathbf{L}_j \mathbf{V}^{-1}) = (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \operatorname{vec} (\mathbf{L}_j)$$

for $j = 1, \dots, l$. Then, the column vector $\Psi_{\gamma} = (\Psi_{\gamma,1}, \dots, \Psi_{\gamma,l})$ can be written as

$$\Psi_{\gamma}(\mathbf{s}, \xi, \sigma) = -\mathbf{L}^T (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \operatorname{vec} (\Psi_{\mathbf{V}}(\mathbf{s}, \xi, \sigma)),$$

with $\Psi_{\mathbf{V}}$ defined in (7.8). $\square$

11.5 Proofs for Section 8

Lemma 2. Suppose that ρ_1 satisfies (R4) and that $\mathbf{V}$ satisfies (V4). Let $\sigma(P)$ be the solution of equation (3.8) and let $\boldsymbol{\xi}(P) \in \mathfrak{D}$ be a local minimum of $R_P(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$. Suppose that $\text{IF}(\mathbf{s}; \sigma, P)$ exists. Let $\boldsymbol{\xi}(P_{h,\mathbf{s}}) \in \mathfrak{D}$ be a local minimum of $R_P(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$ with $P = P_{h,\mathbf{s}}$, and suppose that $\boldsymbol{\xi}(P_{h,\mathbf{s}}) \rightarrow \boldsymbol{\xi}(P)$, as $h \downarrow 0$. Let Λ be defined by (8.2) with Ψ from (7.3) and suppose Λ is continuously differentiable with a non-singular derivative $\mathbf{D}_\xi = \partial\Lambda/\partial\xi$ and derivative $\mathbf{D}_\sigma = \partial\Lambda/\partial\sigma$ at $(\boldsymbol{\xi}(P), \sigma(P))$. Then for $\mathbf{s} \in \mathbb{R}^k \times \mathbb{R}^{kq}$,

$$(i) \text{ IF}(\mathbf{s}; \boldsymbol{\xi}, P) = -\mathbf{D}_\xi^{-1} \left\{ \Psi(\mathbf{s}, \boldsymbol{\xi}(P), \sigma(P)) + \mathbf{D}_\sigma \text{IF}(\mathbf{s}; \sigma, P) \right\}.$$

Let $\boldsymbol{\theta}_1(P)$ and $\boldsymbol{\theta}_1(P_{h,\mathbf{s}})$ be solutions of equation (3.11) and equation (3.11) with $P = P_{h,\mathbf{s}}$, respectively.

(ii) Then

$$\begin{aligned} \text{IF}(\mathbf{s}; \text{vec}\mathbf{V}(\boldsymbol{\theta}_1), P) &= 2\sigma(P)\text{vec}(\mathbf{V}(\boldsymbol{\gamma}(P)))\text{IF}(\mathbf{s}; \sigma, P) \\ &\quad + \sigma^2(P) \frac{\partial \text{vec}(\mathbf{V}(\boldsymbol{\gamma}(P)))}{\partial \boldsymbol{\gamma}^T} \text{IF}(\mathbf{s}; \boldsymbol{\gamma}, P). \end{aligned}$$

In addition, suppose that the $k^2 \times l$ matrix $\mathbf{D}_V = \partial \text{vec}(\mathbf{V}(\boldsymbol{\theta}_1(P)))/\partial \boldsymbol{\theta}^T$ has full rank. Then

$$(iii) \text{ IF}(\mathbf{s}; \boldsymbol{\theta}_1, P) = (\mathbf{D}_V^T \mathbf{D}_V)^{-1} \mathbf{D}_V^T \text{IF}(\mathbf{s}; \text{vec}\mathbf{V}(\boldsymbol{\theta}_1), P).$$

Proof. Denote $\boldsymbol{\xi}_{h,\mathbf{s}} = \boldsymbol{\xi}(P_{h,\mathbf{s}}) = (\boldsymbol{\beta}_1(P_{h,\mathbf{s}}), \boldsymbol{\gamma}(P_{h,\mathbf{s}}))$ and $\sigma_{h,\mathbf{s}} = \sigma(P_{h,\mathbf{s}})$, and write $\boldsymbol{\xi}_P = \boldsymbol{\xi}(P)$ and $\sigma_P = \sigma(P)$. Then $(\boldsymbol{\xi}_{h,\mathbf{s}}, \sigma_{h,\mathbf{s}})$ satisfies the score equation (7.2) for P equal to $P_{h,\mathbf{s}}$. We decompose as follows

$$\begin{aligned} 0 &= \int \Psi(\mathbf{s}, \boldsymbol{\xi}_{h,\mathbf{s}}, \sigma_{h,\mathbf{s}}) dP_{h,\mathbf{s}}(\mathbf{s}) \\ &= (1-h)\Lambda(\boldsymbol{\xi}_{h,\mathbf{s}}, \sigma_{h,\mathbf{s}}) + h \left(\Psi(\mathbf{s}, \boldsymbol{\xi}_{h,\mathbf{s}}, \sigma_{h,\mathbf{s}}) - \Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \right) + h\Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P), \end{aligned} \tag{11.18}$$

where Ψ and Λ are defined by (7.3) and (8.2), respectively. We first determine the order of $\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P$, as $h \downarrow 0$. Because ρ_1 and $\mathbf{V}$ satisfy (R4) and (V4), respectively, it follows that $\Psi(\mathbf{s}, \boldsymbol{\xi}_{h,\mathbf{s}}, \sigma_{h,\mathbf{s}}) \rightarrow \Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)$, as $h \downarrow 0$. Because Λ is continuous differentiable at $(\boldsymbol{\xi}_P, \sigma_P)$, we have that

$$\begin{aligned} \Lambda(\boldsymbol{\xi}_{h,\mathbf{s}}, \sigma_{h,\mathbf{s}}) &= \Lambda(\boldsymbol{\xi}_P, \sigma_{h,\mathbf{s}}) + \frac{\partial \Lambda(\boldsymbol{\xi}_P, \sigma_{h,\mathbf{s}})}{\partial \boldsymbol{\xi}} (\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P) + o(\|\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P\|) \\ &= \Lambda(\boldsymbol{\xi}_P, \sigma_{h,\mathbf{s}}) + \frac{\partial \Lambda(\boldsymbol{\xi}_P, \sigma_P)}{\partial \boldsymbol{\xi}} (\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P) + o(\|\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P\|) \\ &= \Lambda(\boldsymbol{\xi}_P, \sigma_{h,\mathbf{s}}) + \mathbf{D}_\xi (\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P) + o(\|\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P\|). \end{aligned} \tag{11.19}$$

In general $\Lambda(\boldsymbol{\xi}_P, \sigma_{h,\mathbf{s}}) \neq 0$. Since $\text{IF}(\mathbf{s}; \sigma, P)$ exists, we find

$$\begin{aligned} \Lambda(\boldsymbol{\xi}_P, \sigma_{h,\mathbf{s}}) &= \Lambda(\boldsymbol{\xi}_P, \sigma_P) + \frac{\partial \Lambda(\boldsymbol{\xi}_P, \sigma_P)}{\partial \sigma} (\sigma_{h,\mathbf{s}} - \sigma_P) + o(|\sigma_{h,\mathbf{s}} - \sigma_P|) \\ &= h\mathbf{D}_\sigma \text{IF}(\mathbf{s}; \sigma, P) + o(h). \end{aligned} \tag{11.20}$$

Here we also use that $\Lambda(\boldsymbol{\xi}_P, \sigma_P) = 0$, because $\boldsymbol{\xi}_P$ is a solution of (7.2). Inserting (11.19) and (11.20) in (11.18), yields

$$0 = h\mathbf{D}_\sigma \text{IF}(\mathbf{s}; \sigma, P) + \mathbf{D}_\xi (\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P) + h\Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) + o(\|\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P\|) + o(h).$$

Since $\mathbf{D}_\xi$ is non-singular, this means that $\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P = O(h)$. When we insert this in the previous equation and divide by h , we obtain

$$\frac{\boldsymbol{\xi}_{h,\mathbf{s}} - \boldsymbol{\xi}_P}{h} = -\mathbf{D}_\xi^{-1} \{ \mathbf{D}_\sigma \text{IF}(\mathbf{s}; \sigma, P) + \Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \} + o(1).$$

After letting $h \downarrow 0$, this proves the first part of the lemma.

Since $\mathbf{V}$ satisfies (V4), it holds that

$$\text{vec}(\mathbf{V}(\gamma_{h,s})) - \text{vec}(\mathbf{V}(\gamma_P)) = \frac{\partial \text{vec}(\mathbf{V}(\gamma_P))}{\partial \gamma^T} (\gamma_{h,s} - \gamma_P) + o(\|\gamma_{h,s} - \gamma_P\|).$$

From part (i), this means that $\text{IF}(\mathbf{s}; \text{vec} \mathbf{V}(\gamma), P)$ exists, and is given by

$$\text{vec}(\text{IF}(\mathbf{s}; \mathbf{V}(\gamma), P)) = \frac{\partial \text{vec}(\mathbf{V}(\gamma_P))}{\partial \gamma^T} \text{IF}(\mathbf{s}; \gamma, P).$$

Since $\theta_1(P)$ satisfies (3.11), it follows that $\mathbf{V}(\theta_1(P)) = \sigma_P^2 \mathbf{V}(\gamma_P)$, and similarly for $\theta_1(P_{h,s})$. This implies that

$$\begin{aligned} & \text{vec} \mathbf{V}(\theta_1(P_{h,s})) - \text{vec} \mathbf{V}(\theta_1(P)) \\ &= \sigma_{h,s}^2 \text{vec} \mathbf{V}(\gamma_{h,s}) - \sigma_P^2 \text{vec} \mathbf{V}(\gamma_P) \\ &= \text{vec} \mathbf{V}(\gamma_P) (\sigma_{h,s}^2 - \sigma_P^2) + \sigma_{h,s}^2 (\text{vec} \mathbf{V}(\gamma_{h,s}) - \text{vec} \mathbf{V}(\gamma_P)) \\ &= \text{vec} \mathbf{V}(\gamma_P) (2\sigma_P + o(1)) (\sigma_{h,s} - \sigma_P) + (\sigma_P + o(1)) (\text{vec} \mathbf{V}(\gamma_{h,s}) - \text{vec} \mathbf{V}(\gamma_P)). \end{aligned}$$

After dividing by h and letting $h \downarrow 0$, this proves the second part of the lemma.

Finally, since $\mathbf{V}$ satisfies (V4), as before we can write

$$\text{vec}(\mathbf{V}(\theta_1(P_{h,s})) - \text{vec}(\mathbf{V}(\theta_1(P))) = \mathbf{D}_V(\theta_1(P_{h,s}) - \theta_1(P)) + o(\|\theta_1(P_{h,s}) - \theta_1(P)\|).$$

Because $\text{IF}(\mathbf{s}; \text{vec} \mathbf{V}(\theta_1), P)$ exists and $\mathbf{D}_V$ has full rank, it follows that $\theta_1(P_{h,s}) - \theta_1(P) = O(h)$. When insert this in the previous equation, then after dividing by h and letting $h \downarrow 0$, the limit exists and we obtain

$$\text{IF}(\mathbf{s}; \text{vec} \mathbf{V}(\theta_1), P) = \mathbf{D}_V \text{IF}(\mathbf{s}; \theta_1, P).$$

Because $\mathbf{D}_V$ has full rank, we can multiply from the left with $(\mathbf{D}_V^T \mathbf{D}_V)^{-1} \mathbf{D}_V^T$, which proves part three. $\square$

For $\zeta = (\beta, \theta) \in \mathbb{R}^q \times \Theta$, let

$$\Psi_0(\mathbf{s}, \zeta, \sigma) = \rho_0 \left(\frac{\sqrt{(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{\Gamma}(\theta)^{-1} (\mathbf{y} - \mathbf{X}\beta)}}{\sigma} \right) - b_0, \quad (11.21)$$

where $\mathbf{\Gamma}$ is defined in (3.1), and define

$$\Lambda_0(\zeta, \sigma) = \int \Psi_0(\mathbf{s}, \zeta, \sigma) dP(\mathbf{s}). \quad (11.22)$$

Lemma 3. Suppose that ρ_0 satisfies (R4) and that $\mathbf{V}$ satisfies (V4). Let $\zeta_0 = (\beta_0, \theta_0)$ be the pair of initial functionals and suppose that $\text{IF}(\mathbf{s}, \zeta_0, P)$ exists. Let $\sigma(P)$ and $\sigma(P_{h,s})$ be solutions of equation (3.8) and equation (3.8) with $P = P_{h,s}$, respectively, and suppose that $\sigma(P_{h,s}) \rightarrow \sigma(P)$, as $h \downarrow 0$. Let Λ_0 be defined in (11.22) and suppose it is continuously differentiable with derivatives $D_{0,\sigma} = \partial \Lambda_0 / \partial \sigma \neq 0$ and $\mathbf{D}_{0,\zeta} = \partial \Lambda_0 / \partial \zeta \in \mathbb{R}^{q+l}$ at $(\zeta_0(P), \sigma(P))$. Then for $\mathbf{s} \in \mathbb{R}^k \times \mathbb{R}^{kq}$,

$$\text{IF}(\mathbf{s}; \sigma, P) = -D_{0,\sigma}^{-1} \left\{ \Psi_0(\mathbf{s}, \zeta_0(P), \sigma(P)) + \mathbf{D}_{0,\zeta}^T \text{IF}(\mathbf{s}; \zeta_0, P) \right\}.$$

Proof. Denote $\zeta_{h,s} = \zeta_0(P_{h,s})$, $\sigma_{h,s} = \sigma(P_{h,s})$, and write $\zeta_{0,P} = \zeta_0(P)$ and $\sigma_P = \sigma(P)$. By definition, $\sigma_{h,s}$ satisfies equation (3.8) for P equal to $P_{h,s}$. Similar to (11.18), we decompose as follows

$$\begin{aligned} 0 &= (1-h) \Lambda_0(\zeta_{h,s}, \sigma_{h,s}) + h \left(\Psi_0(\mathbf{s}, \zeta_{h,s}, \sigma_{h,s}) - \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P) \right) \\ &\quad + h \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P), \end{aligned} \quad (11.23)$$

where Ψ_0 and Λ_0 are defined by (11.21) and (11.22), respectively. We first determine the order of $\sigma_{h,s} - \sigma_P$, as $h \downarrow 0$. Because ρ_0 satisfies (R4) and $\mathbf{V}$ satisfies (V4), it follows that $\Psi_0(\mathbf{s}, \zeta_{h,s}, \sigma_{h,s}) \rightarrow \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P)$, as $h \downarrow 0$. Since Λ_0 is continuous differentiable at $(\zeta_{0,P}, \sigma_P)$, similar to (11.19) we have that

$$\Lambda_0(\zeta_{h,s}, \sigma_{h,s}) = \Lambda_0(\zeta_{h,s}, \sigma_P) + D_{0,\sigma}(\sigma_{h,s} - \sigma_P) + o(\sigma_{h,s} - \sigma_P).$$

In general $\Lambda_0(\zeta_{h,s}, \sigma_P) \neq \mathbf{0}$, so that the behavior of $\sigma_{h,s} - \sigma_P$ depends on the behavior of

$$\begin{aligned} \Lambda_0(\zeta_{h,s}, \sigma_P) &= \Lambda_0(\zeta_{0,P}, \sigma_P) + \frac{\partial \Lambda_0(\zeta_{0,P}, \sigma_P)}{\partial \zeta}(\zeta_{h,s} - \zeta_{0,P}) + o(\|\zeta_{h,s} - \zeta_{0,P}\|) \\ &= \mathbf{D}_{0,\zeta}^T(\zeta_{h,s} - \zeta_{0,P}) + o(\|\zeta_{h,s} - \zeta_{0,P}\|). \end{aligned}$$

Here we also use that $\Lambda_0(\zeta_{0,P}, \sigma_P) = 0$, because σ_P is a solution of (3.8). Because $\text{IF}(\mathbf{s}; \zeta_0, P)$ exists, we conclude that $\Lambda_0(\zeta_{h,s}, \sigma_P) = O(h)$, and therefore

$$\Lambda_0(\zeta_{h,s}, \sigma_{h,s}) = D_{0,\sigma}(\sigma_{h,s} - \sigma_P) + o(\sigma_{h,s} - \sigma_P) + O(h).$$

When we insert this in the right hand side of (11.23), it follows that $\sigma_{h,s} - \sigma_P = O(h)$. Then, again from (11.23), we find that

$$0 = D_{0,\sigma}(\sigma_{h,s} - \sigma_P) + \mathbf{D}_{0,\zeta}^T(\zeta_{h,s} - \zeta_{0,P}) + h\Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P) + o(h).$$

After dividing by h , it follows that

$$\frac{\sigma_{h,s} - \sigma_P}{h} = -D_{0,\sigma}^{-1} \left\{ \mathbf{D}_{0,\zeta}^T \frac{\zeta_{h,s} - \zeta_{0,P}}{h} + \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P) \right\} + o(1).$$

When we let $h \downarrow 0$ and use that $\text{IF}(\mathbf{s}; \zeta_0, P)$ exists, we conclude that the limit of the left hand side exists and is given by

$$\text{IF}(\mathbf{s}; \sigma, P) = -D_{0,\sigma}^{-1} \left\{ \mathbf{D}_{0,\zeta}^T \text{IF}(\mathbf{s}; \zeta_0, P) + \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P) \right\}.$$

This proves the lemma. $\square$

Proof of Theorem 5

Proof. Denote $\xi_{h,s} = \xi(P_{h,s}) = (\beta_1(P_{h,s}), \gamma(P_{h,s}))$, $\sigma_{h,s} = \sigma(P_{h,s})$, and write $\xi_P = \xi(P)$ and $\sigma_P = \sigma(P)$. Then $(\xi_{h,s}, \sigma_{h,s})$ satisfies the regression score equation in (7.2) for P equal to $P_{h,s}$. Similar to (11.18) we decompose the regression score equation (7.2) as follows

$$\begin{aligned} \mathbf{0} &= \int \Psi_\beta(\mathbf{s}, \xi_{h,s}, \sigma_{h,s}) dP_{h,s}(\mathbf{s}) \\ &= (1-h)\Lambda_\beta(\xi_{h,s}, \sigma_{h,s}) + h \left(\Psi_\beta(\mathbf{s}, \xi_{h,s}, \sigma_{h,s}) - \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P) \right) \\ &\quad + h\Psi_\beta(\mathbf{s}, \xi_P, \sigma_P), \end{aligned} \tag{11.24}$$

where Ψ_β and Λ_β are defined by (7.3) and (8.2), respectively. Because ρ_1 and $\mathbf{V}$ satisfy (R4) and (V1), respectively, it follows that $\Psi_\beta(\mathbf{s}, \xi_{h,s}, \sigma_{h,s}) \rightarrow \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P)$, as $h \downarrow 0$. Because $\partial \Lambda_\beta / \partial \beta$ is continuous at (ξ_P, σ_P) and $(\gamma_{h,s}, \sigma_{h,s}) \rightarrow (\gamma_P, \sigma_P)$, we find that

$$\Lambda_\beta(\xi_{h,s}, \sigma_{h,s}) = \Lambda_\beta(\beta_1(P), \gamma_{h,s}, \sigma_{h,s}) + (D_\beta + o(1))(\beta_1(P_{h,s}) - \beta_1(P)).$$

Because $\beta_1(P)$ is a point of symmetry and Ψ_β is an odd function of $\mathbf{y} - \mathbf{X}\beta$, it follows that $\Lambda_\beta(\beta_1(P), \gamma_{h,s}, \sigma_{h,s}) = \mathbf{0}$. This means that

$$\Lambda_\beta(\xi_{h,s}, \sigma_{h,s}) = D_\beta(\beta_1(P_{h,s}) - \beta_1(P)) + o(\|\beta_1(P_{h,s}) - \beta_1(P)\|).$$

Together with (11.24), we find that

$$\mathbf{0} = \mathbf{D}_\beta(\beta_1(P_{h,\mathbf{s}}) - \beta_1(P)) + h\Psi_\beta(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) + o(\|\beta_1(P_{h,\mathbf{s}}) - \beta_1(P)\|) + o(h).$$

Since $\mathbf{D}_\beta$ is non-singular, this implies that $\beta_1(P_{h,\mathbf{s}}) - \beta_1(P) = O(h)$. When we insert this in the previous equation and divide by h , we obtain

$$\frac{\beta_1(P_{h,\mathbf{s}}) - \beta_1(P)}{h} = -\mathbf{D}_\beta^{-1}\Psi_\beta(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) + o(1).$$

When we let $h \downarrow 0$, this finishes the proof. $\square$

Lemma 4. *Let Λ_0 be defined by (11.22) with Ψ_0 defined in (11.21) and suppose that $\mathbb{E}\|\mathbf{X}\| < \infty$. Suppose that ρ_0 and $\mathbf{V}$ satisfy (R2), (R4) and (V4), respectively. Let $\zeta_0(P) = (\beta_0(P), \theta_0(P))$ be the pair of initial functionals and let $\sigma(P)$ be a solution of (3.8). Let $N \subset \mathbb{R}^q \times \boldsymbol{\Theta} \times (0, \infty)$ be an open neighborhood of $(\zeta_0(P), \sigma(P))$. Then Λ_0 is continuous differentiable at $(\zeta_0(P), \sigma(P))$ and for all $(\zeta, \sigma) \in N$,*

$$\frac{\partial \Lambda_0(\zeta, \sigma)}{\partial \zeta} = \int \frac{\partial \Psi_0(\mathbf{s}, \zeta, \sigma)}{\partial \zeta} dP(\mathbf{s}) \quad \text{and} \quad \frac{\partial \Lambda_0(\zeta, \sigma)}{\partial \sigma} = \int \frac{\partial \Psi_0(\mathbf{s}, \zeta, \sigma)}{\partial \sigma} dP(\mathbf{s}).$$

Proof. Let $(\zeta, \sigma) \in N$. Consider $\zeta \mapsto \Lambda_0(\zeta, \sigma)$ with $\sigma \in (0, \infty)$ fixed. From (11.21) we find

$$\frac{\partial \Psi_0(\mathbf{s}, \zeta, \sigma)}{\partial \beta} = \rho'_0 \left(\frac{d_\Gamma}{\sigma} \right) \frac{1}{2d_\Gamma \sigma} \mathbf{X}^T \boldsymbol{\Gamma}^{-1} (\mathbf{y} - \mathbf{X}\beta) = \frac{1}{2\sigma^2} u_0 \left(\frac{d_\Gamma}{\sigma} \right) \mathbf{X}^T \boldsymbol{\Gamma}^{-1} (\mathbf{y} - \mathbf{X}\beta)$$

where $u_0(s) = \rho_0(s)/s$ and $d_\Gamma^2 = (\mathbf{y} - \mathbf{X}\beta)^T \boldsymbol{\Gamma}^{-1} (\mathbf{y} - \mathbf{X}\beta)$, and where we write $\boldsymbol{\Gamma}$ for $\boldsymbol{\Gamma}(\theta)$, as defined in (3.1). Similar to the proof of Lemma 11.2 in Lopuhaä *et al* [22] we obtain

$$\|\mathbf{X}^T \boldsymbol{\Gamma}^{-1} (\mathbf{y} - \mathbf{X}\beta)\|^2 \leq d_\Gamma^2 \|\mathbf{X}\|^2 \lambda_1(\boldsymbol{\Gamma}^{-1}).$$

This means that

$$\left\| \frac{\partial \Psi_0(\mathbf{s}, \zeta, \sigma)}{\partial \beta} \right\| \leq \frac{1}{2\sigma} \left| u_0 \left(\frac{d_\Gamma}{\sigma} \right) \frac{d_\Gamma}{\sigma} \right| \|\mathbf{X}\| \sqrt{\lambda_1(\boldsymbol{\Gamma}^{-1})}.$$

From (R2) and (R4) it follows that $u_0(s)s = \rho'_0(s)$ is bounded and for (ζ, σ) in the neighborhood N of $(\zeta_0(P), \sigma(P))$, we have that $1/\sigma$ and $\lambda_1(\boldsymbol{\Gamma}(\theta)^{-1})$ are uniformly bounded. This means there exists a universal constant $0 < C_1 < \infty$, such that

$$\left\| \frac{\partial \Psi_0(\mathbf{s}, \zeta, \sigma)}{\partial \beta} \right\| \leq C_1 \|\mathbf{X}\|.$$

Since $\mathbb{E}\|\mathbf{X}\| < \infty$, by dominated convergence, it follows that for (ζ, σ) in the neighborhood N of $(\zeta_0(P), \sigma(P))$, it holds that

$$\frac{\partial \Lambda_0(\zeta, \sigma)}{\partial \beta} = \int \frac{\partial \Psi_0(\mathbf{s}, \zeta, \sigma)}{\partial \beta} dP(\mathbf{s}),$$

and that $\partial \Lambda_0 / \partial \beta$ is continuous at $(\zeta_0(P), \sigma(P))$. Furthermore, from (11.21) we find

$$\frac{\partial \Psi_0(\mathbf{s}, \zeta, \sigma)}{\partial \theta_j} = \frac{1}{2\sigma} \rho'_0 \left(\frac{d_\Gamma}{\sigma} \right) (\mathbf{y} - \mathbf{X}\beta)^T \boldsymbol{\Gamma}^{-1} \frac{\partial \boldsymbol{\Gamma}}{\partial \theta_j} \boldsymbol{\Gamma}^{-1} (\mathbf{y} - \mathbf{X}\beta).$$

for any $j = 1, \dots, l$. Similar to the proof of Lemma 11.2 in Lopuhaä *et al* [22], we find

$$\left| (\mathbf{y} - \mathbf{X}\beta)^T \boldsymbol{\Gamma}^{-1} \frac{\partial \boldsymbol{\Gamma}}{\partial \theta_j} \boldsymbol{\Gamma}^{-1} (\mathbf{y} - \mathbf{X}\beta) \right| \leq d_\Gamma^2 \left\| \frac{\partial \boldsymbol{\Gamma}}{\partial \theta_j} \right\| \lambda_1(\boldsymbol{\Gamma}^{-1}).$$

Furthermore, according to (V4), the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{\Gamma}(\boldsymbol{\theta}) = \mathbf{V}(\boldsymbol{\theta})/|\mathbf{V}(\boldsymbol{\theta})|^{1/k}$ is continuously differentiable. This means that there exists a universal constant $0 < M_1 < \infty$, such that

$$\max_{1 \leq j \leq l} \sup_{(\boldsymbol{\beta}, \boldsymbol{\theta}) \in N} \left\| \frac{\partial \boldsymbol{\Gamma}(\boldsymbol{\theta})}{\partial \theta_j} \right\| \leq M_1. \quad (11.25)$$

We find that

$$\left\| \frac{\partial \Psi_0(\mathbf{s}, \boldsymbol{\zeta}, \sigma)}{\partial \theta_j} \right\| \leq \frac{\sigma M_1}{2} \left| \rho'_0 \left(\frac{d_\Gamma}{\sigma} \right) \right| \frac{d_\Gamma^2}{\sigma^2} \lambda_1(\boldsymbol{\Gamma}^{-1}).$$

From (R2) and (R4), it follows that $\rho'_0(s)s^2$ is bounded and for $(\boldsymbol{\zeta}, \sigma)$ in the neighborhood N of $(\boldsymbol{\zeta}_0(P), \sigma(P))$, we have that σ and $\lambda_1(\boldsymbol{\Gamma}(\boldsymbol{\theta})^{-1})$ are uniformly bounded. This means there exists a universal constant $0 < C_2 < \infty$, such that

$$\left\| \frac{\partial \Psi_0(\mathbf{s}, \boldsymbol{\zeta}, \sigma)}{\partial \theta_j} \right\| \leq C_2,$$

for all $j = 1, \dots, l$. By dominated convergence, it follows that for $(\boldsymbol{\zeta}, \sigma)$ in the neighborhood N of $(\boldsymbol{\zeta}_0(P), \sigma(P))$, it holds that

$$\frac{\partial \Lambda_0(\boldsymbol{\zeta}, \sigma)}{\partial \boldsymbol{\theta}} = \int \frac{\partial \Psi_0(\mathbf{s}, \boldsymbol{\zeta}, \sigma)}{\partial \boldsymbol{\theta}} dP(\mathbf{s}),$$

and that $\partial \Lambda_0 / \partial \boldsymbol{\theta}$ is continuous at $(\boldsymbol{\zeta}_0(P), \sigma(P))$. Finally, from (11.21) we obtain

$$\frac{\partial \Psi_0(\mathbf{s}, \boldsymbol{\zeta}, \sigma)}{\partial \sigma} = -\frac{1}{\sigma} \rho'_0 \left(\frac{d_\Gamma}{\sigma} \right) \left(\frac{d_\Gamma}{\sigma} \right).$$

From (R2) and (R4), it follows that $\rho'_0(s)s$ is bounded, and for $(\boldsymbol{\zeta}, \sigma)$ in the neighborhood N of $(\boldsymbol{\zeta}_0(P), \sigma(P))$, we have that $1/\sigma$ is uniformly bounded. This means there exists a universal constant $0 < C_3 < \infty$, such that

$$\left\| \frac{\partial \Psi_0(\mathbf{s}, \boldsymbol{\zeta}, \sigma)}{\partial \sigma} \right\| \leq C_3.$$

By dominated convergence, it follows that for $(\boldsymbol{\zeta}, \sigma)$ in the neighborhood N of $(\boldsymbol{\zeta}_0(P), \sigma(P))$, it holds that

$$\frac{\partial \Lambda_0(\boldsymbol{\zeta}, \sigma)}{\partial \sigma} = \int \frac{\partial \Psi_0(\mathbf{s}, \boldsymbol{\zeta}, \sigma)}{\partial \sigma} dP(\mathbf{s}),$$

and that $\partial \Lambda_0 / \partial \sigma$ is continuous at $(\boldsymbol{\zeta}_0(P), \sigma(P))$. $\square$

For convenience we state the following result about spherically contoured densities, e.g., see Lemma 5.1 in [17]. This lemma uses the commutation matrix $\mathbf{K}_{k,k}$, which is the $k^2 \times k^2$ block matrix with the (i, j) -block being equal to the $k \times k$ matrix $\boldsymbol{\Delta}_{ji}$ consisting of zero's except a 1 at entry (j, i) . A useful property (e.g., see [24, Section 3.7]) is that for any $k \times k$ matrix $\mathbf{A}$, it holds that

$$\mathbf{K}_{k,k} \text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}^T). \quad (11.26)$$

Lemma 5. *Suppose that $\mathbf{z}$ has a k -variate elliptical contoured density defined in (8.4), with parameters $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_k$. Then $\mathbf{u} = \mathbf{z}/\|\mathbf{z}\|$ is independent of $\|\mathbf{z}\|$, has mean zero and covariance matrix $(1/k)\mathbf{I}_k$. Furthermore, $\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\mathbf{u}\mathbf{u}^T \mathbf{u}] = \mathbf{0}$ and*

$$\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\text{vec}(\mathbf{u}\mathbf{u}^T) \text{vec}(\mathbf{u}\mathbf{u}^T)^T] = \sigma_1(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}) + \sigma_2 \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T,$$

where $\sigma_1 = \sigma_2 = (k(k+2))^{-1}$.

Proof. See e.g. the proof of Lemma 5.1 in Lopushaä [17]. $\square$

Lemma 6. Suppose that P satisfies (E) for some $\zeta^* = (\beta^*, \theta^*) \in \mathbb{R}^q \times \Theta$ and suppose that $\mathbb{E}\|\mathbf{X}\| < \infty$. Suppose that ρ_0 and $\mathbf{V}$ satisfy (R2), (R4) and (V4), respectively. Let $\zeta_0 = (\beta_0, \theta_0)$ be the pair of initial functionals satisfying $(\beta_0(P), \theta_0(P)) = (\beta^*, \theta^*)$. Let $\sigma(P)$ be the unique solution of (3.8) and let Λ_0 be defined in (11.22) with Ψ_0 from (11.21). Then,

$$\mathbf{D}_{0,\zeta} = \frac{\partial \Lambda_0(\zeta_0(P), \sigma(P))}{\partial \zeta} = \mathbf{0},$$

and

$$D_{0,\sigma} = \frac{\partial \Lambda_0(\zeta_0(P), \sigma(P))}{\partial \sigma} = -\frac{1}{\sigma(P)} \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|],$$

where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma(P)$.

Proof. Write $\zeta_{0,P} = (\beta_{0,P}, \theta_{0,P}) = (\beta_0(P), \theta_0(P))$ and $\sigma_P = \sigma(P)$. Because ρ_0 and $\mathbf{V}$ satisfy (R2), (R4) and (V4), respectively, and $\mathbb{E}\|\mathbf{X}\| < \infty$, according to Lemma 4, we have that Λ_0 is continuously differentiable at $(\zeta_{0,P}, \sigma_P)$ and that we may interchange integration and differentiation in $\partial \Lambda_0 / \partial \beta$ at $(\zeta_{0,P}, \sigma_P)$. With $u_0(s) = \rho'_0(s)/s$, we find that

$$\mathbf{D}_{0,\beta} = \frac{\partial \Lambda_0(\zeta_{0,P}, \sigma_P)}{\partial \beta} = -\frac{1}{2\sigma_P} \mathbb{E} \left[u_0 \left(\frac{d_{\Gamma,0}}{\sigma_P} \right) \mathbf{X}^T \Gamma(\theta_{0,P})^{-1} (\mathbf{y} - \mathbf{X}\beta_{0,P}) \right],$$

where $d_{\Gamma,0} = d(\mathbf{y}, \mathbf{X}\beta_{0,P}, \Gamma(\theta_{0,P}))$, as defined in (7.1), with Γ defined in (3.1). Since $\beta_{0,P} = \beta^*$ is a point of symmetry of P , it follows that

$$\mathbf{D}_{0,\beta} = \mathbf{0}. \quad (11.27)$$

According to Lemma 4, we may also interchange integration and differentiation in $\partial \Lambda_0 / \partial \theta$ at $(\zeta_{0,P}, \sigma_P)$. For any $j = 1, \dots, l$, we find that

$$\begin{aligned} D_{0,j} &= \frac{\partial \Lambda_0(\zeta_{0,P}, \sigma_P)}{\partial \theta_j} \\ &= -\frac{1}{2\sigma_P} \mathbb{E} \left[u_0 \left(\frac{d_{\Gamma,0}}{\sigma_P} \right) \mathbf{e}_{0,P}^T \Gamma(\theta_{0,P})^{-1} \frac{\partial \Gamma(\theta_{0,P})}{\partial \theta_j} \Gamma(\theta_{0,P})^{-1} \mathbf{e}_{0,P} \right], \end{aligned} \quad (11.28)$$

where $\mathbf{e}_{0,P} = \mathbf{y} - \mathbf{X}\beta_{0,P}$. Since $(\beta_{0,P}, \theta_{0,P}) = (\beta^*, \theta^*)$ and $\Gamma(\theta_{0,P}) = \mathbf{V}(\theta^*)/|\mathbf{V}(\theta^*)|^{1/k} = \Sigma/|\Sigma|^{1/k}$, it follows that $d_{\Gamma,0} = d_\Gamma^*$, where

$$(d_\Gamma^*)^2 = |\Sigma|^{1/k} (\mathbf{y} - \mathbf{X}\beta^*)^T \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta^*). \quad (11.29)$$

Hence, the expectation on the right hand side of (11.28) can be written as

$$|\Sigma|^{2/k} \mathbb{E} \left[\mathbb{E} \left[u_0 \left(\frac{d_\Gamma^*}{\sigma_P} \right) (\mathbf{e}^*)^T \Sigma^{-1} \frac{\partial \Gamma(\theta^*)}{\partial \theta_j} \Sigma^{-1} \mathbf{e}^* \middle| \mathbf{X} \right] \right]$$

where $\mathbf{e}^* = \mathbf{y} - \mathbf{X}\beta^*$. The inner expectation is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$, which has the same distribution as $\Sigma^{1/2} \mathbf{z} + \mathbf{X}\beta^*$, where $\mathbf{z}$ has spherical density $f_{\mathbf{0}, \mathbf{I}_k}$. This means that the inner expectation can be written as

$$\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[u_0(c_\sigma \|\mathbf{z}\|) \mathbf{z}^T \Sigma^{-1/2} \frac{\partial \Gamma(\theta^*)}{\partial \theta_j} \Sigma^{-1/2} \mathbf{z} \right],$$

where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma_P$. Next, let $\mathbf{u} = \mathbf{z}/\|\mathbf{z}\|$ and apply Lemma 5. It follows that this expectation is equal to

$$\begin{aligned} &\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_0(c_\sigma \|\mathbf{z}\|) \|\mathbf{z}\|^2] \text{tr} \left(\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\mathbf{u}\mathbf{u}^T] \Sigma^{-1/2} \frac{\partial \Gamma(\theta^*)}{\partial \theta_j} \Sigma^{-1/2} \right) \\ &= \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_0(c_\sigma \|\mathbf{z}\|) \|\mathbf{z}\|^2] \frac{1}{k} \text{tr} \left(\Sigma^{-1/2} \frac{\partial \Gamma(\theta^*)}{\partial \theta_j} \Sigma^{-1/2} \right). \end{aligned}$$

Because for each $j = 1, \dots, l$,

$$\frac{\partial \Gamma}{\partial \theta_j} = \frac{\partial \mathbf{V}/|\mathbf{V}|^{1/k}}{\partial \theta_j} = -\frac{1}{k}|\mathbf{V}|^{-1/k} \text{tr} \left(\Sigma^{-1} \frac{\partial \mathbf{V}}{\partial \theta_j} \right) \mathbf{V} + |\mathbf{V}|^{-1/k} \frac{\partial \mathbf{V}}{\partial \theta_j},$$

together with $\mathbf{V}(\boldsymbol{\theta}^*) = \Sigma$, it follows that

$$\text{tr} \left(\Sigma^{-1/2} \frac{\partial \Gamma(\boldsymbol{\theta}^*)}{\partial \theta_j} \Sigma^{-1/2} \right) = 0,$$

for all $j = 1, 2, \dots, l$. This means that

$$\mathbf{D}_{0,\boldsymbol{\theta}} = \frac{\partial \Lambda_0(\zeta_{0,P}, \sigma_P)}{\partial \boldsymbol{\theta}} = \mathbf{0}. \quad (11.30)$$

Together with (11.27) this proves part one.

According to Lemma 4, we may also interchange integration and differentiation in $\partial \Lambda_0 / \partial \sigma$. We find that

$$D_{0,\sigma} = \frac{\partial \Lambda_0(\zeta_{0,P}, \sigma_P)}{\partial \sigma} = -\frac{1}{\sigma_P} \mathbb{E} \left[\rho'_0 \left(\frac{d_{\Gamma,0}}{\sigma_P} \right) \frac{d_{\Gamma,0}}{\sigma_P} \right],$$

where $d_{\Gamma,0} = d(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}_{0,P}, \Gamma(\boldsymbol{\theta}_{0,P}))$, as defined in (7.1), with Γ defined in (3.1). As before, it follows that $d_{\Gamma,0} = d_{\Gamma}^*$, where d_{Γ}^* is defined in (11.29), so that

$$D_{0,\sigma} = -\frac{1}{\sigma_P} \mathbb{E} \left[\rho'_0 \left(\frac{d_{\Gamma}^*}{\sigma_P} \right) \frac{d_{\Gamma}^*}{\sigma_P} \right] = -\frac{1}{\sigma_P} \mathbb{E} \left[\mathbb{E} \left[\rho'_0 \left(\frac{d_{\Gamma}^*}{\sigma_P} \right) \frac{d_{\Gamma}^*}{\sigma_P} \middle| \mathbf{X} \right] \right].$$

Then, the inner expectation on the right hand side is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$, which has the same distribution as $\Sigma^{1/2} \mathbf{z} + \mathbf{X}\boldsymbol{\beta}^*$, where $\mathbf{z}$ has spherical density $f_{\mathbf{0}, \mathbf{I}_k}$. This means that

$$D_{0,\sigma} = -\frac{1}{\sigma_P} \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|] < 0, \quad (11.31)$$

where $c_\sigma = |\Sigma|^{1/(2k)} / \sigma_P$. □

Lemma 7. Suppose that P satisfies (E) for some $(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*) \in \mathbb{R}^q \times \boldsymbol{\Theta}$ and suppose that $\mathbb{E}\|\mathbf{X}\| < \infty$. Suppose that ρ_0 satisfies (R2) and (R4), and that $\mathbf{V}$ satisfies (V4). Let $\zeta_0 = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ be the pair of initial functionals satisfying $(\boldsymbol{\beta}_0(P), \boldsymbol{\theta}_0(P)) = (\boldsymbol{\beta}^*, \boldsymbol{\theta}^*)$, and suppose that $\text{IF}(\mathbf{s}, \zeta_0, P)$ exists. Let $\sigma(P_{h,\mathbf{s}})$ be the solution of equation (3.8) with $P = P_{h,\mathbf{s}}$, and suppose that for all $\mathbf{s} \in \mathbb{R}^k \times \mathbb{R}^{kq}$, $\sigma(P_{h,\mathbf{s}}) \rightarrow \sigma(P)$, as $h \downarrow 0$, where $\sigma(P)$ is a solution of (3.8). Suppose that $\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|] > 0$, where $c_\sigma = |\Sigma|^{1/(2k)} / \sigma(P)$. Then, for $\mathbf{s}_0 \in \mathbb{R}^k \times \mathbb{R}^{kq}$,

$$\text{IF}(\mathbf{s}_0; \sigma, P) = \frac{\sigma(P)}{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|]} \left\{ \rho_0(c_\sigma \|\mathbf{z}_0\|) - b_0 \right\},$$

where $\mathbf{z}_0 = \Sigma^{-1/2}(\mathbf{y}_0 - \mathbf{X}_0 \boldsymbol{\beta}^*)$.

Proof. From Lemmas 4 and 6, we have that Λ_0 is continuously differentiable at $(\zeta_0(P), \sigma(P))$ with $\mathbf{D}_{0,\zeta} = \mathbf{0}$ and $D_{0,\sigma} = -\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|] / \sigma(P) < 0$. Since $(\boldsymbol{\beta}_0(P), \boldsymbol{\theta}_0(P)) = (\boldsymbol{\beta}^*, \boldsymbol{\theta}^*)$ and $\Gamma(\boldsymbol{\theta}_0(P)) = \mathbf{V}(\boldsymbol{\theta}^*) / |\mathbf{V}(\boldsymbol{\theta}^*)|^{1/k} = \Sigma / |\Sigma|^{1/k}$, it follows that

$$\Psi_0(\mathbf{s}_0, \zeta_0(P), \sigma(P)) = \rho_0 \left(\frac{d(\mathbf{y}_0, \mathbf{X}_0 \boldsymbol{\beta}^*, \Sigma / |\Sigma|^{1/k})}{\sigma(P)} \right) - b_0 = \rho_0(c_\sigma \|\mathbf{z}_0\|) - b_0, \quad (11.32)$$

where $\mathbf{z}_0 = \Sigma^{-1/2}(\mathbf{y}_0 - \mathbf{X}_0 \boldsymbol{\beta}^*)$. The lemma now follows immediately from Lemma 3. □

Lemma 8. Suppose that ρ_1 satisfies (R2) and (R4). Let $\sigma(P)$ be a solution of (3.8) and let $\Psi = (\Psi_\beta, \Psi_\gamma)$, as defined in (7.3). Then there exist $0 < C_1 < \infty$, only depending on P and $\sigma(P)$, such that $\|\Psi_\beta(\mathbf{s}, \boldsymbol{\xi}(P), \sigma(P))\| \leq C_1 \|\mathbf{X}\|$. If in addition, $\mathbf{V}$ satisfies (V4), then there exist $0 < C_2 < \infty$, only depending on P and $\sigma(P)$, such that $\|\Psi_\gamma(\mathbf{s}, \boldsymbol{\xi}(P), \sigma(P))\| \leq C_2$.

Proof. The proof is completely similar to that of Lemma 11.2 in Lopuhaä *et al* [22]. $\square$

Lemma 9. Let Λ be defined by (8.2) with Ψ defined in (7.3) and let $\mathbb{E}\|\mathbf{X}\|^2 < \infty$. Suppose that ρ_1 satisfies (R2) and (R5) and $\mathbf{V}$ satisfies (V5). Let $\sigma(P)$ be a solution of (3.8) and let $\xi(P) \in \mathfrak{D}$ be a local minimum of $R_P(\beta, \mathbf{V}(\gamma))$. Let $N \subset \mathbb{R}^k \times \Theta \times (0, \infty)$ be an open neighborhood of $(\xi(P), \sigma(P))$. Then Λ is continuous differentiable at $(\xi(P), \sigma(P))$ and for all $(\xi, \sigma) \in N$,

$$\frac{\partial \Lambda(\xi, \sigma)}{\partial \xi} = \int \frac{\partial \Psi(\mathbf{s}, \xi, \sigma)}{\partial \xi} dP(\mathbf{s}) \quad \text{and} \quad \frac{\partial \Lambda(\xi, \sigma)}{\partial \sigma} = \int \frac{\partial \Psi(\mathbf{s}, \xi, \sigma)}{\partial \sigma} dP(\mathbf{s}).$$

Proof. Let $(\xi, \sigma) \in N$. Consider $\xi \mapsto \Lambda(\xi, \sigma)$ with $\sigma \in (0, \infty)$ fixed. The proof of

$$\frac{\partial \Lambda(\xi, \sigma)}{\partial \xi} = \int \frac{\partial \Psi(\mathbf{s}, \xi, \sigma)}{\partial \xi} dP(\mathbf{s}),$$

and that $\partial \Lambda / \partial \xi$ is continuous at $(\xi(P), \sigma(P))$, is completely similar to that of Lemma 11.3 in Lopuhaä *et al* [22], taking into account that σ is uniformly bounded away from zero and infinity.

Next consider $\sigma \mapsto \Lambda(\xi, \sigma)$ with $\xi \in \mathbb{R}^k \times \Theta$ fixed. From (7.3), we find that

$$\begin{aligned} \frac{\partial \Psi_{\beta}(\mathbf{s}, \xi, \sigma)}{\partial \sigma} &= -u'_1 \left(\frac{d}{\sigma} \right) \frac{d}{\sigma^2} \mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) \\ \frac{\partial \Psi_{\gamma, j}(\mathbf{s}, \xi, \sigma)}{\partial \sigma} &= -u'_1 \left(\frac{d}{\sigma} \right) \frac{d}{\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \mathbf{H}_{1, j} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta), \end{aligned}$$

for $j = 1, \dots, l$, where $d^2 = (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$ and $\mathbf{H}_j$ is defined in (7.4). Taking into account that σ is uniformly bounded away from zero and infinity, similar to the proof of Lemma 11.3 in Lopuhaä *et al* [22], we obtain

$$\left\| \frac{\partial \Psi_{\beta}(\mathbf{s}, \xi, \sigma)}{\partial \sigma} \right\| \leq C_1 \|\mathbf{X}\|^2 \quad \text{and} \quad \left\| \frac{\partial \Psi_{\gamma, j}(\mathbf{s}, \xi, \sigma)}{\partial \sigma} \right\| \leq C_2,$$

for constants $0 < C_1, C_2 < \infty$ only depending on P . Hence, it follows by dominated convergence that for (ξ, σ) in the neighborhood N of $(\xi(P), \sigma(P))$, it holds that

$$\frac{\partial \Lambda(\xi, \sigma)}{\partial \sigma} = \int \frac{\partial \Psi(\mathbf{s}, \xi, \sigma)}{\partial \sigma} dP(\mathbf{s}),$$

and that $\partial \Lambda / \partial \sigma$ is continuous at $(\xi(P), \sigma(P))$. $\square$

Lemma 10. Suppose that P satisfies (E) for some $(\beta^*, \theta^*) \in \mathbb{R}^q \times \Theta$ and that $\mathbb{E}\|\mathbf{X}\|^2 < \infty$. Suppose that ρ_1 satisfies (R2) and (R5) and that $\mathbf{V}$ satisfies (V5) and has a linear structure (7.5). Let $\sigma(P)$ be the solution of (3.8) and let $\xi(P) = (\beta_1(P), \gamma(P)) \in \mathfrak{D}$ be a local minimum of $R_P(\beta, \mathbf{V}(\gamma))$. Suppose that $\beta_1(P) = \beta^*$ and $\mathbf{V}(\gamma(P)) = \Sigma / |\Sigma|^{1/k}$. Let Λ be defined by (8.2) and (8.3) with Ψ from (7.7). Then

$$\mathbf{D}_{\xi} = \frac{\partial \Lambda(\xi(P), \sigma(P))}{\partial \xi} = \begin{pmatrix} \mathbf{D}_{\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\gamma} \end{pmatrix},$$

where

$$\mathbf{D}_{\beta} = \frac{\partial \Lambda_{\beta}(\xi(P), \sigma(P))}{\partial \beta} = -\alpha_1 |\Sigma|^{1/k} \mathbb{E} [\mathbf{X}^T \Sigma^{-1} \mathbf{X}], \quad (11.33)$$

with α_1 defined in (8.5), and

$$\mathbf{D}_{\gamma} = \frac{\partial \Lambda_{\gamma}(\xi(P), \sigma(P))}{\partial \gamma} = \omega_1 \mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} - \omega_2 \mathbf{L}^T \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma^{-1})^T \mathbf{L}, \quad (11.34)$$

where $\mathbf{L} = \partial \text{vec}(\mathbf{V}(\gamma(P))) / \partial \gamma^T$ is the $k^2 \times l$ matrix given in (7.6), $\omega_1 = \sigma^2(P) |\Sigma|^{2/k} \gamma_1$ and $\omega_2 = \omega_1 / k + |\Sigma|^{2/k}$, with γ_1 defined in (8.5).

Proof. For convenience, write $\xi_P = (\beta_{1,P}, \gamma_P) = (\beta_1(P), \gamma(P))$, $\mathbf{V}_P = \mathbf{V}(\gamma(P))$, and $\sigma_P = \sigma(P)$. Write $\partial\Lambda/\partial\xi$ as the block matrix

$$\frac{\partial\Lambda(\xi_P, \sigma_P)}{\partial\xi} = \begin{pmatrix} \frac{\partial\Lambda_\beta(\xi_P, \sigma_P)}{\partial\beta} & \frac{\partial\Lambda_\beta(\xi_P, \sigma_P)}{\partial\gamma} \\ \frac{\partial\Lambda_\gamma(\xi_P, \sigma_P)}{\partial\beta} & \frac{\partial\Lambda_\gamma(\xi_P, \sigma_P)}{\partial\gamma} \end{pmatrix}, \quad (11.35)$$

where Λ_β and Λ_γ are defined in (8.3) with Ψ_β and Ψ_γ from (7.7). Because ρ_1 and $\mathbf{V}$ satisfy (R2), (R5), and (V5), and $\mathbb{E}\|\mathbf{X}\|^2 < \infty$, according to Lemma 9 we may interchange integration and differentiation in $\partial\Lambda_\beta/\partial\gamma$ and $\partial\Lambda_\gamma/\partial\beta$. It can be seen that these are expectations of an odd function of $\mathbf{y} - \mathbf{X}\beta_{1,P}$, which means that they are equal to zero, as $\beta_{1,P} = \beta^*$ is a point of symmetry of P . Therefore

$$\mathbf{D}_\xi = \frac{\partial\Lambda(\xi_P, \sigma_P)}{\partial\xi} = \begin{pmatrix} \mathbf{D}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\gamma \end{pmatrix}. \quad (11.36)$$

It remains to determine $\mathbf{D}_\beta = \partial\Lambda_\beta(\xi_P, \sigma_P)/\partial\beta$ and $\mathbf{D}_\gamma = \partial\Lambda_\gamma(\xi_P, \sigma_P)/\partial\gamma$. According to Lemma 9, we have that Λ is continuous differentiable at (ξ_P, σ_P) and that we may interchange integration and differentiation in $\partial\Lambda_\beta/\partial\beta$, where Λ_β is defined in (8.3) with Ψ_β from (7.7). We obtain

$$\begin{aligned} \mathbf{D}_\beta &= \int \frac{\partial}{\partial\beta} \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P) dP(\mathbf{s}) \\ &= -\mathbb{E} \left[u'_1 \left(\frac{d_P}{\sigma_P} \right) \frac{\mathbf{X}^T \mathbf{V}_P^{-1} \mathbf{e}_P \mathbf{e}_P^T \mathbf{V}_P^{-1} \mathbf{X}}{\sigma_P d_P} + u_1 \left(\frac{d_P}{\sigma_P} \right) \mathbf{X}^T \mathbf{V}_P^{-1} \mathbf{X} \right] \\ &= -\mathbb{E} \left[\mathbb{E} \left[u'_1 \left(\frac{d_P}{\sigma_P} \right) \frac{\mathbf{X}^T \mathbf{V}_P^{-1} \mathbf{e}_P \mathbf{e}_P^T \mathbf{V}_P^{-1} \mathbf{X}}{\sigma_P d_P} + u_1 \left(\frac{d_P}{\sigma_P} \right) \mathbf{X}^T \mathbf{V}_P^{-1} \mathbf{X} \middle| \mathbf{X} \right] \right], \end{aligned} \quad (11.37)$$

where $d_P^2 = \mathbf{e}_P^T \mathbf{V}_P^{-1} \mathbf{e}_P$ and $\mathbf{e}_P = \mathbf{y} - \mathbf{X}\beta_{1,P}$. The inner expectation on the right hand side is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$, which can be written as

$$\mathbf{X}^T \mathbf{V}_P^{-1/2} \mathbb{E} \left[u'_1 \left(\frac{d_P}{\sigma_P} \right) \frac{\mathbf{V}_P^{-1/2} \mathbf{e}_P \mathbf{e}_P^T \mathbf{V}_P^{-1/2}}{\sigma_P d_P} + u_1 \left(\frac{d_P}{\sigma_P} \right) \mathbf{I}_k \middle| \mathbf{X} \right] \mathbf{V}_P^{-1/2} \mathbf{X}.$$

Because $\beta_{1,P} = \beta^*$ and $\mathbf{V}_P = \Sigma/|\Sigma|^{1/k}$, the previous expression is equal to

$$|\Sigma|^{1/k} \mathbf{X}^T \Sigma^{-1/2} \mathbb{E} \left[u'_1 \left(\frac{d^*}{\sigma_P} \right) \frac{|\Sigma|^{1/k} \Sigma^{-1/2} \mathbf{e}^* (\mathbf{e}^*)^T \Sigma^{-1/2}}{\sigma_P d^*} + u_1 \left(\frac{d^*}{\sigma_P} \right) \mathbf{I}_k \middle| \mathbf{X} \right] \Sigma^{-1/2} \mathbf{X},$$

where $(d^*)^2 = |\Sigma|^{1/k} (\mathbf{y} - \mathbf{X}\beta^*)^T \Sigma^{-1} (\mathbf{y} - \mathbf{X}\beta^*)$ and $\mathbf{e}^* = \mathbf{y} - \mathbf{X}\beta^*$. Note that $\mathbf{y} \mid \mathbf{X}$ has the same distribution as $\Sigma^{1/2} \mathbf{z} + \mathbf{X}\beta^*$, where $\mathbf{z}$ has a spherical density $f_{\mathbf{0}, \mathbf{I}_k}$, so that the expression in the previous display is equal to

$$|\Sigma|^{1/k} \mathbf{X}^T \Sigma^{-1/2} \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[u'_1(c_\sigma \|\mathbf{z}\|) \frac{c_\sigma}{\|\mathbf{z}\|} \mathbf{z} \mathbf{z}^T + u_1(c_\sigma \|\mathbf{z}\|) \mathbf{I}_k \right] \Sigma^{-1/2} \mathbf{X},$$

where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma_P$. Let $\mathbf{u} = \mathbf{z}/\|\mathbf{z}\|$. Then with Lemma 5 we find

$$\begin{aligned} &\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[u'_1(c_\sigma \|\mathbf{z}\|) \frac{c_\sigma}{\|\mathbf{z}\|} \mathbf{z} \mathbf{z}^T + u_1(c_\sigma \|\mathbf{z}\|) \mathbf{I}_k \right] \\ &= \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u'_1(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|] \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\mathbf{u} \mathbf{u}^T] + \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|)] \mathbf{I}_k = \alpha_1 \mathbf{I}_k, \end{aligned}$$

where

$$\begin{aligned}\alpha_1 &= \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\frac{1}{k} u'_1(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\| + u_1(c_\sigma \|\mathbf{z}\|) \right] \\ &= \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\left(1 - \frac{1}{k} \right) \frac{\rho'_1(c_\sigma \|\mathbf{z}\|)}{c_\sigma \|\mathbf{z}\|} + \frac{1}{k} \rho''_1(c_\sigma \|\mathbf{z}\|) \right].\end{aligned}$$

We conclude that

$$\frac{\partial \Lambda_{\boldsymbol{\beta}}(\boldsymbol{\xi}_P, \sigma_P)}{\partial \boldsymbol{\beta}} = -\alpha_1 |\boldsymbol{\Sigma}|^{1/k} \mathbb{E} [\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}].$$

Next, we determine $\partial \Lambda_{\boldsymbol{\gamma}}(\boldsymbol{\xi}_P, \sigma_P) / \partial \boldsymbol{\gamma}$. From (7.7) we have

$$\Psi_{\boldsymbol{\gamma}, j} = -\text{vec}(\mathbf{V}^{-1} \mathbf{L}_j \mathbf{V}^{-1})^T \text{vec}(\Psi_{\mathbf{V}}),$$

for all $j = 1, 2, \dots, l$, where $\Psi_{\mathbf{V}}$ is defined in (7.8). Because $|\mathbf{V}_P| = 1$, we have

$$\begin{aligned}\int \Psi_{\mathbf{V}}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) dP(\mathbf{s}) &= \mathbb{E} \left[k u_1 \left(\frac{d_P}{\sigma_P} \right) \mathbf{e}_P \mathbf{e}_P^T - v_1 \left(\frac{d_P}{\sigma_P} \right) \sigma_P^2 \mathbf{V}_P \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[k u_1 \left(\frac{d_P}{\sigma_P} \right) \mathbf{e}_P \mathbf{e}_P^T - v_1 \left(\frac{d_P}{\sigma_P} \right) \sigma_P^2 \mathbf{V}_P \middle| \mathbf{X} \right] \right].\end{aligned}$$

Similar to the reasoning before, the inner expectation can be written as

$$\begin{aligned}\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} &\left[k u_1(c_\sigma \|\mathbf{z}\|) \boldsymbol{\Sigma}^{1/2} \mathbf{z} \mathbf{z}^T \boldsymbol{\Sigma}^{1/2} - v_1(c_\sigma \|\mathbf{z}\|) \sigma_P^2 \boldsymbol{\Sigma} / |\boldsymbol{\Sigma}|^{1/k} \right] \\ &= k \boldsymbol{\Sigma}^{1/2} \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|) \mathbf{z} \mathbf{z}^T] \boldsymbol{\Sigma}^{1/2} - \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|) \|\mathbf{z}\|^2] \boldsymbol{\Sigma} \\ &= k \boldsymbol{\Sigma}^{1/2} \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|) \|\mathbf{z}\|^2] \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\mathbf{u} \mathbf{u}^T] \boldsymbol{\Sigma}^{1/2} - \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|) \|\mathbf{z}\|^2] \boldsymbol{\Sigma} \\ &= \mathbf{0},\end{aligned}$$

since $\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\mathbf{u} \mathbf{u}^T] = (1/k) \mathbf{I}_k$, according to Lemma 5. Hence we conclude that

$$\int \Psi_{\mathbf{V}}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) dP(\mathbf{s}) = \mathbf{0}.$$

Since, we may interchange integration and differentiation in $\partial \Lambda_{\boldsymbol{\gamma}} / \partial \boldsymbol{\gamma}$, according to Lemma 9, this means that for each $j, s = 1, \dots, l$,

$$\frac{\partial \Lambda_{\boldsymbol{\gamma}, j}(\boldsymbol{\xi}_P, \sigma_P)}{\partial \gamma_s} = -\text{vec}(\mathbf{V}_P^{-1} \mathbf{L}_j \mathbf{V}_P^{-1})^T \text{vec} \left(\int \frac{\partial \Psi_{\mathbf{V}}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)}{\partial \gamma_s} dP(\mathbf{s}) \right), \quad (11.38)$$

where $\Psi_{\mathbf{V}}$ is defined in (7.8). We have

$$\frac{\partial \Psi_{\mathbf{V}}}{\partial \gamma_s} = \frac{\partial}{\partial \gamma_s} k u_1 \left(\frac{d}{\sigma} \right) (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})(\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^T - \frac{\partial}{\partial \gamma_s} v_1 \left(\frac{d}{\sigma} \right) \sigma^2 \mathbf{V} + \frac{\partial}{\partial \gamma_s} (\log |\mathbf{V}|) \mathbf{V}. \quad (11.39)$$

Because $\mathbf{V}$ satisfies (7.5), it follows that $\partial \mathbf{V} / \partial \gamma_s = \mathbf{L}_s$. Similar to (11.37), for the first term in (11.39) we have at $(\boldsymbol{\xi}_P, \sigma_P)$:

$$\begin{aligned}&\int \frac{\partial}{\partial \gamma_s} k u_1 \left(\frac{d_P}{\sigma_P} \right) \mathbf{e}_P \mathbf{e}_P^T dP(\mathbf{s}) \\ &= \mathbb{E} \left[\mathbb{E} \left[-k u'_1 \left(\frac{d_P}{\sigma_P} \right) \frac{1}{2 \sigma_P d_P} \mathbf{e}_P^T \mathbf{V}_P^{-1} \frac{\partial \mathbf{V}_P}{\partial \gamma_s} \mathbf{V}_P^{-1} \mathbf{e}_P \cdot \mathbf{e}_P \mathbf{e}_P^T \middle| \mathbf{X} \right] \right],\end{aligned}$$

where $d_P^2 = \mathbf{e}_P^T \mathbf{V}_P^{-1} \mathbf{e}_P$ and $\mathbf{e}_P = \mathbf{y} - \mathbf{X} \boldsymbol{\beta}_{1,P}$. Because $\boldsymbol{\beta}_{1,P} = \boldsymbol{\beta}^*$ and $\mathbf{V}_P = \boldsymbol{\Sigma} / |\boldsymbol{\Sigma}|^{1/k}$, the inner expectation on the right hand side can be written as

$$\begin{aligned}&-k \mathbb{E} \left[u'_1 \left(\frac{d^*}{\sigma_P} \right) \frac{|\boldsymbol{\Sigma}|^{2/k}}{2 \sigma_P d^*} (\mathbf{e}^*)^T \boldsymbol{\Sigma}^{-1} \mathbf{L}_s \boldsymbol{\Sigma}^{-1} \mathbf{e}^* \cdot \mathbf{e}^* (\mathbf{e}^*)^T \right] \\ &= -\sigma_P^2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\frac{k u'_1(c_\sigma \|\mathbf{z}\|) (c_\sigma \|\mathbf{z}\|)^3}{2} \mathbf{u}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{L}_s \boldsymbol{\Sigma}^{-1/2} \mathbf{u} \cdot \boldsymbol{\Sigma}^{1/2} \mathbf{u} \mathbf{u}^T \boldsymbol{\Sigma}^{1/2} \right],\end{aligned}$$

where $\mathbf{z} = \Sigma^{-1/2}(\mathbf{y} - \mathbf{X}\beta^*)$ and $\mathbf{u} = \mathbf{z}/\|\mathbf{z}\|$. According to Lemma 5, the expectation of the right hand side is equal to

$$-\sigma_P^2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\frac{k u_1' (c_\sigma \|\mathbf{z}\|) (c_\sigma \|\mathbf{z}\|)^3}{2} \right] \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\mathbf{u}^T \Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \mathbf{u} \Sigma^{1/2} \mathbf{u} \mathbf{u}^T \Sigma^{1/2} \right]. \quad (11.40)$$

For the second term in (11.39) we get at (ξ_P, σ_P) :

$$\begin{aligned} & - \int \frac{\partial}{\partial \gamma_s} v_1 \left(\frac{d_P}{\sigma_P} \right) \sigma_P^2 \mathbf{V}_P dP(\mathbf{s}) \\ & = \mathbb{E} \left[\mathbb{E} \left[v_1' \left(\frac{d_P}{\sigma_P} \right) \frac{1}{2\sigma_P d_P} \mathbf{e}_P^T \mathbf{V}_P^{-1} \mathbf{L}_s \mathbf{V}_P^{-1} \mathbf{e}_P \cdot \sigma_P^2 \mathbf{V}_P - v_1 \left(\frac{d_P}{\sigma_P} \right) \sigma_P^2 \mathbf{L}_s \middle| \mathbf{X} \right] \right], \end{aligned}$$

where $d_P^2 = \mathbf{e}_P^T \mathbf{V}_P^{-1} \mathbf{e}_P$ and $\mathbf{e}_P = \mathbf{y} - \mathbf{X}\beta_{1,P}$. As before, the inner expectation can be written as

$$\begin{aligned} & \sigma_P^2 \mathbb{E} \left[v_1' \left(\frac{d^*}{\sigma_P} \right) \frac{|\Sigma|^{1/k}}{2\sigma_P d^*} (\mathbf{e}^*)^T \Sigma^{-1} \mathbf{L}_s \Sigma^{-1} \mathbf{e}^* \cdot \Sigma - v_1 \left(\frac{d^*}{\sigma_P} \right) \sigma_P^2 \mathbf{L}_s \right] \\ & = \sigma_P^2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\frac{v_1' (c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|}{2} \mathbf{u}^T \Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \mathbf{u} \cdot \Sigma - \sigma_P^2 v_1 (c_\sigma \|\mathbf{z}\|) \mathbf{L}_s \right]. \end{aligned}$$

With Lemma 5, for the expectation of the second term in (11.39) we get

$$\begin{aligned} & \sigma_P^2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\frac{v_1' (c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|}{2} \right] \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\mathbf{u}^T \Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \mathbf{u} \right] \Sigma \\ & \quad - \sigma_P^2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [v_1 (c_\sigma \|\mathbf{z}\|)] \mathbf{L}_s. \end{aligned} \quad (11.41)$$

For the third term in (11.39) we get at (ξ_P, σ_P) :

$$\begin{aligned} & \frac{\partial}{\partial \gamma_s} (\log |\mathbf{V}_P|) \mathbf{V}_P = \left(\frac{\partial \log |\mathbf{V}_P|}{\partial \gamma_s} \right) \mathbf{V}_P + (\log |\mathbf{V}_P|) \frac{\partial \mathbf{V}}{\partial \gamma_s} \\ & = \text{tr} (\mathbf{V}_P^{-1} \mathbf{L}_s) \mathbf{V}_P = \text{tr} (\Sigma^{-1} \mathbf{L}_s) \Sigma, \end{aligned} \quad (11.42)$$

using that $|\mathbf{V}_P| = 1$. It follows that

$$\begin{aligned} & \int \frac{\partial \Psi_{\mathbf{V}}(\mathbf{s}, \xi_P, \sigma_P)}{\partial \gamma_s} dP(\mathbf{s}) \\ & = -\sigma_P^2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\frac{k u_1' (c_\sigma \|\mathbf{z}\|) (c_\sigma \|\mathbf{z}\|)^3}{2} \right] \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\mathbf{u}^T \Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \mathbf{u} \Sigma^{1/2} \mathbf{u} \mathbf{u}^T \Sigma^{1/2} \right] \\ & \quad + \sigma_P^2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\frac{v_1' (c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|}{2} \right] \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\mathbf{u}^T \Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \mathbf{u} \right] \Sigma \\ & \quad - \sigma_P^2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [v_1 (c_\sigma \|\mathbf{z}\|)] \mathbf{L}_s + \text{tr} (\Sigma^{-1} \mathbf{L}_s) \Sigma. \end{aligned} \quad (11.43)$$

In view of (11.38) and (11.43), for the first term in $\partial \Lambda_{\gamma,j} / \partial \gamma_s$ we obtain

$$\begin{aligned} & \text{vec}(\mathbf{V}_P^{-1} \mathbf{L}_j \mathbf{V}_P^{-1})^T \text{vec} \left(\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\mathbf{u}^T \Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \mathbf{u} \Sigma^{1/2} \mathbf{u} \mathbf{u}^T \Sigma^{1/2} \right] \right) \\ & = |\Sigma|^{2/k} \text{vec}(\Sigma^{-1} \mathbf{L}_j \Sigma^{-1})^T \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\text{vec} \left(\Sigma^{1/2} \mathbf{u} \mathbf{u}^T \Sigma^{1/2} \right) \mathbf{u}^T \Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \mathbf{u} \right] \\ & = |\Sigma|^{2/k} \text{vec}(\Sigma^{-1} \mathbf{L}_j \Sigma^{-1})^T (\Sigma^{1/2} \otimes \Sigma^{1/2}) \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\text{vec}(\mathbf{u} \mathbf{u}^T) \text{vec}(\mathbf{u} \mathbf{u}^T)^T \right] \\ & \quad \text{vec} \left(\Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \right) \\ & = |\Sigma|^{2/k} \text{vec}(\Sigma^{-1/2} \mathbf{L}_j \Sigma^{-1/2})^T \frac{1}{k(k+2)} \left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k} + \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \right) \\ & \quad \text{vec} \left(\Sigma^{-1/2} \mathbf{L}_s \Sigma^{-1/2} \right), \end{aligned}$$

using Lemma 5. Application of properties (11.26) and (11.17), yields

$$\begin{aligned} & \text{vec}(\mathbf{V}_P^{-1} \mathbf{L}_j \mathbf{V}_P^{-1})^T \text{vec} \left(\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\mathbf{u}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{L}_s \boldsymbol{\Sigma}^{-1/2} \mathbf{u} \boldsymbol{\Sigma}^{1/2} \mathbf{u} \mathbf{u}^T \boldsymbol{\Sigma}^{1/2} \right] \right) \\ &= \frac{|\boldsymbol{\Sigma}|^{2/k}}{k(k+2)} \left(2\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1} \mathbf{L}_s) + \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j) \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \right). \end{aligned}$$

It follows that the first term in $\partial \Lambda_{\gamma, j} / \partial \gamma_s$ is equal to

$$\sigma_P^2 |\boldsymbol{\Sigma}|^{2/k} \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u'_1(c_\sigma \|\mathbf{z}\|)(c_\sigma \|\mathbf{z}\|)^3]}{2(k+2)} \left(2\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1} \mathbf{L}_s) + \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j) \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \right). \quad (11.44)$$

Similarly, for the second term in $\partial \Lambda_{\gamma, j} / \partial \gamma_s$ we obtain

$$\begin{aligned} & \text{vec}(\mathbf{V}_P^{-1} \mathbf{L}_j \mathbf{V}_P^{-1})^T \text{vec} \left(\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \left[\mathbf{u}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{L}_s \boldsymbol{\Sigma}^{-1/2} \mathbf{u} \right] \boldsymbol{\Sigma} \right) \\ &= |\boldsymbol{\Sigma}|^{2/k} \text{vec}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1})^T \text{tr} \left(\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\mathbf{u} \mathbf{u}^T] \boldsymbol{\Sigma}^{-1/2} \mathbf{L}_s \boldsymbol{\Sigma}^{-1/2} \right) \text{vec}(\boldsymbol{\Sigma}) \\ &= |\boldsymbol{\Sigma}|^{2/k} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \text{vec}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1})^T \text{vec}(\boldsymbol{\Sigma}) \\ &= |\boldsymbol{\Sigma}|^{2/k} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j). \end{aligned}$$

It follows that the second term in $\partial \Lambda_{\gamma, j} / \partial \gamma_s$ is equal to

$$-\sigma_P^2 |\boldsymbol{\Sigma}|^{2/k} \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [v'_1(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|]}{2k} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j). \quad (11.45)$$

Finally, the third term in $\partial \Lambda_{\gamma, j} / \partial \gamma_s$ is equal to

$$\begin{aligned} & \sigma_P^2 |\boldsymbol{\Sigma}|^{2/k} \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [v_1(c_\sigma \|\mathbf{z}\|)] \text{vec}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1})^T \text{vec}(\mathbf{L}_s) \\ &= \sigma_P^2 |\boldsymbol{\Sigma}|^{2/k} \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [v_1(c_\sigma \|\mathbf{z}\|)] \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1} \mathbf{L}_s), \end{aligned} \quad (11.46)$$

and the fourth term in $\partial \Lambda_{\gamma, j} / \partial \gamma_s$ is equal to

$$-|\boldsymbol{\Sigma}|^{2/k} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \text{vec}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1})^T \text{vec}(\boldsymbol{\Sigma}) = -|\boldsymbol{\Sigma}|^{2/k} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j). \quad (11.47)$$

We conclude that $\partial \Lambda_{\gamma, j}(\boldsymbol{\xi}_P) / \partial \gamma_s$ consists of a term $\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1} \mathbf{L}_s)$ from (11.44) and (11.46) with coefficient

$$\begin{aligned} \omega_1 &= \sigma_P^2 |\boldsymbol{\Sigma}|^{2/k} \left(\frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u'_1(c_\sigma \|\mathbf{z}\|)(c_\sigma \|\mathbf{z}\|)^3]}{k+2} + \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [v_1(c_\sigma \|\mathbf{z}\|)] \right) \\ &= \sigma_P^2 |\boldsymbol{\Sigma}|^{2/k} \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho''(c_\sigma \|\mathbf{z}\|)(c_\sigma \|\mathbf{z}\|)^2 + (k+1)\rho'(c_\sigma \|\mathbf{z}\|)c_\sigma \|\mathbf{z}\|]}{k+2}, \end{aligned}$$

and a term $-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j)$ from (11.44), (11.45), and (11.47) with coefficient

$$\begin{aligned} \omega_2 &= \sigma_P^2 |\boldsymbol{\Sigma}|^{2/k} \left(\frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u'_1(c_\sigma \|\mathbf{z}\|)(c_\sigma \|\mathbf{z}\|)^3]}{2(k+2)} - \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [v'_1(c_\sigma \|\mathbf{z}\|)c_\sigma \|\mathbf{z}\|]}{2k} \right) + |\boldsymbol{\Sigma}|^{2/k} \\ &= \sigma_P^2 |\boldsymbol{\Sigma}|^{2/k} \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho''(c_\sigma \|\mathbf{z}\|)(c_\sigma \|\mathbf{z}\|)^2 + (k+1)\rho'(c_\sigma \|\mathbf{z}\|)c_\sigma \|\mathbf{z}\|]}{k(k+2)} + |\boldsymbol{\Sigma}|^{2/k} \\ &= \frac{\omega_1}{k} + |\boldsymbol{\Sigma}|^{2/k}. \end{aligned}$$

From the definition of $\mathbf{L}$ in (7.6) it follows that the $l \times l$ matrix with entries

$$\frac{\partial \Lambda_{\gamma, j}(\boldsymbol{\xi}_P, \sigma_P)}{\partial \gamma_s} = \omega_1 \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j \boldsymbol{\Sigma}^{-1} \mathbf{L}_s) - \omega_2 \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_s) \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{L}_j),$$

is the matrix

$$\mathbf{D}_\gamma = \frac{\partial \Lambda_\gamma(\boldsymbol{\xi}_P, \sigma_P)}{\partial \gamma} = \omega_1 \mathbf{L}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{L} - \omega_2 \mathbf{L}^T \text{vec}(\boldsymbol{\Sigma}^{-1}) \text{vec}(\boldsymbol{\Sigma}^{-1})^T \mathbf{L}.$$

This proves the lemma. $\square$

Lemma 11. Suppose the conditions of Lemma 10 hold. Let $\mathbf{D}_\beta$ be defined in (11.33) with α_1 from (8.5). When $\alpha_1 \neq 0$ and $\mathbf{X}$ has full rank with probability one, then $\mathbf{D}_\beta$ is non-singular with inverse

$$\mathbf{D}_\beta^{-1} = -\frac{1}{\alpha_1 |\boldsymbol{\Sigma}|^{1/k}} (\mathbb{E} [\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}])^{-1}.$$

Let $\mathbf{D}_\gamma$ be defined in (11.34) with $\omega_1 = \sigma^2(P) |\boldsymbol{\Sigma}|^{2/k} \gamma_1$ and $\omega_2 = \omega_1/k + |\boldsymbol{\Sigma}|^{2/k}$, where γ_1 is defined in (8.5) and $\mathbf{L}$ is defined in (7.6). When $\gamma_1 \neq 0$ and $\mathbf{L}$ has full rank, then $\mathbf{D}_\gamma$ is non-singular with inverse

$$\mathbf{D}_\gamma^{-1} = a(\mathbf{E}^T \mathbf{E})^{-1} + b(\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E} (\mathbf{E}^T \mathbf{E})^{-1},$$

where $\mathbf{E} = (\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) \mathbf{L}$, $a = 1/\omega_1$, and $b = \omega_2/(\omega_1(\omega_1 - k\omega_2))$.

Proof. Because $\mathbf{X}$ has full rank with probability one, it follows that $\mathbb{E}[\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}]$ is non-singular. Since $\alpha_1 \neq 0$, this proves part one. Then consider $\mathbf{D}_\gamma$ as given in (11.34). With $\mathbf{E} = (\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) \mathbf{L}$, we can write

$$\mathbf{D}_\gamma = \omega_1 \mathbf{E}^T \mathbf{E} - \omega_2 \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E},$$

which is of the form $\mathbf{A} + \mathbf{u}\mathbf{v}^T$. Because $\mathbf{L}$ has full rank and $\omega_1 = \sigma^2(P) |\boldsymbol{\Sigma}|^{2/k} \gamma_1 \neq 0$, it follows that $\mathbf{A} = \omega_1 \mathbf{E}^T \mathbf{E} = \omega_1 \mathbf{L}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{L}$ is non-singular. Furthermore, since $\mathbf{V}$ has a linear structure, we have that $\text{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}^*$, which implies $\mathbf{E}\boldsymbol{\theta}^* = \text{vec}(\mathbf{I}_k)$ and $(\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) = \boldsymbol{\theta}^*$. This means that $1 + \mathbf{v}^T \mathbf{A}^{-1} \mathbf{u} = (\omega_1 - k\omega_2)/\omega_1 = k|\boldsymbol{\Sigma}|^{2/k}/\omega_1 \neq 0$. It then follows from the Sherman-Morrisson formula that $\mathbf{D}_\gamma$ is non-singular and has inverse

$$\mathbf{D}_\gamma^{-1} = a(\mathbf{E}^T \mathbf{E})^{-1} + b(\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E} (\mathbf{E}^T \mathbf{E})^{-1},$$

where $a = 1/\omega_1$ and $b = \omega_2/(\omega_1(\omega_1 - k\omega_2))$. $\square$

Lemma 12. Suppose P satisfies (E) for some $(\boldsymbol{\beta}^*, \boldsymbol{\theta}^*) \in \mathbb{R}^q \times \boldsymbol{\Theta}$ and $\mathbb{E}\|\mathbf{X}\|^2 < \infty$. Suppose that ρ_1 satisfy (R2) and (R5), and suppose that $\mathbf{V}$ satisfy (V5). Let $\sigma(P)$ be the solution of (3.8) and let $\boldsymbol{\xi}(P) = (\boldsymbol{\beta}_1(P), \boldsymbol{\gamma}(P)) \in \mathfrak{D}$ be a local minimum of $R_P(\boldsymbol{\beta}, \mathbf{V}(\boldsymbol{\gamma}))$ that satisfies $\boldsymbol{\beta}_1(P) = \boldsymbol{\beta}^*$ and $\mathbf{V}(\boldsymbol{\gamma}(P)) = \boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/k}$. Then

$$\mathbf{D}_\sigma = \frac{\partial \Lambda(\boldsymbol{\xi}(P), \sigma(P))}{\partial \sigma} = \mathbf{0}.$$

Proof. For convenience, write $\boldsymbol{\xi}_P = (\boldsymbol{\beta}_{1,P}, \boldsymbol{\gamma}_P) = (\boldsymbol{\beta}_1(P), \boldsymbol{\gamma}(P))$, $\mathbf{V}_P = \mathbf{V}(\boldsymbol{\gamma}(P))$, and $\sigma_P = \sigma(P)$. Consider Λ_β as defined in (8.3) with Ψ_β from (7.7). Because ρ_1 and $\mathbf{V}$ satisfy (R2), (R5), and (V5), and $\mathbb{E}\|\mathbf{X}\|^2 < \infty$, according to Lemma 9, we may interchange differentiation and integration in $\partial \Lambda_\beta / \partial \sigma$. We find

$$\begin{aligned} \frac{\partial \Lambda_\beta(\boldsymbol{\xi}_P, \sigma_P)}{\partial \sigma} &= \int \frac{\partial \Psi_\beta(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)}{\partial \sigma} dP(\mathbf{s}) \\ &= -\mathbb{E} \left[u'_1 \left(\frac{d_P}{\sigma_P} \right) \frac{d_P}{\sigma_P^2} \mathbf{X}^T \mathbf{V}_P^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{1,P}) \right]. \end{aligned}$$

where $d_P^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{1,P})^T \mathbf{V}_P^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{1,P})$. Because $\boldsymbol{\beta}_{1,P} = \boldsymbol{\beta}^*$ is a point of symmetry, it follows that

$$\mathbf{D}_{\beta, \sigma} = \frac{\partial \Lambda_\beta(\boldsymbol{\xi}_P, \sigma_P)}{\partial \sigma} = \mathbf{0}. \quad (11.48)$$

According to Lemma 9, we may also interchange differentiation and integration in $\partial\Lambda_\gamma/\partial\sigma$, where Λ_γ is defined by (8.3) with Ψ_γ from (7.7). For all $j = 1, \dots, l$ we obtain

$$\frac{\partial\Lambda_{\gamma,j}(\xi_P, \sigma_P)}{\partial\sigma} = -\text{vec}(\mathbf{V}_P^{-1}\mathbf{L}_j\mathbf{V}_P^{-1})^T \text{vec}\left(\int \frac{\partial\Psi_{\mathbf{V}}(\mathbf{s}, \xi_P, \sigma_P)}{\partial\sigma} dP(\mathbf{s})\right),$$

where $\Psi_{\mathbf{V}}$ is defined in (7.8). Since $v_1(s) = u_1(s)s^2$, we have

$$\int \frac{\partial\Psi_{\mathbf{V}}(\mathbf{s}, \xi_P, \sigma_P)}{\partial\sigma} dP(\mathbf{s}) = -\mathbb{E}\left[\mathbb{E}\left[ku'_1\left(\frac{d_P}{\sigma_P}\right)\frac{d_P}{\sigma_P^2}\mathbf{e}_P\mathbf{e}_P^T - u'_1\left(\frac{d_P}{\sigma_P}\right)\frac{d_P^3}{\sigma_P^2}\mathbf{V}_P \middle| \mathbf{X}\right]\right],$$

where $d_P^2 = \mathbf{e}_P^T\mathbf{V}_P^{-1}\mathbf{e}_P$ and $\mathbf{e}_P = \mathbf{y} - \mathbf{X}\beta_{1,P}$. The inner expectation on the right hand side is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$. Because $\beta_{1,P} = \beta^*$ and $\mathbf{V}_P = \Sigma/|\Sigma|^{1/k}$, it follows that $d_P^2 = |\Sigma|^{1/k}(\mathbf{y} - \mathbf{X}\beta^*)^T\Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta^*)$. Furthermore, $\mathbf{y} \mid \mathbf{X}$ has the same distribution as $\Sigma^{1/2}\mathbf{z} + \mathbf{X}\beta^*$, where $\mathbf{z}$ has a spherical density $f_{\mathbf{0},\mathbf{I}_k}$. This means that the inner expectation is equal to

$$\begin{aligned} & \mathbb{E}\left[ku'_1\left(\frac{d_P}{\sigma_P}\right)\frac{d_P}{\sigma_P^2}\mathbf{e}_P\mathbf{e}_P^T - u'_1\left(\frac{d_P}{\sigma_P}\right)\frac{d_P^3}{\sigma_P^2}\mathbf{V}_P \middle| \mathbf{X}\right] \\ &= \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[ku'_1(c_\sigma\|\mathbf{z}\|)\frac{c_\sigma\|\mathbf{z}\|}{\sigma_P}\Sigma^{1/2}\mathbf{z}\mathbf{z}^T\Sigma^{1/2} - u'_1(c_\sigma\|\mathbf{z}\|)(c_\sigma\|\mathbf{z}\|)^3\frac{\sigma_P}{|\Sigma|^{1/k}}\Sigma\right] \\ &= \frac{\sigma_P}{|\Sigma|^{1/k}}\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[ku'_1(c_\sigma\|\mathbf{z}\|)(c_\sigma\|\mathbf{z}\|)^3\Sigma^{1/2}\mathbf{u}\mathbf{u}^T\Sigma^{1/2} - u'_1(c_\sigma\|\mathbf{z}\|)(c_\sigma\|\mathbf{z}\|)^3\Sigma\right] \end{aligned}$$

where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma_P$ and $\mathbf{u} = \mathbf{z}/\|\mathbf{z}\|$. Because $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[\mathbf{u}\mathbf{u}^T] = (1/k)\mathbf{I}_k$, according to Lemma 5, the right hand side is equal to

$$\begin{aligned} & \frac{k\sigma_P}{|\Sigma|^{1/k}}\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[u'_1(c_\sigma\|\mathbf{z}\|)(c_\sigma\|\mathbf{z}\|)^3]\Sigma^{1/2}\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[\mathbf{u}\mathbf{u}^T]\Sigma^{1/2} \\ & - \frac{\sigma_P}{|\Sigma|^{1/k}}\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[u'_1(c_\sigma\|\mathbf{z}\|)(c_\sigma\|\mathbf{z}\|)^3]\Sigma = \mathbf{0}. \end{aligned}$$

We conclude that

$$\mathbf{D}_{\gamma,\sigma} = \frac{\partial\Lambda_\gamma(\xi_P, \sigma_P)}{\partial\sigma} = \mathbf{0}. \quad (11.49)$$

Together with (11.48) this proves the lemma. $\square$

Proof of Theorem 6

Proof. Write $\xi_P = (\beta_{1,P}, \gamma_P) = (\beta_1(P), \gamma(P))$, $\mathbf{V}_P = \mathbf{V}(\gamma(P))$, and $\sigma_P = \sigma(P)$. Because $(\beta_{1,P}, \mathbf{V}_P) = (\beta^*, \Sigma/|\Sigma|^{1/k})$, we have

$$d_{0,P}^2 = (\mathbf{y}_0 - \mathbf{X}_0\beta_{1,P})^T\mathbf{V}_P^{-1}(\mathbf{y}_0 - \mathbf{X}_0\beta_{1,P}) = |\Sigma|^{1/k}\|\mathbf{z}_0\|^2,$$

where $\mathbf{z}_0 = \Sigma^{-1/2}(\mathbf{y}_0 - \mathbf{X}_0\beta^*)$. This means that

$$\Psi_\beta(\mathbf{s}_0, \xi_P, \sigma_P) = u_1\left(\frac{d_{0,P}}{\sigma_P}\right)\mathbf{X}_0^T\mathbf{V}_P^{-1}(\mathbf{y}_0 - \mathbf{X}_0\beta^*) = |\Sigma|^{1/k}u_1(c_\sigma\|\mathbf{z}_0\|)\mathbf{X}_0^T\Sigma^{-1/2}\mathbf{z}_0,$$

where $c_\sigma = |\Sigma|^{1/(2k)}/\sigma_P$. From Lemmas 9 and 10, we have that Λ_β is continuously differentiable at (ξ_P, σ_P) , with a derivative given by $\mathbf{D}_\beta = -\alpha_1|\Sigma|^{1/k}\mathbb{E}[\mathbf{X}^T\Sigma^{-1}\mathbf{X}]$, which is non-singular according to Lemma 11. Because $\beta_{1,P} = \beta^*$ is a point of symmetry, from Theorem 5 we obtain

$$\begin{aligned} \text{IF}(\mathbf{s}_0; \beta_1, P) &= -\mathbf{D}_\beta^{-1}\Psi_\beta(\mathbf{s}_0; \xi_P, \sigma_P) \\ &= \frac{u_1(c_\sigma\|\mathbf{z}_0\|)}{\alpha_1}(\mathbb{E}[\mathbf{X}^T\Sigma^{-1}\mathbf{X}])^{-1}\mathbf{X}_0^T\Sigma^{-1/2}\mathbf{z}_0. \end{aligned}$$

This proves part one.

For part two, we apply Lemma 2(i). Consider Λ as defined in (8.2) with Ψ defined in (7.7). From Lemmas 9 and 10, we have that Λ is continuously differentiable at (ξ_P, σ_P) , with a derivative given by

$$\mathbf{D}_\xi = \begin{pmatrix} \mathbf{D}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\gamma \end{pmatrix},$$

where $\mathbf{D}_\beta$ and $\mathbf{D}_\gamma$ are given in (11.33) and (11.34). According to Lemma 11, $\mathbf{D}_\beta$ and $\mathbf{D}_\gamma$ are non-singular, which implies that $\mathbf{D}_\xi$ is non-singular. Furthermore, from Lemma 12 we have $\mathbf{D}_\sigma = \partial\Lambda(\xi_P, \sigma_P)/\partial\sigma = \mathbf{0}$. From Lemma 2(i), this means that

$$\text{IF}(\mathbf{s}_0; \gamma, P) = -\mathbf{D}_\gamma^{-1} \Psi_\gamma(\mathbf{s}_0; \xi_P, \sigma_P),$$

where $\mathbf{D}_\gamma^{-1}$ is given in Lemma 11. The remaining derivation of the expression for $\text{IF}(\mathbf{s}_0; \gamma, P)$ runs along the same line of reasoning as in (the second part of) the proof of Corollary 8.4 in Lopuhaä *et al* [22]. Using that $\log |\mathbf{V}_P| = 0$, we find that

$$\begin{aligned} & \text{IF}(\mathbf{s}_0; \gamma, P) \\ &= -\mathbf{D}_\gamma^{-1} \text{vec}(\Psi_\gamma(\mathbf{s}_0; \xi_P, \sigma_P)) \\ &= \mathbf{D}_\gamma^{-1} \mathbf{L}^T (\mathbf{V}_P^{-1} \otimes \mathbf{V}_P^{-1}) \text{vec} \left\{ k u_1 \left(\frac{d_{0,P}}{\sigma_P} \right) \mathbf{e}_{0,P} \mathbf{e}_{0,P}^T - v_1 \left(\frac{d_{0,P}}{\sigma_P} \right) \sigma_P^2 \mathbf{V}_P \right\} \\ &= k |\Sigma|^{2/k} u_1(c_\sigma \|\mathbf{z}_0\|) \mathbf{D}_\gamma^{-1} \mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec} \{ \mathbf{e}_0^* (\mathbf{e}_0^*)^T \} \\ &\quad - \sigma_P^2 |\Sigma|^{1/k} v_1(c_\sigma \|\mathbf{z}_0\|) \mathbf{D}_\gamma^{-1} \mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(\Sigma), \end{aligned} \quad (11.50)$$

where $d_{0,P}^2 = \mathbf{e}_{0,P}^T \mathbf{V}_P^{-1} \mathbf{e}_{0,P}$ with $\mathbf{e}_{0,P} = \mathbf{y}_0 - \mathbf{X}_0 \beta_{1,P}$, and where $\mathbf{z}_0 = \Sigma^{-1/2} \mathbf{e}_0^*$ with $\mathbf{e}_0^* = \mathbf{y}_0 - \mathbf{X}_0 \beta^*$, and where $c_\sigma = \sigma(P)/|\Sigma|^{1/(2k)}$. Consider the first term on the right hand side of (11.50). We have that

$$\mathbf{D}_\gamma^{-1} \mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec} \{ \mathbf{e}_0^* (\mathbf{e}_0^*)^T \} = \mathbf{D}_\gamma^{-1} \mathbf{L}^T (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T).$$

From Lemma 11 we obtain

$$\begin{aligned} & \mathbf{D}_\gamma^{-1} \mathbf{L}^T (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \\ &= a (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T + b (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E} (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T, \end{aligned} \quad (11.51)$$

where $\mathbf{E} = (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \mathbf{L}$, $a = 1/\omega_1$, and $b = \omega_2/(\omega_1(\omega_1 - k\omega_2))$. This implies that

$$\begin{aligned} & \mathbf{D}_\gamma^{-1} \mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(\mathbf{e}_0^* (\mathbf{e}_0^*)^T) \\ &= a (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) \\ &\quad + b (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E} (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T). \end{aligned} \quad (11.52)$$

The first term on the right hand side of (11.52) is equal to

$$a \left(\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} \right)^{-1} \mathbf{L}^T (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T).$$

Since $\mathbf{V}$ has a linear structure, we have $\text{vec}(\Sigma) = |\Sigma|^{1/k} \text{vec}(\mathbf{V}_P) = |\Sigma|^{1/k} \mathbf{L} \gamma_P$. This means that

$$\begin{aligned} \mathbf{E} \gamma_P &= (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \mathbf{L} \gamma_P \\ &= (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \text{vec}(\Sigma) / |\Sigma|^{1/k} = \text{vec}(\mathbf{I}_k) / |\Sigma|^{1/k}, \end{aligned} \quad (11.53)$$

and

$$(\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) = |\Sigma|^{1/k} \gamma_P. \quad (11.54)$$

It follows that the second term on the right hand side of (11.52) is equal to

$$\begin{aligned} b \left(|\Sigma|^{1/k} \gamma_P \right) \left(|\Sigma|^{1/k} \gamma_P \right)^T \mathbf{E}^T \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) &= b |\Sigma|^{1/k} \gamma_P \text{vec}(\mathbf{I}_k)^T \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) \\ &= b |\Sigma|^{1/k} \gamma_P \text{tr}(\mathbf{z}_0 \mathbf{z}_0^T) = b |\Sigma|^{1/k} \gamma_P \|\mathbf{z}_0\|^2. \end{aligned}$$

It follows that the first term on the right hand side of (11.50) is equal to

$$\begin{aligned} ka |\Sigma|^{2/k} u_1(c_\sigma \|\mathbf{z}_0\|) \left(\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} \right)^{-1} \mathbf{L}^T \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) \\ + kb |\Sigma|^{3/k} u_1(c_\sigma \|\mathbf{z}_0\|) \|\mathbf{z}_0\|^2 \gamma_P. \end{aligned}$$

Next, consider the second term on the right hand side of (11.50). From (11.51), together with (11.53) and (11.54), we have

$$\begin{aligned} \mathbf{D}_\gamma^{-1} \mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(\Sigma) \\ &= \mathbf{D}_\gamma^{-1} \mathbf{L}^T (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \text{vec}(\mathbf{I}_k) \\ &= a (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) + b (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E} (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \\ &= a |\Sigma|^{1/k} \gamma_P + b (|\Sigma|^{1/k} \gamma_P) \text{vec}(\mathbf{I}_k)^T \mathbf{E} (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \\ &= a |\Sigma|^{1/k} \gamma_P + b (|\Sigma|^{1/k} \gamma_P) |\Sigma|^{1/k} \gamma_P^T \mathbf{E}^T \text{vec}(\mathbf{I}_k) \\ &= a |\Sigma|^{1/k} \gamma_P + b (|\Sigma|^{1/k} \gamma_P) \text{vec}(\mathbf{I}_k)^T \text{vec}(\mathbf{I}_k) \\ &= (a + kb) |\Sigma|^{1/k} \gamma_P. \end{aligned}$$

It follows that the second term on the right hand side of (11.50) is equal to

$$-\sigma_P^2 |\Sigma|^{2/k} v_1(c_\sigma \|\mathbf{z}_0\|) (a + kb) \gamma_P.$$

Putting things together, we find that $\text{IF}(\mathbf{s}_0; \gamma, P)$ is equal to

$$\begin{aligned} ka |\Sigma|^{2/k} u_1(c_\sigma \|\mathbf{z}_0\|) \left(\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} \right)^{-1} \mathbf{L}^T \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) \\ + kb |\Sigma|^{3/k} u_1(c_\sigma \|\mathbf{z}_0\|) \|\mathbf{z}_0\|^2 \gamma_P - \sigma_P^2 |\Sigma|^{2/k} v_1(c_\sigma \|\mathbf{z}_0\|) (a + kb) \gamma_P. \end{aligned}$$

The term with γ_P has coefficient

$$\begin{aligned} kb |\Sigma|^{3/k} u_1(c_\sigma \|\mathbf{z}_0\|) \|\mathbf{z}_0\|^2 - \sigma_P^2 |\Sigma|^{2/k} v_1(c_\sigma \|\mathbf{z}_0\|) (a + kb) \\ = kb \sigma_P^2 |\Sigma|^{2/k} v_1(c_\sigma \|\mathbf{z}_0\|) - \sigma_P^2 |\Sigma|^{2/k} v_1(c_\sigma \|\mathbf{z}_0\|) (a + kb) \\ = -a \sigma_P^2 |\Sigma|^{2/k} v_1(c_\sigma \|\mathbf{z}_0\|), \end{aligned}$$

using that $v_1(s) = u_1(s)s^2$. Note that $a = 1/\omega_1 = 1/(\sigma_P^2 |\Sigma|^{2/k} \gamma_1)$. It follows that

$$\begin{aligned} \text{IF}(\mathbf{s}_0; \gamma, P) &= \frac{ku_1(c_\sigma \|\mathbf{z}_0\|)}{\sigma_P^2 \gamma_1} \left(\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L} \right)^{-1} \mathbf{L}^T \left(\Sigma^{-1/2} \otimes \Sigma^{-1/2} \right) \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T) \\ &\quad - \frac{v_1(c_\sigma \|\mathbf{z}_0\|)}{\gamma_1} \gamma_P. \end{aligned}$$

Finally, due to linearity of $\mathbf{V}$, we have

$$\mathbf{L} \gamma_P = \mathbf{V}(\gamma_P) = \frac{\Sigma}{|\Sigma|^{1/k}} = \frac{\mathbf{V}(\boldsymbol{\theta}^*)}{|\Sigma|^{1/k}} = \frac{\mathbf{L} \boldsymbol{\theta}^*}{|\Sigma|^{1/k}}. \quad (11.55)$$

Since $\mathbf{L}$ has full rank, we can multiply from the left by $(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T$, which implies that $\gamma_P = \boldsymbol{\theta}^* / |\Sigma|^{1/k}$. This proves the theorem. $\square$

Proof of Corollary 4

Proof. Write $(\gamma_P, \sigma_P) = (\gamma(P), \sigma(P))$ and $(\gamma_{h,s_0}, \sigma_{h,s_0}) = (\gamma(P_{h,s_0}), \sigma(P_{h,s_0}))$. Because $\theta_1(P)$ is a solution of (3.11) and $\mathbf{V}$ has a linear structure, it follows that

$$\mathbf{L}\theta_1(P) = \mathbf{V}(\theta_1(P)) = \sigma_P^2 \mathbf{V}(\gamma_P) = \sigma_P^2 \mathbf{L}\gamma_P. \quad (11.56)$$

Since $\mathbf{L}$ has full rank, we can multiply from the left by $(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T$, which implies that $\theta_1(P) = \sigma_P^2 \gamma_P$ and a similar property holds for $\theta_1(P_{h,s_0})$. We find that

$$\begin{aligned} \theta_1(P_{h,s_0}) - \theta_1(P) &= \sigma_{h,s_0}^2 (\gamma_{h,s_0} - \gamma_P) + \gamma_P (\sigma_{h,s_0}^2 - \sigma_P^2) \\ &= \sigma_{h,s_0}^2 (\gamma_{h,s_0} - \gamma_P) + \gamma_P (\sigma_{h,s_0} + \sigma_P) (\sigma_{h,s_0} - \sigma_P). \end{aligned}$$

According to Lemma 7 and Theorem 6, $\text{IF}(\mathbf{s}_0; \sigma, P)$ and $\text{IF}(\mathbf{s}_0; \gamma, P)$ exist. Together with $\sigma_{h,s_0} \rightarrow \sigma_P$, we obtain

$$\text{IF}(\mathbf{s}_0; \theta_1, P) = \sigma_P^2 \text{IF}(\mathbf{s}_0; \gamma, P) + 2\sigma_P \gamma_P \text{IF}(\mathbf{s}_0; \sigma, P).$$

Because $\mathbf{V}$ has a linear structure, from (11.55), we have $\gamma_P = \theta^* / |\Sigma|^{1/k}$, after multiplication from the left by $(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T$. The corollary now follows from the expressions obtained in Lemma 7 and Theorem 6. $\square$

11.6 Influence functions of covariance MM-functionals

We provide some details about the influence functions of the covariance MM-functionals in Examples 1, 2, and 3 for the situation where the distribution P satisfies (E).

Example 1 (Linear Mixed Effects model). *For the influence function of covariance MM-functionals in linear mixed effects models, nothing seems to be available yet. For model (2.3), the expression for the influence function of the variance component MM-functional now follows from Corollary 4 and the one for the covariance MM-functional and the corresponding shape component follow from (8.6) and (8.8), respectively. From (7.5) and (7.6), it follows that*

$$\mathbf{L} = \begin{bmatrix} \text{vec}(\mathbf{I}_k) & \text{vec}(\mathbf{Z}_1 \mathbf{Z}_1^T) & \cdots & \text{vec}(\mathbf{Z}_r \mathbf{Z}_r^T) \end{bmatrix}. \quad (11.57)$$

Furthermore, it can be seen that $\mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \mathbf{L}$ is equal to the matrix $\mathbf{Q}$ with entries

$$Q_{ij} = \text{tr}(\mathbf{Z}_i \mathbf{Z}_i^T \Sigma^{-1} \mathbf{Z}_j \mathbf{Z}_j^T \Sigma^{-1}), \quad i, j = 0, 1, \dots, r, \quad (11.58)$$

where $\mathbf{Z}_0 = \mathbf{I}_k$, and that $\mathbf{L}^T (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \text{vec}(\mathbf{z}_0 \mathbf{z}_0^T)$ is equal to the vector $\mathbf{U}$ with coordinates

$$U_i = (\mathbf{y}_0 - \mathbf{X}_0 \beta^*)^T \Sigma^{-1} \mathbf{Z}_i \mathbf{Z}_i^T \Sigma^{-1} (\mathbf{y}_0 - \mathbf{X}_0 \beta^*), \quad i = 0, 1, \dots, r,$$

This implies that the influence function of the variance component MM-functional is given by

$$\text{IF}(\mathbf{s}_0, \theta_1, P) = \alpha_C(c_\sigma \|\mathbf{z}_0\|) \mathbf{Q}^{-1} \mathbf{U} - \beta_C(c_\sigma \|\mathbf{z}_0\|) \theta^*,$$

with α_C and β_C defined in (8.7), and where $\mathbf{z}_0 = \Sigma^{-1/2} (\mathbf{y}_0 - \mathbf{X}_0 \beta^*)$ and $c_\sigma = |\Sigma|^{1/(2k)} / \sigma(P)$. From (8.6) we find that the influence functional of the covariance MM-functional is given by

$$\text{IF}(\mathbf{s}_0, \text{vec}(\mathbf{V}(\theta_1)), P) = \alpha_C(c_\sigma \|\mathbf{z}_0\|) \mathbf{L} \mathbf{Q}^{-1} \mathbf{U} - \beta_C(c_\sigma \|\mathbf{z}_0\|) \text{vec}(\Sigma).$$

From (8.8), we obtain the influence function of the shape MM-functional

$$\text{IF}(\mathbf{s}_0, \Gamma(\theta_1), P) = \frac{\alpha_C(c_\sigma \|\mathbf{z}_0\|)}{\sigma^2(P)} \left\{ \mathbf{L} \mathbf{Q}^{-1} \mathbf{U} - \frac{\|\mathbf{z}_0\|^2}{k} \text{vec}(\Sigma) \right\},$$

and in view of Remark 8.1, from Theorem 6 we obtain the influence function of the direction MM-functional

$$\text{IF}(\mathbf{s}_0, \gamma, P) = \frac{\alpha_C(c_\sigma \|\mathbf{z}_0\|)}{\sigma^2(P)} \left\{ \mathbf{Q}^{-1} \mathbf{U} - \frac{\|\mathbf{z}_0\|^2}{k} \boldsymbol{\theta}^* \right\}.$$

When $c_\sigma = 1$, then the influence function of the covariance shape MM-functional coincides with that of the shape component of the covariance S-functional defined with ρ_1 , and similarly for the direction component of the variance component S-functional defined with ρ_1 , see Lopuhaä et al [22].

Example 2 (Multivariate Linear Regression). For the multivariate linear regression model (2.4), Kudraszow and Maronna [15] do not consider the influence function of the covariance MM-functional. In this model, the matrix $\mathbf{L}$ is equal to the duplication matrix $\mathcal{D}_k$, which satisfies

$$\mathcal{D}_k (\mathcal{D}_k^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \mathcal{D}_k)^{-1} \mathcal{D}_k^T = \frac{1}{2} (\mathbf{I}_{k^2} + \mathbf{K}_{k,k}) (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}), \quad (11.59)$$

(e.g., see Magnus and Neudecker [24, Ch. 3, Sec. 8]). Together with (8.6) we find that the expression for the covariance MM-functional is given by

$$\text{IF}(\mathbf{s}_0, \mathbf{V}(\boldsymbol{\theta}_1), P) = \alpha_C(c_\sigma \|\mathbf{z}_0\|) (\mathbf{y}_0 - \mathbf{B}^T \mathbf{x}_0)(\mathbf{y}_0 - \mathbf{B}^T \mathbf{x}_0)^T - \beta_C(c_\sigma \|\mathbf{z}_0\|) \boldsymbol{\Sigma}, \quad (11.60)$$

with α_C and β_C defined in (8.7), $\mathbf{z}_0 = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_0 - \mathbf{B}^T \mathbf{x}_0)$, and $c_\sigma = |\boldsymbol{\Sigma}|^{1/(2k)}/\sigma(P)$. From (8.8), it follows that the influence function of the shape MM-functional is given by

$$\text{IF}(\mathbf{s}_0; \boldsymbol{\Gamma}(\boldsymbol{\theta}_1), P) = \frac{\alpha_C(c_\sigma \|\mathbf{z}_0\|)}{\sigma^2(P)} \left\{ (\mathbf{y}_0 - \mathbf{B}^T \mathbf{x}_0)(\mathbf{y}_0 - \mathbf{B}^T \mathbf{x}_0)^T - \frac{\|\mathbf{z}_0\|^2}{k} \boldsymbol{\Sigma} \right\}. \quad (11.61)$$

When $c_\sigma = 1$, this also coincides with the influence function of shape component of the covariance S-functional defined with ρ_1 in the multivariate linear regression model. This is confirmed by application of formula (8.3) in Kent and Tyler [14] to the expression found in Theorem 2 of Van Aelst and Willems [35].

Example 3 (Multivariate Location and Scatter). For the multivariate location-scatter model, we also have $\mathbf{L} = \mathcal{D}_k$. Since this model is a special case of the multivariate linear regression model (2.4) by taking $\mathbf{x}_i = 1$ and $\mathbf{B}^T = \boldsymbol{\mu}$, the expression for the influence function of the covariance MM-functional can be obtained from (11.60) and the influence function of the covariance shape MM-functional from (11.61), by replacing $\mathbf{B}^T \mathbf{x}_0$ by $\boldsymbol{\mu}$. When $c_\sigma = 1$, this also coincides with the influence function of shape component of the covariance S-functional defined with ρ_1 in the multivariate location-scatter model. This was already observed by Salibián-Barrera et al [31]. Finally, again there is a connection with the CM-functionals considered in Kent and Tyler [14], whose influence function depends on a parameter λ_0 . By using (11.59), one finds that the expression in (11.60), with $\mathbf{B}^T \mathbf{x}_0 = \boldsymbol{\mu}$, is similar to the expression for the influence function of the covariance CM-functional, for the case that $\lambda_0 = \lambda_L$ (see Kent and Tyler [14] for details), and they both coincide when $c_\sigma = 1/\sqrt{\lambda_0}$ and $\rho_0 = \rho_1$.

11.7 Proofs for Section 9

Lemma 13. Suppose that ρ_1 satisfies (R1)-(R4), such that $u_1(s)$ is of bounded variation, and suppose that $\mathbf{V}$ satisfies (V4). Let σ_n and $\sigma(P)$ be solutions of (3.2) and (3.8), respectively, and let $\boldsymbol{\xi}_n = (\beta_{1,n}, \gamma_n)$ and $\boldsymbol{\xi}(P) = (\beta_1(P), \gamma(P))$ be local minima of $R_n(\beta, \mathbf{V}(\gamma))$ and $R_P(\beta, \mathbf{V}(\gamma))$, respectively. Suppose that $(\boldsymbol{\xi}_n, \sigma_n) \rightarrow (\boldsymbol{\xi}(P), \sigma(P))$, in probability. Let Λ be defined in (8.2) with Ψ defined in (7.3). Suppose that Λ is continuously differentiable with a non-singular derivative $\mathbf{D}_\xi = \partial \Lambda / \partial \boldsymbol{\xi}$ and derivative $\mathbf{D}_\sigma = \partial \Lambda / \partial \sigma$ at $(\boldsymbol{\xi}(P), \sigma(P))$, and suppose that $\mathbb{E}\|\mathbf{s}\|^2 < \infty$. Then

$$\begin{aligned} \boldsymbol{\xi}_n - \boldsymbol{\xi}(P) &= -\mathbf{D}_\xi^{-1} \left\{ \mathbf{D}_\sigma(\sigma_n - \sigma(P)) + \int \Psi(\mathbf{s}, \boldsymbol{\xi}(P), \sigma(P)) d(\mathbb{P}_n - P)(\mathbf{s}) \right\} \\ &\quad + o_P(\|\boldsymbol{\xi}_n - \boldsymbol{\xi}(P)\|) + o_P(|\sigma_n - \sigma(P)|) + o_P(1/\sqrt{n}), \end{aligned}$$

Proof. The proof is similar to that of Theorem 9.1 in Lopuhaä *et al* [22]. Due to (3.12), we have that $\xi_n = \xi(\mathbb{P}_n) = (\beta_{1,n}, \gamma_n)$ and $\sigma_n = \sigma(\mathbb{P}_n)$. This means that (ξ_n, σ_n) satisfies score equation (7.2) for the MM-functionals at $\mathbb{P}_n$, that is

$$\int \Psi(\mathbf{s}, \xi_n, \sigma_n) d\mathbb{P}_n(\mathbf{s}) = \mathbf{0},$$

with Ψ defined in (7.3). Writing $\xi_P = (\beta_{1,P}, \gamma_P) = (\beta_1(P), \gamma(P))$, we decompose as follows

$$\begin{aligned} \mathbf{0} &= \int \Psi(\mathbf{s}, \xi_n, \sigma_n) dP(\mathbf{s}) + \int \Psi(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) \\ &\quad + \int (\Psi(\mathbf{s}, \xi_n, \sigma_n) - \Psi(\mathbf{s}, \xi_P, \sigma_P)) d(\mathbb{P}_n - P)(\mathbf{s}). \end{aligned} \quad (11.62)$$

We start by showing that the third term on the right hand side of (11.62) is of order $o_P(1/\sqrt{n})$. Since $\Psi = (\Psi_\beta, \Psi_\gamma)$, with $\Psi_\gamma = (\Psi_{\gamma,1}, \dots, \Psi_{\gamma,l})$, it suffices to show that

$$\int (\Psi_\beta(\mathbf{s}, \xi_n, \sigma_n) - \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P)) d(\mathbb{P}_n - P)(\mathbf{s}) = o_P(1/\sqrt{n}), \quad (11.63)$$

$$\int (\Psi_{\gamma,j}(\mathbf{s}, \xi_n, \sigma_n) - \Psi_{\gamma,j}(\mathbf{s}, \xi_P, \sigma_P)) d(\mathbb{P}_n - P)(\mathbf{s}) = o_P(1/\sqrt{n}), \quad (11.64)$$

for $j = 1, \dots, l$. From (7.3) we have that

$$\Psi_\beta(\mathbf{s}, \xi, \sigma) = u_1 \left(\frac{d}{\sigma} \right) \mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) = u_1(d(\mathbf{y}, \mathbf{X}\beta, \sigma^2 \mathbf{V})) \mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta).$$

Since $\mathbf{V}$ satisfies (V1) we have that $\sigma_n^2 \mathbf{V}(\gamma_n) \rightarrow \sigma_P^2 \mathbf{V}(\gamma_P)$, in probability. Then (11.63) follows from equation (96) in the proof of Lemma 11.8 in Lopuhaä *et al* [23]. From (7.3), we also have

$$\Psi_{\gamma,j}(\mathbf{s}, \xi, \sigma) = u_1 \left(\frac{d}{\sigma} \right) (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} \mathbf{H}_{1,j} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) - \text{tr} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \gamma_j} \right) \log |\mathbf{V}|,$$

for $j = 1, \dots, l$, where $\mathbf{H}_{1,j} = \mathbf{H}_{1,j}(\gamma)$ is defined in (7.4). Because $|\mathbf{V}(\gamma_n)| = 1$, it follows that

$$\Psi_{\gamma,j}(\mathbf{s}, \xi_n, \sigma_n) = u_1(d_n) (\mathbf{y} - \mathbf{X}\beta_{1,n})^T \mathbf{V}(\gamma_n)^{-1} \mathbf{H}_{1,j}(\gamma_n) \mathbf{V}(\gamma_n)^{-1} (\mathbf{y} - \mathbf{X}\beta_{1,n}),$$

where $d_n^2 = (\mathbf{y} - \mathbf{X}\beta_{1,n})^T (\sigma_n^2 \mathbf{V}(\gamma_n))^{-1} (\mathbf{y} - \mathbf{X}\beta_{1,n})$ and similarly for $\Psi_{\gamma,j}(\mathbf{s}, \xi_P, \sigma_P)$. Because $\mathbf{V}$ satisfies (V4) we have that $\sigma_n^2 \mathbf{V}(\gamma_n) \rightarrow \sigma_P^2 \mathbf{V}(\gamma_P)$ and $\mathbf{H}_{1,j}(\gamma_n) \rightarrow \mathbf{H}_{1,j}(\gamma_P)$, in probability. This implies that (11.64) follows from equation (97) in the proof of Lemma 11.8 in Lopuhaä *et al* [23].

Then, from (11.62) we can write

$$\mathbf{0} = \Lambda(\xi_n, \sigma_n) + \int \Psi(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}).$$

Because Λ is continuously differentiable at (ξ_P, σ_P) , it follows that

$$\begin{aligned} \Lambda(\xi_n, \sigma_n) &= \Lambda(\xi_P, \sigma_n) + \frac{\partial \Lambda}{\partial \xi}(\xi_P, \sigma_n)(\xi_n - \xi_P) + o_P(\|\xi_n - \xi_P\|) \\ &= \Lambda(\xi_P, \sigma_n) + \left(\frac{\partial \Lambda}{\partial \xi}(\xi_P, \sigma_P) + o_P(1) \right) (\xi_n - \xi_P) + o_P(\|\xi_n - \xi_P\|). \end{aligned}$$

Furthermore, since ξ_P is a solution of (7.2), we find

$$\begin{aligned} \Lambda(\xi_P, \sigma_n) &= \Lambda(\xi_P, \sigma_P) + \frac{\partial \Lambda}{\partial \sigma}(\xi_P, \sigma_P)(\sigma_n - \sigma_P) + o_P(|\sigma_n - \sigma_P|) \\ &= \frac{\partial \Lambda}{\partial \sigma}(\xi_P, \sigma_P)(\sigma_n - \sigma_P) + o_P(|\sigma_n - \sigma_P|). \end{aligned}$$

Putting things together, we obtain

$$\begin{aligned} \mathbf{0} &= \mathbf{D}_\sigma(\sigma_n - \sigma_P) + \mathbf{D}_\xi(\xi_n - \xi_P) + \int \Psi(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) \\ &\quad + o_P(\|\xi_n - \xi_P\|) + o_P(|\sigma_n - \sigma_P|) + o_P(1/\sqrt{n}). \end{aligned}$$

Because $\mathbf{D}_\xi$ is non-singular, it follows that

$$\begin{aligned} \xi_n - \xi_P &= -\mathbf{D}_\xi^{-1} \left\{ \mathbf{D}_\sigma(\sigma_n - \sigma_P) + \int \Psi(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) \right\} \\ &\quad + o_P(\|\xi_n - \xi_P\|) + o_P(|\sigma_n - \sigma_P|) + o_P(1/\sqrt{n}). \end{aligned}$$

This proves the lemma. $\square$

Lemma 14. Suppose that ρ_0 satisfies (R1)-(R2) and $\mathbf{V}$ satisfies (V4). Let $\zeta_{0,n} = (\beta_{0,n}, \theta_{0,n})$ and $\zeta_0(P) = (\beta_0(P), \theta_0(P))$ be the pairs of initial estimators and corresponding functionals, and let σ_n and $\sigma(P)$ be solutions of (3.2) and (3.8), respectively. Suppose that $(\zeta_{0,n}, \sigma_n) \rightarrow (\zeta_0(P), \sigma(P))$, in probability. Let Λ_0 be defined in (11.22) with Ψ_0 defined in (11.21), and suppose that Λ_0 is continuously differentiable with derivatives $D_{0,\sigma} = \partial\Lambda_0/\partial\sigma \neq 0$ and $\mathbf{D}_{0,\zeta} = \partial\Lambda_0/\partial\zeta$ at $(\zeta(P), \sigma(P))$. Then

$$\begin{aligned} \sigma_n - \sigma(P) &= -D_{0,\sigma}^{-1} \left\{ \mathbf{D}_{0,\zeta}^T(\zeta_{0,n} - \zeta_0(P)) + \int \Psi_0(\mathbf{s}, \zeta_0(P), \sigma(P)) d(\mathbb{P}_n - P)(\mathbf{s}) \right\} \\ &\quad + o(\|\zeta_{0,n} - \zeta_0(P)\|) + o_P(|\sigma_n - \sigma(P)|) + o_P(1/\sqrt{n}). \end{aligned}$$

Proof. Denote $\zeta_{0,n} = (\beta_{0,n}, \theta_{0,n})$ and let σ_n and $\sigma(P)$ be solutions of (3.2) and (3.8), respectively. Then, according to (3.2), $(\zeta_{0,n}, \sigma_n)$ satisfies

$$\int \Psi_0(\mathbf{s}, \zeta_{0,n}, \sigma_n) d\mathbb{P}_n(\mathbf{s}) = 0.$$

Writing $\zeta_{0,P} = (\beta_{0,P}, \theta_{0,P})$ and $\sigma_P = \sigma(P)$, we decompose as follows

$$\begin{aligned} \mathbf{0} &= \int \Psi_0(\mathbf{s}, \zeta_{0,n}, \sigma_n) dP(\mathbf{s}) + \int \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) \\ &\quad + \int (\Psi_0(\mathbf{s}, \zeta_{0,n}, \sigma_n) - \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P)) d(\mathbb{P}_n - P)(\mathbf{s}). \end{aligned} \tag{11.65}$$

Consider the third term on the right hand side, where we can write

$$\Psi_0(\mathbf{s}, \zeta, \sigma) = \rho_0(d(\mathbf{y}, \mathbf{X}\beta, \sigma^2\Gamma(\theta))) - b_0,$$

where $d(\mathbf{y}, \mathbf{X}\beta, \sigma^2\Gamma(\theta))$ is defined in (7.1) and $\Gamma(\theta) = \mathbf{V}(\theta)/|\mathbf{V}(\theta)|^{1/k}$. Because $\mathbf{V}$ satisfies (V4) we have that $\sigma_n^2\Gamma(\theta_{0,n}) \rightarrow \sigma_P^2\Gamma(\theta_{0,P})$, in probability. Since ρ_0 and $\mathbf{V}$ satisfy the conditions needed to establish (98) in the proof of Lemma 11.8 in Lopuhaä *et al* [23], it follows that

$$\int (\Psi_0(\mathbf{s}, \zeta_{0,n}, \sigma_n) - \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P)) d(\mathbb{P}_n - P)(\mathbf{s}) = o_P(1/\sqrt{n}).$$

Then from (11.65) we can write

$$0 = \Lambda_0(\zeta_{0,n}, \sigma_n) + \int \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}).$$

Because the partial derivative $\partial\Lambda_0/\partial\sigma$ is continuous at $(\zeta_{0,P}, \sigma_P)$, it follows that

$$\begin{aligned} \Lambda_0(\zeta_{0,n}, \sigma_n) &= \Lambda_0(\zeta_{0,n}, \sigma_P) + \frac{\partial\Lambda_0}{\partial\sigma}(\zeta_{0,n}, \sigma_P)(\sigma_n - \sigma_P) + o_P(|\sigma_n - \sigma_P|) \\ &= \Lambda(\zeta_{0,n}, \sigma_P) + D_{0,\sigma}(\sigma_n - \sigma_P) + o_P(|\sigma_n - \sigma_P|). \end{aligned}$$

Furthermore, since σ_P is a solution of (3.8), we find

$$\begin{aligned}\Lambda(\zeta_{0,n}, \sigma_P) &= \Lambda(\zeta_{0,P}, \sigma_P) + \frac{\partial \Lambda_0}{\partial \zeta}(\zeta_{0,P}, \sigma_P)(\zeta_{0,n} - \zeta_{0,P}) + o_P(\|\zeta_{0,n} - \zeta_{0,P}\|) \\ &= \mathbf{D}_{0,\zeta}^T(\zeta_{0,n} - \zeta_{0,P}) + o_P(\|\zeta_{0,n} - \zeta_{0,P}\|).\end{aligned}$$

Putting things together gives

$$\begin{aligned}0 &= \mathbf{D}_{0,\zeta}^T(\zeta_{0,n} - \zeta_{0,P}) + D_{0,\sigma}(\sigma_n - \sigma_P) + \int \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) \\ &\quad + o_P(1/\sqrt{n}) + o_P(\|\zeta_{0,n} - \zeta_{0,P}\|) + o_P(|\sigma_n - \sigma_P|).\end{aligned}$$

Because $D_{0,\sigma} \neq 0$, it follows that

$$\begin{aligned}\sigma_n - \sigma_P &= -D_{0,\sigma}^{-1} \left\{ \mathbf{D}_{0,\zeta}^T(\zeta_{0,n} - \zeta_{0,P}) + \int \Psi_0(\mathbf{s}, \zeta_{0,P}, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) \right\} \\ &\quad + o_P(1/\sqrt{n}) + o_P(\|\zeta_{0,n} - \zeta_{0,P}\|) + o_P(|\sigma_n - \sigma_P|).\end{aligned}$$

This proves the lemma. $\square$

Proof of Theorem 7

Proof. Write $\xi_P = (\beta_{1,P}, \gamma_P) = (\beta_1(P), \gamma(P))$ and $\sigma_P = \sigma(P)$. Similar to (11.62) we decompose as follows

$$\begin{aligned}\mathbf{0} &= \int \Psi_\beta(\mathbf{s}, \xi_n, \sigma_n) dP(\mathbf{s}) + \int \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) \\ &\quad + \int (\Psi_\beta(\mathbf{s}, \xi_n, \sigma_n) - \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P)) d(\mathbb{P}_n - P)(\mathbf{s}).\end{aligned}$$

From (7.3) we have that

$$\Psi_\beta(\mathbf{s}, \xi, \sigma) = u_1 \left(\frac{d}{\sigma} \right) \mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) = u_1(d(\mathbf{y}, \mathbf{X}\beta, \sigma^2 \mathbf{V})) \mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta).$$

Since $\mathbf{V}$ satisfies (V1) we have that $\sigma_n^2 \mathbf{V}(\gamma_n) \rightarrow \sigma_P^2 \mathbf{V}(\gamma_P)$, in probability. Then similar to the proof of equation (96) in Lemma 11.8 in Lopuhaä *et al* [23], it follows

$$\int (\Psi_\beta(\mathbf{s}, \xi_n, \sigma_n) - \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P)) d(\mathbb{P}_n - P)(\mathbf{s}) = o_P(1/\sqrt{n}).$$

This means we can write

$$\mathbf{0} = \Lambda_\beta(\xi_n, \sigma_n) + \int \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}). \quad (11.66)$$

Since $\partial \Lambda_\beta / \partial \beta$ is continuous at (ξ_P, σ_P) and $(\gamma_n, \sigma_n) \rightarrow (\gamma_P, \sigma_P)$, in probability, it follows that

$$\Lambda_\beta(\xi_n, \sigma_n) = \Lambda_\beta(\beta_{1,P}, \gamma_n, \sigma_n) + \left(\frac{\partial \Lambda_\beta}{\partial \beta}(\beta_{1,P}, \gamma_P, \sigma_P) + o_P(1) \right) (\beta_{1,n} - \beta_{1,P}).$$

Because $\beta_{1,P}$ is a point of symmetry and Ψ_β is an odd function of $\mathbf{y} - \mathbf{X}\beta$, it follows that $\Lambda_\beta(\beta_{1,P}, \gamma_n, \sigma_n) = \mathbf{0}$, so that

$$\Lambda_\beta(\xi_n, \sigma_n) = \mathbf{D}_\beta(\beta_{1,n} - \beta_{1,P}) + o(\|\beta_{1,n} - \beta_{1,P}\|).$$

Together with (11.66), we obtain

$$\mathbf{0} = \mathbf{D}_\beta(\beta_n - \beta_{1,P}) + \int \Psi_\beta(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o(\|\beta_{1,n} - \beta_{1,P}\|) + o_P(1/\sqrt{n}).$$

Since ρ_1 satisfies (R2) and (R4), according to Lemma 8 there exist a constant $C_1 > 0$ only depending on P and ρ_1 , such that $\|\Psi_\beta(\mathbf{s}; \boldsymbol{\xi}_P, \sigma_P)\| \leq C_1 \|\mathbf{X}\|$. Since $\mathbb{E}\|\mathbf{X}\|^2 < \infty$, the central limit theorem applies to the second term on the right hand side. From this, we first conclude that $\beta_{1,n} - \beta_{1,P} = O_P(1/\sqrt{n})$. After inserting this in the previous equality and use that $\mathbf{D}_\beta$ is non-singular, this implies

$$\beta_{1,n} - \beta_{1,P} = -\mathbf{D}_\beta^{-1} \int \Psi_\beta(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}).$$

This finishes the proof. $\square$

Lemma 15. Suppose that P satisfies (E) for some $(\beta^*, \theta^*) \in \mathbb{R}^q \times \Theta$ and suppose that $\mathbb{E}\|\mathbf{X}\| < \infty$. Suppose that ρ_0 and $\mathbf{V}$ satisfy (R2), (R4) and (V4), respectively. Let $\zeta_{0,n} = (\beta_{0,n}, \theta_{0,n})$ be the pair of initial estimators and let $\zeta_0 = (\beta_0, \theta_0)$ be the corresponding functional satisfying $(\beta_0(P), \theta_0(P)) = (\beta^*, \theta^*)$, and suppose that $\zeta_{0,n} - \zeta_0(P) = O_P(1/\sqrt{n})$. Let σ_n be a solution of (3.2) and suppose that $\sigma_n \rightarrow \sigma(P)$, in probability, where $\sigma(P)$ is a solution of (3.8). Suppose that $\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_0(c_\sigma \|\mathbf{z}\|)c_\sigma \|\mathbf{z}\|] > 0$, where $c_\sigma = |\boldsymbol{\Sigma}|^{1/(2k)}/\sigma(P)$. Then $\sqrt{n}(\sigma_n - \sigma(P))$ is asymptotically normal with mean zero and variance

$$\frac{\sigma^2(P) \mathbb{E}[(\rho_0(c_\sigma \|\mathbf{z}\|) - b_0)^2]}{(\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_0(c_\sigma \|\mathbf{z}\|)c_\sigma \|\mathbf{z}\|])^2}.$$

Proof. We apply Lemma 14. Consider Λ_0 as defined in (11.22) with Ψ_0 from (11.21). From Lemma 4, we have that Λ_0 is continuously differentiable at $(\zeta_0(P), \sigma(P))$, with derivatives $\mathbf{D}_{0,\zeta} = \mathbf{0}$ and $D_{0,\sigma} = -\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_0(c_\sigma \|\mathbf{z}\|)c_\sigma \|\mathbf{z}\|]/\sigma(P) < 0$, according to Lemma 6. Since $\zeta_{0,n} - \zeta_0(P) = O_P(1/\sqrt{n})$, from Lemma 14, it then follows that

$$\begin{aligned} \sigma_n - \sigma(P) &= -D_{0,\sigma}^{-1} \int \Psi_0(\mathbf{s}, \zeta_0(P), \sigma(P)) d(\mathbb{P}_n - P)(\mathbf{s}) \\ &\quad + o_P(\sigma_n - \sigma(P)) + o_P(1/\sqrt{n}), \end{aligned} \tag{11.67}$$

Since ρ_0 is bounded, the central limit theorem applies to the first term on the right hand side of (11.67). We first conclude that $\sigma_n - \sigma(P) = O_P(1/\sqrt{n})$ and after inserting this in (11.67), we find that $\sqrt{n}(\sigma_n - \sigma(P))$ is asymptotically normal with mean zero and variance $\mathbb{E}[\Psi_0(\mathbf{s}, \zeta_0(P), \sigma(P))^2] / D_{0,\sigma}^2$. Since $(\beta_0(P), \theta_0(P)) = (\beta^*, \theta^*)$ and $\boldsymbol{\Gamma}(\theta_0(P)) = \mathbf{V}(\theta^*)/|\mathbf{V}(\theta^*)|^{1/k} = \boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/k}$, it follows that

$$\mathbb{E}[\Psi_0(\mathbf{s}, \zeta_0(P), \sigma(P))^2] = \mathbb{E}\left[\mathbb{E}\left[\left(\rho_0\left(\frac{d_\Gamma^*}{\sigma(P)}\right) - b_0\right)^2 \middle| \mathbf{X}\right]\right],$$

where d_Γ^* is defined in (11.29). The inner expectation on the right hand side is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$, which has the same distribution as $\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \mathbf{X}\beta^*$, where $\mathbf{z}$ has spherical density $f_{\mathbf{0}, \mathbf{I}_k}$. This implies that the inner expectation on the right hand side is equal to $\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[(\rho_0(c_\sigma \|\mathbf{z}\|) - b_0)^2]$, where $c_\sigma = |\boldsymbol{\Sigma}|^{1/(2k)}/\sigma(P)$. This proves the lemma. $\square$

Proof of Theorem 8

Proof. Write $\boldsymbol{\xi}_P = (\beta_{1,P}, \gamma_P) = (\beta_1(P), \gamma(P))$, $\mathbf{V}_P = \mathbf{V}(\gamma_P)$, and $\sigma_P = \sigma(P)$. We apply Lemma 13. Consider Λ as defined in (8.2) with Ψ as defined in (7.7). From Lemmas 9 and 10, we have that Λ is continuously differentiable at $(\boldsymbol{\xi}_P, \sigma_P)$ with derivative

$$\mathbf{D}_\xi = \frac{\partial \Lambda(\boldsymbol{\xi}_P, \sigma_P)}{\partial \xi} = \begin{pmatrix} \mathbf{D}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\gamma \end{pmatrix},$$

where $\mathbf{D}_\beta$ and $\mathbf{D}_\gamma$ are given in (11.33) and (11.34). According to Lemma 11, we have that $\mathbf{D}_\beta$ and $\mathbf{D}_\gamma$ are non-singular, so that $\mathbf{D}_\xi$ is non-singular. Furthermore, from Lemma 12, we have that $\mathbf{D}_\sigma = \partial \Lambda(\boldsymbol{\xi}_P, \sigma_P)/\partial \sigma = \mathbf{0}$. Since $\sigma_n - \sigma_P = O_P(1/\sqrt{n})$, Lemma 13 yields that

$$\boldsymbol{\xi}_n - \boldsymbol{\xi}_P = -\mathbf{D}_\xi^{-1} \int \Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(\|\boldsymbol{\xi}_n - \boldsymbol{\xi}(P)\|) + o_P(1/\sqrt{n}).$$

According to Lemma 8, there exist constants $C_1, C_2 > 0$ only depending on P and $\sigma(P)$, such that $\|\Psi(\mathbf{s}; \boldsymbol{\xi}_P, \sigma_P)\| \leq C_1 + C_2 \|\mathbf{X}\|$. Since $\mathbb{E}\|\mathbf{X}\|^2 < \infty$, the central limit theorem applies to the second term on the right hand side. From this we first obtain that $\boldsymbol{\xi}_n - \boldsymbol{\xi}_P = O_P(1/\sqrt{n})$. After inserting this in the previous equation, we find that

$$\boldsymbol{\xi}_n - \boldsymbol{\xi}_P = -\mathbf{D}_{\boldsymbol{\xi}}^{-1} \int \Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}). \quad (11.68)$$

It follows that $\sqrt{n}(\boldsymbol{\xi}_n - \boldsymbol{\xi}_P)$ is asymptotically normal with mean zero and variance

$$\mathbf{D}_{\boldsymbol{\xi}}^{-1} \mathbb{E} [\Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T] \mathbf{D}_{\boldsymbol{\xi}}^{-1}. \quad (11.69)$$

Because $\Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T$ is an odd function of $\mathbf{y} - \mathbf{X}\beta_{1,P}$ and $\beta_{1,P} = \beta^*$ is a point of symmetry, it follows that $\mathbb{E}[\Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T] = \mathbf{0}$. Hence, also $\mathbb{E}[\Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T]$ is a block matrix with $\mathbb{E}[\Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T]$ and $\mathbb{E}[\Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T]$ on the main diagonal. We conclude that the limiting variance (11.69) of $\sqrt{n}(\boldsymbol{\xi}_n - \boldsymbol{\xi}_P)$ is a block matrix. This proves that $\sqrt{n}(\beta_{1,n} - \beta_{1,P})$ and $\sqrt{n}(\gamma_n - \gamma_P)$ are asymptotically independent.

Moreover, it also follows that $\sqrt{n}(\beta_{1,n} - \beta_{1,P})$ is asymptotically normal with mean zero and variance

$$\mathbf{D}_{\beta}^{-1} \mathbb{E}[\Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T] \mathbf{D}_{\beta}^{-1}, \quad (11.70)$$

and $\sqrt{n}(\gamma_n - \gamma_P)$ is asymptotically normal with mean zero and variance

$$\mathbf{D}_{\gamma}^{-1} \mathbb{E}[\Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T] \mathbf{D}_{\gamma}^{-1}. \quad (11.71)$$

Because $\mathbf{V}_P = \boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/k}$, we can write

$$\Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) = |\boldsymbol{\Sigma}|^{1/k} u_1 \left(\frac{d^*}{\sigma_P} \right) \mathbf{X}^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\beta^*),$$

where $(d^*)^2 = |\boldsymbol{\Sigma}|^{1/k} (\mathbf{y} - \mathbf{X}\beta^*)^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\beta^*)$ and $u_1(s) = \rho'_1(s)/s$. We find that

$$\begin{aligned} & \mathbb{E} [\Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\beta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T] \\ &= |\boldsymbol{\Sigma}|^{2/k} \mathbb{E} \left[\mathbf{X}^T \mathbb{E} \left[u_1 \left(\frac{d^*}{\sigma_P} \right)^2 \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\beta^*) (\mathbf{y} - \mathbf{X}\beta^*)^T \boldsymbol{\Sigma}^{-1} \middle| \mathbf{X} \right] \mathbf{X} \right]. \end{aligned}$$

The inner expectation on the right hand side is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$, which has the same distribution as $\boldsymbol{\Sigma}^{1/2} \mathbf{z} + \mathbf{X}\beta^*$, where $\mathbf{z}$ has spherical density $f_{\mathbf{0}, \mathbf{I}_k}$. Therefore, similar to the proof of Theorem 6, the inner conditional expectation can be written as

$$\boldsymbol{\Sigma}^{-1/2} \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_{\sigma} \|\mathbf{z}\|)^2 \|\mathbf{z}\|^2] \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\mathbf{u} \mathbf{u}^T] \boldsymbol{\Sigma}^{-1/2} = \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_{\sigma} \|\mathbf{z}\|)^2 \|\mathbf{z}\|^2]}{k} \boldsymbol{\Sigma}^{-1},$$

where $c_{\sigma} = |\boldsymbol{\Sigma}|^{1/(2k)}/\sigma_P$. Together with (11.33) and $u_1(s) = \rho'_1(s)/s$, this implies that the asymptotic variance (11.70) of $\sqrt{n}(\beta_{1,n} - \beta_{1,P})$ is given by

$$\frac{\sigma_P^2}{|\boldsymbol{\Sigma}|^{1/k}} \frac{\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [\rho'_1(c_{\sigma} \|\mathbf{z}\|)^2]}{k \alpha_1^2} (\mathbb{E} [\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}])^{-1},$$

where α_1 is defined in (8.5).

Next, consider the limiting variance (11.71) of $\sqrt{n}(\gamma_n - \gamma(P))$. According to Lemma 11, we have

$$\mathbf{D}_{\gamma}^{-1} = a(\mathbf{E}^T \mathbf{E})^{-1} + b(\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E}(\mathbf{E}^T \mathbf{E})^{-1}, \quad (11.72)$$

where $\mathbf{E} = (\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) \mathbf{L}$, and $a = 1/\omega_1$ and $b = \omega_2(\omega_1(\omega_1 - k\omega_2))$, where ω_1 and ω_2 are given in Lemma 10. The rest of the proof is similar to the (second part of the) proof of Corollary 9.2 in

Lopuhaä *et al* [22]. To make the connection with the proof of Corollary 9.2 in Lopuhaä *et al* [22], note that $\mathbf{V}_P = \Sigma/|\Sigma|^{1/k}$, so that

$$\Psi_\gamma(\mathbf{s}, \xi_P, \sigma_P) = -|\Sigma|^{2/k} \mathbf{L}^T (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(\Psi_{\mathbf{V}}(\mathbf{s}, \xi_P, \sigma_P)),$$

and

$$\begin{aligned} & \mathbb{E} \left[\text{vec}(\Psi_\gamma(\mathbf{s}, \xi_P, \sigma_P)) \text{vec}(\Psi_\gamma(\mathbf{s}, \xi_P, \sigma_P))^T \right] \\ &= |\Sigma|^{4/k} \mathbf{E}^T \mathbb{E} \left[\text{vec}(\Sigma^{-1/2} \Psi_{\mathbf{V}}(\mathbf{s}, \xi_P, \sigma_P) \Sigma^{-1/2}) \text{vec}(\Sigma^{-1/2} \Psi_{\mathbf{V}}(\mathbf{s}, \xi_P, \sigma_P) \Sigma^{-1/2})^T \right] \mathbf{E}, \end{aligned}$$

where $\Psi_{\mathbf{V}}$ is given in (7.8). Furthermore, with $\mathbf{z} = \Sigma^{-1/2}(\mathbf{y} - \mathbf{X}\beta^*)$ and $\mathbf{u} = \mathbf{z}/\|\mathbf{z}\|$, we can write

$$\begin{aligned} \Sigma^{-1/2} \Psi_{\mathbf{V}}(\mathbf{s}, \xi_P, \sigma_P) \Sigma^{-1/2} &= k u_1(c_\sigma \|\mathbf{z}\|) \mathbf{z} \mathbf{z}^T - v_1(c_\sigma \|\mathbf{z}\|) \frac{\sigma_P^2}{|\Sigma|^{1/k}} \Sigma \\ &= k u_1(c_\sigma \|\mathbf{z}\|) \|\mathbf{z}\|^2 \mathbf{u} \mathbf{u}^T - v_1(c_\sigma \|\mathbf{z}\|) \frac{1}{c_\sigma^2} \Sigma \\ &= \frac{1}{c_\sigma^2} \left\{ k u_1(c_\sigma \|\mathbf{z}\|) (c_\sigma \|\mathbf{z}\|)^2 \mathbf{u} \mathbf{u}^T - v_1(c_\sigma \|\mathbf{z}\|) \Sigma \right\}. \end{aligned}$$

Similar to the proof of Corollary 9.2 in Lopuhaä *et al* [22], we obtain

$$\mathbb{E} [\Psi_\gamma(\mathbf{s}, \xi_P, \sigma_P) \Psi_\gamma(\mathbf{s}, \xi_P, \sigma_P)^T] = \sigma_P^4 |\Sigma|^{2/k} \left\{ 2\delta_1 \mathbf{E}^T \mathbf{E} + \delta_2 \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E} \right\},$$

where

$$\begin{aligned} \delta_1 &= \frac{k \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|)^2 (c_\sigma \|\mathbf{z}\|)^4]}{k+2} \\ \delta_2 &= \frac{k \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|)^2 (c_\sigma \|\mathbf{z}\|)^4]}{k+2} - 2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|) v_1(c_\sigma \|\mathbf{z}\|) (c_\sigma \|\mathbf{z}\|)^2] \\ &\quad + \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [v_1(\|\mathbf{z}\|)^2]. \end{aligned}$$

Because $v_1(s) = u_1(s)s^2$, we find that

$$\frac{k}{k+2} u_1(s)^2 s^4 - 2 u_1(s) v_1(s) s^2 + v_1(s)^2 = -\frac{2}{k+2} u_1(s)^2 s^4.$$

This means

$$\begin{aligned} \delta_1 &= \frac{k \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|)^2 (c_\sigma \|\mathbf{z}\|)^4]}{k+2} \\ \delta_2 &= -\frac{2 \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(c_\sigma \|\mathbf{z}\|)^2 (c_\sigma \|\mathbf{z}\|)^4]}{k+2} = -\frac{2}{k} \delta_1. \end{aligned} \tag{11.73}$$

Together with (11.72), as in the proof of Corollary 9.2 in Lopuhaä *et al* [22], it follows that

$$\begin{aligned} & \mathbf{D}_\gamma^{-1} \mathbb{E} [\Psi_\gamma(\mathbf{s}, \xi_P, \sigma_P) \Psi_\gamma(\mathbf{s}, \xi_P, \sigma_P)^T] \mathbf{D}_\gamma^{-1} \\ &= 2\sigma_1 (\mathbf{E}^T \mathbf{E})^{-1} + \sigma_2 (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) \text{vec}(\mathbf{I}_k)^T \mathbf{E} (\mathbf{E}^T \mathbf{E})^{-1}, \end{aligned}$$

where

$$\sigma_1 = \sigma_P^4 |\Sigma|^{2/k} a^2 \delta_1 = \frac{\delta_1}{|\Sigma|^{2/k} \gamma_1^2},$$

and

$$\sigma_2 = \sigma_P^4 |\Sigma|^{2/k} \{2b(2a + kb)\delta_1 + (a + kb)^2 \delta_2\} = -2\sigma_P^4 |\Sigma|^{2/k} a^2 \delta_1 / k = -2\sigma_1 / k.$$

We obtain

$$\begin{aligned}\sigma_1 &= \frac{k\mathbb{E}_{\mathbf{0}, \mathbf{I}_k} [u_1(\|c_\sigma \mathbf{z}\|)^2 (c_\sigma \|\mathbf{z}\|)^4]}{(k+2)\gamma_1^2 |\Sigma|^{2/k}}; \\ \sigma_2 &= -\frac{2\sigma_1}{k}.\end{aligned}\tag{11.74}$$

Since $\mathbf{V}$ has a linear structure, we have $\text{vec}(\Sigma) = \mathbf{L}\boldsymbol{\theta}^*$ and $\mathbf{E}\boldsymbol{\theta}^* = \text{vec}(\mathbf{I}_k)$ and

$$(\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \text{vec}(\mathbf{I}_k) = \boldsymbol{\theta}^*.$$

We find that the limiting variance (11.71) is equal to $2\sigma_1(\mathbf{L}^T(\Sigma^{-1} \otimes \Sigma^{-1})\mathbf{L})^{-1} + \sigma_2 \boldsymbol{\theta}^* (\boldsymbol{\theta}^*)^T$. This finishes the proof. $\square$

Proof of Corollary 5

Proof. Write $\zeta_{0,P} = \zeta_0(P)$, $\xi_P = (\beta_{1,P}, \gamma_P) = (\beta_1(P), \gamma(P))$, $\mathbf{V}_P = \mathbf{V}(\gamma_P)$, and $\sigma_P = \sigma(P)$. Because $\boldsymbol{\theta}_{1,n}$ and $\boldsymbol{\theta}_1(P)$ are solutions of (3.6) and (3.11), respectively, similar to the reasoning in the proof of Corollary 4, we have that

$$\begin{aligned}\boldsymbol{\theta}_{1,n} - \boldsymbol{\theta}_1(P) &= \sigma_n^2 \gamma_n - \sigma_P^2 \gamma_P = \sigma_n^2 (\gamma_n - \gamma_P) + \gamma_P (\sigma_n^2 - \sigma_P^2) \\ &= \sigma_n^2 (\gamma_n - \gamma_P) + \gamma_P (\sigma_n + \sigma_P) (\sigma_n - \sigma_P).\end{aligned}\tag{11.75}$$

We will apply Lemmas 13 and 14. From Lemma 4, we have that Λ_0 is continuously differentiable at $(\zeta_{0,P}, \sigma_P)$ with derivatives $\mathbf{D}_{0,\zeta} = \mathbf{0}$ and $D_{0,\sigma} < 0$, according to Lemma 6. Since $\zeta_{0,n} - \zeta_{0,P} = O_P(1/\sqrt{n})$, it follows from Lemma 14 that

$$\begin{aligned}\sigma_n - \sigma(P) &= -D_{0,\sigma}^{-1} \int \Psi_0(\mathbf{s}, \zeta_0(P), \sigma(P)) d(\mathbb{P}_n - P)(\mathbf{s}) \\ &\quad + o_P(\sigma_n - \sigma_P) + o_P(1/\sqrt{n}).\end{aligned}$$

Since ρ_0 is bounded, the central limit theorem applies to the first term on the right hand side of (11.76) and we find that $\sigma_n - \sigma_P = O_P(1/\sqrt{n})$. Therefore,

$$\sigma_n - \sigma(P) = -D_{0,\sigma}^{-1} \int \Psi_0(\mathbf{s}, \zeta_0(P), \sigma(P)) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}).\tag{11.76}$$

From Lemma 9, we have that Λ is continuous differentiable at (ξ_P, σ_P) , with derivative $\mathbf{D}_\xi = \partial \Lambda(\xi_P, \sigma_P) / \partial \xi$ given in Lemma 10, which is non-singular according to Lemma 11. Furthermore, according to Lemma 12, we have $\mathbf{D}_\sigma = \partial \Lambda(\xi_P, \sigma_P) / \partial \sigma = \mathbf{0}$. Since ρ_1 satisfies (R1)-(R4), such that $u_1(s)$ is of bounded variation, and $\mathbf{V}$ satisfies (V4), we may apply Lemma 13 and obtain

$$\xi_n - \xi_P = -\mathbf{D}_\xi^{-1} \int \Psi(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(\|\xi_n - \xi(P)\|) + o_P(1/\sqrt{n}).$$

As in the proof of Theorem 8, we first obtain that $\xi_n - \xi_P = O_P(1/\sqrt{n})$ and then conclude that

$$\xi_n - \xi_P = -\mathbf{D}_\xi^{-1} \int \Psi(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}).$$

From the block structure of $\mathbf{D}_\xi$ established in Lemma 10, it then follows that

$$\gamma_n - \gamma_P = -\mathbf{D}_\gamma^{-1} \int \Psi_\gamma(\mathbf{s}, \xi_P, \sigma_P) d(\mathbb{P}_n - P)(\mathbf{s}) + o_P(1/\sqrt{n}).\tag{11.77}$$

In particular, this implies that $\gamma_n - \gamma_P = O_P(1/\sqrt{n})$, so that from (11.75) we obtain

$$\sqrt{n}(\boldsymbol{\theta}_{1,n} - \boldsymbol{\theta}_1(P)) = \sigma_P^2 \sqrt{n}(\gamma_n - \gamma_P) + 2\sigma_P \gamma_P \sqrt{n}(\sigma_n - \sigma_P) + o_P(1).$$

From expansions (11.76) and (11.77), we conclude that if we define

$$\Psi_{\theta}(\mathbf{s}, \boldsymbol{\xi}, \sigma) = \sigma_P^2 \mathbf{D}_{\gamma}^{-1} \Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) + 2\sigma_P \gamma_P D_{0,\sigma}^{-1} \Psi_0(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P),$$

then $\sqrt{n}(\boldsymbol{\theta}_{1,n} - \boldsymbol{\theta}_1(P))$ is asymptotically normal with mean zero and variance

$$\mathbb{E} [\Psi_{\theta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\theta}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T].$$

As in the proof of Theorem 8,

$$\Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) = -|\boldsymbol{\Sigma}|^{2/k} \mathbf{L}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \text{vec}(\Psi_{\mathbf{V}}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)),$$

with

$$\Psi_{\mathbf{V}}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) = k u_1 \left(\frac{d^*}{\sigma_P} \right) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)^T - v_1 \left(\frac{d^*}{\sigma_P} \right) \frac{\sigma_P^2}{|\boldsymbol{\Sigma}|^{1/k}} \boldsymbol{\Sigma}.$$

where $(d^*)^2 = |\boldsymbol{\Sigma}|^{1/k} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)$, and from the proof of Lemma 15

$$\Psi_0(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) = \rho_0 \left(\frac{d_{\Gamma}^*}{\sigma_P} \right) - b_0,$$

where $d_{\Gamma}^* = d^*$. This means that

$$\begin{aligned} & \mathbb{E}[\Psi_{\mathbf{V}}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_0(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)] \\ &= \mathbb{E} \left[\mathbb{E} \left[\left(k u_1 \left(\frac{d^*}{\sigma_P} \right) \mathbf{e}^* (\mathbf{e}^*)^T - \frac{\sigma_P^2}{|\boldsymbol{\Sigma}|^{1/k}} v_1 \left(\frac{d^*}{\sigma_P} \right) \boldsymbol{\Sigma} \right) \left(\rho_0 \left(\frac{d^*}{\sigma_P} \right) - b_0 \right) \middle| \mathbf{X} \right] \right], \end{aligned}$$

where $\mathbf{e}^* = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*$. The inner expectation is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$, which has the same distribution as $\boldsymbol{\Sigma}^{1/2} \mathbf{z} + \mathbf{X}\boldsymbol{\beta}^*$, where $\mathbf{z}$ has spherical density $f_{0, \mathbf{I}_k}$. Therefore, if we denote $c_{\sigma} = |\boldsymbol{\Sigma}|^{1/(2k)}/\sigma_P$, then the inner expectation can be written as

$$\begin{aligned} & \mathbb{E}_{0, \mathbf{I}_k} \left[\left(k u_1 (c_{\sigma} \|\mathbf{z}\|) \boldsymbol{\Sigma}^{1/2} \mathbf{z} \mathbf{z}^T \boldsymbol{\Sigma}^{1/2} - v_1 (\|\mathbf{z}\|) \frac{1}{c_{\sigma}^2} \boldsymbol{\Sigma} \right) (\rho_0 (c_{\sigma} \|\mathbf{z}\|) - b_0) \right] \\ &= \frac{1}{c_{\sigma}^2} \mathbb{E}_{0, \mathbf{I}_k} \left[k u_1 (c_{\sigma} \|\mathbf{z}\|) (c_{\sigma} \|\mathbf{z}\|)^2 (\rho_0 (c_{\sigma} \|\mathbf{z}\|) - b_0) \boldsymbol{\Sigma}^{1/2} \mathbf{u} \mathbf{u}^T \boldsymbol{\Sigma}^{1/2} \right] \\ &\quad - \frac{1}{c_{\sigma}^2} \mathbb{E}_{0, \mathbf{I}_k} [v_1 (c_{\sigma} \|\mathbf{z}\|) (\rho_0 (c_{\sigma} \|\mathbf{z}\|) - b_0)] \boldsymbol{\Sigma} \\ &= \frac{k}{c_{\sigma}^2} \mathbb{E}_{0, \mathbf{I}_k} [v_1 (c_{\sigma} \|\mathbf{z}\|) (\rho_0 (c_{\sigma} \|\mathbf{z}\|) - b_0)] \boldsymbol{\Sigma}^{1/2} \mathbb{E}_{0, \mathbf{I}_k} [\mathbf{u} \mathbf{u}^T] \boldsymbol{\Sigma}^{1/2} \\ &\quad - \frac{1}{c_{\sigma}^2} \mathbb{E}_{0, \mathbf{I}_k} [v_1 (c_{\sigma} \|\mathbf{z}\|) (\rho_0 (c_{\sigma} \|\mathbf{z}\|) - b_0)] \boldsymbol{\Sigma} \\ &= \frac{k}{c_{\sigma}^2} \mathbb{E}_{0, \mathbf{I}_k} [v_1 (c_{\sigma} \|\mathbf{z}\|) (\rho_0 (c_{\sigma} \|\mathbf{z}\|) - b_0)] \frac{1}{k} \boldsymbol{\Sigma} \\ &\quad - \frac{1}{c_{\sigma}^2} \mathbb{E}_{0, \mathbf{I}_k} [v_1 (c_{\sigma} \|\mathbf{z}\|) (\rho_0 (c_{\sigma} \|\mathbf{z}\|) - b_0)] \boldsymbol{\Sigma} = \mathbf{0}. \end{aligned}$$

We find that

$$\begin{aligned} \mathbb{E} [\Psi_{\theta}(\mathbf{s}, \boldsymbol{\xi}, \sigma) \Psi_{\theta}(\mathbf{s}, \boldsymbol{\xi}, \sigma)^T] &= \sigma_P^4 \mathbf{D}_{\gamma}^{-1} \mathbb{E} [\Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P) \Psi_{\gamma}(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^T] \mathbf{D}_{\gamma}^{-1} \\ &\quad + 4\sigma_P^2 D_{0,\sigma}^{-1} \mathbb{E} [\Psi_0(\mathbf{s}, \boldsymbol{\xi}_P, \sigma_P)^2] D_{0,\sigma}^{-1} \gamma_P \gamma_P^T, \end{aligned}$$

which is a linear combination of the asymptotic variances of $\sqrt{n}(\gamma_n - \gamma_P)$ and $\sqrt{n}(\sigma_n - \sigma_P)$:

$$\begin{aligned} & \mathbb{E} [\Psi_{\theta}(\mathbf{s}, \boldsymbol{\xi}, \sigma) \Psi_{\theta}(\mathbf{s}, \boldsymbol{\xi}, \sigma)^T] \\ &= \sigma_P^4 \text{AVAR}(\sqrt{n}(\gamma_n - \gamma_P)) + 4\sigma_P^2 \text{AVAR}(\sqrt{n}(\sigma_n - \sigma_P)) \gamma_P \gamma_P^T. \end{aligned}$$

Since $\mathbf{V}$ has a linear structure, similar to the proof of Corollary 4, from (11.55) we have $\gamma_P = \boldsymbol{\theta}^*/|\boldsymbol{\Sigma}|^{1/k}$. Then, from Theorem 8 and Lemma 15, we find that the right hand side is equal to

$$\frac{2\sigma_1}{c_\sigma^2} \left\{ \left(\mathbf{L}^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{L} \right)^{-1} - \frac{1}{k} \boldsymbol{\theta}^* (\boldsymbol{\theta}^*)^T \right\} + \frac{4}{c_\sigma^2} \frac{\mathbb{E}[(\rho_0(c_\sigma \|\mathbf{z}\|) - b_0)^2]}{(\mathbb{E}_{\mathbf{0}, \mathbf{I}_k}[\rho'_0(c_\sigma \|\mathbf{z}\|) c_\sigma \|\mathbf{z}\|])^2} \boldsymbol{\theta}^* (\boldsymbol{\theta}^*)^T.$$

This finishes the proof. $\square$

11.8 Limiting distributions of covariance estimators

We provide some details about the limiting variances of the covariance MM-estimators in Examples 1, 2, and 3 for the situation where the distribution P satisfies (E).

Example 1 (Linear Mixed Effects model). *For linear mixed effects models, nothing seems to be available about the limiting distribution of covariance MM-estimators. For model (2.3), the limiting distribution of the variance component MM-estimator now follows from Corollary 5 and the limiting distributions of the covariance MM-estimator and the corresponding shape component follow from (9.4) and (9.5), respectively. This implies that $\sqrt{n}(\boldsymbol{\theta}_{1,n} - \boldsymbol{\theta}_1(P))$ is asymptotically normal with mean zero and variance*

$$\frac{2\sigma_1}{c_\sigma^2} \mathbf{Q}^{-1} + \left(-\frac{2\sigma_1}{kc_\sigma^2} + \sigma_3 \right) \boldsymbol{\theta}^* (\boldsymbol{\theta}^*)^T,$$

where σ_1 and σ_3 are defined in (9.2) and (9.3), and where $\mathbf{Q}$ is the matrix with entries given in (11.58). From (9.4) we find that $\sqrt{n}(\text{vec}(\mathbf{V}(\boldsymbol{\theta}_{1,n})) - \text{vec}(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and variance

$$\frac{2\sigma_1}{c_\sigma^2} \mathbf{L} \mathbf{Q}^{-1} \mathbf{L}^T + \left(-\frac{2\sigma_1}{kc_\sigma^2} + \sigma_3 \right) \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})^T,$$

where $\mathbf{L}$ is given by (11.57). For the shape component $\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \mathbf{V}(\boldsymbol{\theta})/|\mathbf{V}(\boldsymbol{\theta})|^{1/k}$, from (9.5) we obtain that $\sqrt{n}(\text{vec}(\boldsymbol{\Gamma}(\boldsymbol{\theta}_{1,n})) - \text{vec}(\boldsymbol{\Gamma}(\boldsymbol{\theta}_1(P))))$ is asymptotically normal with mean zero and variance

$$\frac{2\sigma_1}{c_\sigma^2 |\boldsymbol{\Sigma}|^{2/k}} \left\{ \mathbf{L} \mathbf{Q}^{-1} \mathbf{L}^T - \frac{1}{k} \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})^T \right\}.$$

When $c_\sigma = 1$, the limiting distribution of the covariance shape MM-estimator in the linear mixed effects model (2.3) coincides with that of the shape component corresponding to the covariance S -estimator defined with ρ_1 , see Lopuhaä et al [22], and similarly for the direction component of the variance component MM-estimator.

Example 2 (Multivariate Linear Regression). *For the multivariate linear regression model (2.4), Kudraszow and Maronna [15] do not consider the limiting distribution of the covariance MM-estimator. In this model, the matrix $\mathbf{L}$ is equal to the duplication matrix $\mathcal{D}_k$, which satisfies (11.59). The limiting distribution of the covariance MM-estimator and the corresponding shape component follow from (9.4) and (9.5), respectively. By using (11.59), the limiting variance of the covariance MM-estimator becomes*

$$\frac{\sigma_1}{c_\sigma^2} (\mathbf{I}_{k^2} + \mathbf{K}_{k,k}) (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + \left(-\frac{2\sigma_1}{kc_\sigma^2} + \sigma_3 \right) \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})^T, \quad (11.78)$$

whereas the covariance shape estimator has limiting variance

$$\frac{\sigma_1}{c_\sigma^2 |\boldsymbol{\Sigma}|^{2/k}} \left\{ (\mathbf{I}_{k^2} + \mathbf{K}_{k,k}) (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) - \frac{2}{k} \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})^T \right\}, \quad (11.79)$$

where σ_1 and σ_3 are defined in (9.2) and (9.3). When $c_\sigma = 1$, the limit behavior of the covariance shape MM-estimator coincides with that of the shape component of the covariance S -estimator defined with ρ_1 in the multivariate linear regression model.

Example 3 (Multivariate Location and Scatter). *For the multivariate location-scatter model, we also have $\mathbf{L} = \mathcal{D}_k$. Since this model is a special case of the multivariate linear regression model (2.4) by taking $\mathbf{x}_i = 1$ and $\mathbf{B}^T = \boldsymbol{\mu}$, the limiting distributions of the covariance MM-estimator and the corresponding shape component are the same as that of their counterparts in the multivariate linear regression model and the limiting variances have the same expressions as (11.78) and (11.79). When $c_\sigma = 1$, the behavior of the covariance shape MM-estimator coincides with that of the covariance shape S-estimator defined with ρ_1 . This was already observed by Salibián-Barrera et al [31], whose formula (9) matches with the expression in (11.79) with $c_\sigma = 1$. Finally, also here there is a connection with the CM-estimators considered in Kent and Tyler [14], whose limiting distribution depends on a parameter λ_0 . By using (11.59), it can be seen that the limiting distribution of $\sqrt{n}(\text{vec}(\mathbf{V}(\boldsymbol{\theta}_{1,n})) - \text{vec}(\boldsymbol{\Sigma}))$ is similar to that of the covariance CM-estimator for the particular case that $\lambda_0 = \lambda_L$ (see Kent and Tyler [14] for details), and that they both coincide when $c_\sigma = 1/\sqrt{\lambda_0}$ and $\rho_0 = \rho_1$.*