Asymptotics of estimators for structured covariance matrices

Lopuhaä, Hendrik Paul

DOI
10.1016/j.jmva.2025.105443

Publication date
2025

Document Version
Final published version

Published in
Journal of Multivariate Analysis

Journal of Multivariate Analysis 208 (2025) 105443

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Asymptotics of estimators for structured covariance matrices

Hendrik Paul Lopuhaä

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

ARTICLE INFO

AMS 2020 subject classifications:
primary 62E20
62H10
62H12
secondary 62F35
62J05

Keywords:
Asymptotic variance
Influence function
Linear model with structured covariance
S-estimators

ABSTRACT

We show that the limiting variance of a sequence of estimators for a structured covariance matrix has a general form, that for linear covariance structures appears as the variance of a scaled projection of a random matrix that is of radial type, and a similar result is obtained for the corresponding sequence of estimators for the vector of variance components. These results are illustrated by the limiting behavior of estimators for a differentiable covariance structure in a variety of multivariate statistical models. We also derive a characterization for the influence function of corresponding functionals. Furthermore, we derive the limiting distribution and influence function of scale invariant mappings of such estimators and their corresponding functionals. As a consequence, the asymptotic relative efficiency of different estimators for the shape component of a structured covariance matrix can be compared by means of a single scalar and the gross error sensitivity of the corresponding influence functions can be compared by means of a single index. Similar results are obtained for estimators of the normalized vector of variance components. We apply our results to investigate how the efficiency, gross error sensitivity, and breakdown point of S-estimators for the normalized variance components are affected simultaneously by varying their cutoff value.

1. Introduction

Covariance matrices describe the relationships and variability between different variables in a dataset. When there is a known structure or pattern in these relationships, structured covariance matrices can be estimated to capture and represent that structure. The use of structured covariance matrices is a valuable tool for modeling the underlying patterns and dependencies in multivariate data. It provides a more nuanced understanding of the relationships between variables, especially in scenarios where variables exhibit specific structures or patterns of correlation. Structured covariance matrices are commonly used in the analysis of repeated measures, longitudinal data, and multivariate data with a known underlying structure. They are particularly useful when there are dependencies or correlations among different measurements or variables and are widely used in various fields, including biology, medicine, psychology, and social sciences.

When a covariance matrix is unstructured and can be any positive definite symmetric matrix $\boldsymbol{\Sigma}$, then the limiting behavior of covariance estimators $\mathbf{V}_n$ for $\boldsymbol{\Sigma}$ is well understood. For example, if $\mathbf{V}_n$ is based on a sample $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^k$ from a distribution with an elliptically contoured density $|\boldsymbol{\Sigma}|^{-1/2} g((\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}))$, then typically $\sqrt{n}(\mathbf{V}_n - \boldsymbol{\Sigma})$ converges in distribution to a random matrix $\mathbf{N}$ that has a multivariate normal distribution with mean zero and variance

$$\mathrm{var}\{\mathrm{vec}(\mathbf{N})\} = \sigma_1(\mathbf{I}_{k^2} + \mathbf{K}_{k,k})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + \sigma_2 \mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top, \tag{1}$$

for some $\sigma_1 \geq 0$ and $\sigma_2 \geq -2\sigma_1/k$, where $\otimes$ denotes the Kronecker product, $\mathbf{K}_{k,k}$ is the commutation matrix, and vec is the operator that stacks the columns of a matrix. This form of limiting variance appears for many covariance estimators. Tyler [27] gives several

E-mail address: h.p.lopuhaa@tudelft.nl.

https://doi.org/10.1016/j.jmva.2025.105443
Received 1 July 2024; Received in revised form 14 March 2025; Accepted 17 March 2025

Available online 24 March 2025

examples, including the sample covariance matrix, and nicely explains that this general form will always appear when $\mathbf{N}$ is of radial type with respect to $\boldsymbol{\Sigma}$.

The situation becomes different, when estimating a structured covariance matrix $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta})$, where $\mathbf{V}(\cdot)$ is a known covariance structure depending on a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_\ell)^\top$ of unknown variance components. Asymptotic results for the maximum likelihood estimator of variance components in linear models with Gaussian errors having a structured covariance matrix $\mathbf{V}(\boldsymbol{\theta})$, can be found in Hartley and Rao [8], Miller [23], and Mardia and Marshall [21]. When scaled appropriately, the maximum likelihood estimator $\boldsymbol{\theta}_n$ is shown to be asymptotically normal with mean $\boldsymbol{\theta}$ and variance $\mathbf{J}^{-1}$, where $\mathbf{J}_{ij} = \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{L}_i\boldsymbol{\Sigma}^{-1}\mathbf{L}_j)/2$, for $i, j \in \{1, \ldots, \ell\}$, with $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta})$ and $\mathbf{L}_i = \partial\mathbf{V}(\boldsymbol{\theta})/\partial\theta_i$. By employing the vec-notation, the limiting variance of $\boldsymbol{\theta}_n$ can be expressed as

$$2\left(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{L}\right)^{-1},$$

where $\mathbf{L}$ is the matrix with columns $\mathrm{vec}(\mathbf{L}_1), \ldots, \mathrm{vec}(\mathbf{L}_\ell)$. According to the delta method the limiting variance of $\mathrm{vec}(\mathbf{V}(\boldsymbol{\theta}_n))$ is then given by

$$2\mathbf{L}\left(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{L}\right)^{-1}\mathbf{L}^\top.$$

Similar results have been obtained in Lopuhaä et al. [17] for the class of S-estimators based on observations that follow a linear model with a structured covariance $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta})$, where $\mathbf{V}$ is a linear function of $\boldsymbol{\theta}$. Under appropriate conditions, it holds that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta})$ is asymptotically normal with mean zero and variance

$$2\sigma_1\left(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1} + \sigma_2\boldsymbol{\theta}\boldsymbol{\theta}^\top, \tag{2}$$

and $\sqrt{n}(\mathbf{V}(\boldsymbol{\theta}_n) - \boldsymbol{\Sigma})$ converges in distribution to a random matrix $\mathbf{M}$, that has a multivariate normal distribution with mean zero and variance

$$\mathrm{var}\{\mathrm{vec}(\mathbf{M})\} = 2\sigma_1\mathbf{L}\left(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1}\mathbf{L}^\top + \sigma_2\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top. \tag{3}$$

One of the objective of this paper is to show that this type of general form will always appear when $\mathbf{M}$ is a scaled projection on the column space of $\mathbf{L}$, of a random matrix that is of radial type with respect to $\boldsymbol{\Sigma}$. Moreover, we provide several examples of covariance estimators that exhibit this kind of limiting behavior.

Another objective concerns the asymptotic behavior of estimators for scale invariant mappings H of positive definite symmetric matrices. For affine equivariant covariance estimators $\mathbf{V}_n$ with asymptotic variance (1), Tyler [28] shows that $H(\mathbf{V}_n)$ has an asymptotic variance that only depends on the scalar σ_1. When dealing with a structured covariance matrix, the covariance estimators are typically not affine equivariant and have asymptotic variance (3). The second objective of this paper is to show that Tyler's result for affine equivariant covariance estimators, remains true for estimators of a structured covariance matrix. Moreover, we will establish a similar result for scale invariant mappings $H(\boldsymbol{\theta}_n)$ of estimators for the vector of variance components.

An example of a scale invariant mapping is the shape component $\mathbf{V}/|\mathbf{V}|^{1/k}$. A consequence of our results is that the asymptotic relative efficiency of estimators of the shape of a structured covariance can be compared simply by comparing the corresponding values for σ_1. For affine equivariant covariance estimators, this was already observed by Kent and Tyler [12] and Salibián et al. [25]. Similar properties will be shown to hold for the direction component $\boldsymbol{\theta}/\|\boldsymbol{\theta}\|$ corresponding to the vector of variance components.

A final objective of this paper concerns the influence function of structured covariance functionals. For affine equivariant covariance functionals, Croux and Haesbroeck [5] show that the influence function at the multivariate normal is characterized by two real-valued functions. Structured covariance functionals, however, are not necessarily affine equivariant. We will show that such a characterization remains valid for structured covariance functionals at any elliptically contoured distribution, and similarly for the variance components functional. A nice consequence is that the influence function of scale invariant mappings H of a structured covariance functional $\mathbf{V}(\boldsymbol{\theta}(\cdot))$ or of $\boldsymbol{\theta}(\cdot)$ itself, is characterized by a single real-valued function. As such, the gross-error-sensitivity (GES) is proportional to a single index, which can be used to compare the GES of different shape functionals or different direction functionals. Kent and Tyler [12] already observed such a property for the shape component of affine equivariant covariance functionals, see also Salibián et al. [25].

Except that our results have a merit of their own, they also enable the construction of MM-estimators with auxiliary scale in linear mixed effects models and other linear models with structured covariances. These estimators inherit the robustness of S-estimators considered in Lopuhaä et al. [17] and, in contrast to the simpler version considered in Lopuhaä [16], improve both the efficiency of the estimator of the fixed effects as well as the efficiency of the estimator of the covariance shape component and of the direction of the vector of variance components. Investigation of this version of MM-estimators will be postponed to a future manuscript, in which we will extend similar results that are already available for unstructured covariances in the multivariate location-scale model, see Tatsuoka and Tyler [26] or Salibián-Barrera et al. [25], and in the multivariate regression model, see Kudraszow and Maronna [13].

The paper is organized as follows. In Section 2 we show that the general forms of (2) and (3) can be derived solely using a scaled projection of a random matrix that is of radial type. In Section 3 we investigate the limiting behavior of estimators of a differentiable covariance structure in a variety of multivariate models. In the special case of a linear covariance structure, we establish that these estimators asymptotically behave the same as a scaled projection of a sequence of affine equivariant covariance estimators that are asymptotically of radial type. In Section 4 we derive the limiting distribution of scale invariant mappings of estimators of a linear covariance structure that are asymptotically normal, and similarly for scale invariant mappings of estimators of the vector of variance components. In Section 5 we derive a characterization for the influence function of structured covariance functionals and the corresponding functional of variance components, and of scale invariant mappings thereof. In Section 6 we apply our results to

investigate how the efficiency, GES, and breakdown point of S-estimators of the variance components are affected simultaneously, when we vary the cut-off value of the rho-function that defines the S-estimator. All proofs are postponed to an Appendix A at the end of the paper.

2. Projection of a random matrix of radial type

A random matrix $\mathbf{R}$ is said to be of *radial type*, if for any orthogonal matrix $\mathbf{O}$, the distribution of $\mathbf{ORO}^\top$ is the same as that of $\mathbf{R}$. The covariance structure of random matrices with a radial distribution was first given by Mallows [20] in index form. Tyler [27] gave the covariance structure in matrix form and provided necessary conditions on its parameters. A random matrix $\mathbf{N}$ is said to be of radial type with respect to the positive definite symmetric matrix $\boldsymbol{\Sigma}$, if $\boldsymbol{\Sigma}^{-1/2}\mathbf{N}\boldsymbol{\Sigma}^{-1/2}$ has a radial distribution. If the first two moments of $\mathbf{N}$ exist, then according to Corollary 1 in Tyler [27], the variance of $\mathbf{N}$ is given by (1).

Consider a $k \times k$ structured covariance matrix $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta})$, where $\mathbf{V}$ is a known covariance structure that is a function of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_\ell)^\top$, a vector of unknown variance components, which is an element of the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^\ell$. Define the $k^2 \times \ell$ matrix

$$\mathbf{L} = \big[\mathrm{vec}(\mathbf{L}_1) \quad \cdots \quad \mathrm{vec}(\mathbf{L}_\ell)\big], \quad \mathbf{L}_j = \partial \mathbf{V}/\partial\theta_j, \quad j \in \{1, \ldots, \ell\}. \tag{4}$$

A special case is when $\mathbf{V}$ is linear, that is

$$\boldsymbol{\Sigma} = \theta_1\mathbf{L}_1 + \cdots + \theta_\ell\mathbf{L}_\ell. \tag{5}$$

In this case we can write $\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}$. Furthermore, let $\boldsymbol{\Pi}_L$ be the projection of a vector $\mathbf{x} \in \mathbb{R}^{k^2}$ on the column space of $\mathbf{L}$, re-scaled by $\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}$, that is

$$\boldsymbol{\Pi}_L\mathbf{x} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\mathrm{argmin}}\,(\mathbf{x} - \mathbf{L}\boldsymbol{\theta})^\top(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{x} - \mathbf{L}\boldsymbol{\theta}). \tag{6}$$

We then have the following theorem.

Theorem 1. *Let $\mathbf{N}$ be a random matrix that is of radial type with respect to a positive definite symmetric matrix $\boldsymbol{\Sigma}$. Suppose that $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta})$, for some $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$, and that $\mathbf{V}$ is differentiable such that $\mathbf{L}$, as defined in (4), is of full column rank. Let $\boldsymbol{\Pi}_L$ be defined in (6) and define the random matrix $\mathbf{M}$ by $\mathrm{vec}(\mathbf{M}) = \boldsymbol{\Pi}_L\mathrm{vec}(\mathbf{N})$.*

(i) *If the first two moments of $\mathbf{N}$ exist, then there exist constants η, σ_1 and σ_2 with $\sigma_1 \geq 0$ and $\sigma_2 \geq -2\sigma_1/k$, such that $\mathbb{E}[\mathrm{vec}(\mathbf{M})] = \eta\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma})$, and*

$$\mathit{var}(\mathrm{vec}(\mathbf{M})) = 2\sigma_1\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top + \sigma_2\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top\boldsymbol{\Pi}_L^\top. \tag{7}$$

(ii) *Let $\boldsymbol{\theta}_L \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$ be such that $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_L$. If $\mathbf{T} \in \mathbb{R}^\ell$ is the random vector, such that $\mathrm{vec}(\mathbf{M}) = \mathbf{LT}$, then $\mathbb{E}[\mathbf{T}] = \eta\boldsymbol{\theta}_L$ and*

$$\mathit{var}(\mathbf{T}) = 2\sigma_1\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1} + \sigma_2\boldsymbol{\theta}_L\boldsymbol{\theta}_L^\top. \tag{8}$$

Note that the constants η, σ_1 and σ_2 have nothing to do with the projection $\boldsymbol{\Pi}_L$, but are inherited from the variance (1) of the radial random matrix $\mathbf{N}$. Their existence is guaranteed by Corollary 1 in Tyler [27]. When $\mathbf{V}$ is linear, the expressions in Theorem 1 simplify to the ones in (2) and (3).

Corollary 1. *Let $\mathbf{N}$ and $\mathbf{M}$ be defined in Theorem 1. Under the assumptions of Theorem 1, assume in addition that $\mathbf{V}$ satisfies (5).*

(i) *If the first two moments of $\mathbf{N}$ exist, then there exist constants η, σ_1 and σ_2 with $\sigma_1 \geq 0$ and $\sigma_2 \geq -2\sigma_1/k$, such that $\mathbb{E}[\mathrm{vec}(\mathbf{M})] = \eta\mathrm{vec}(\boldsymbol{\Sigma})$ and $\mathit{var}(\mathrm{vec}(\mathbf{M}))$ is given by (3).*

(ii) *If $\mathbf{T} \in \mathbb{R}^\ell$ is the random vector, such that $\mathrm{vec}(\mathbf{M}) = \mathbf{LT}$, then $\mathbb{E}[\mathbf{T}] = \eta\boldsymbol{\theta}$ and $\mathit{var}(\mathbf{T})$ is given by (2).*

Examples of multivariate statistical models with a structured covariance matrix are linear mixed effects models. But also linear models with errors generated by some autoregressive time series may correspond to a structured covariance matrix. When $\boldsymbol{\Sigma}$ is unstructured and can be any positive definite symmetric covariance matrix, it can also be seen as a linear covariance structure $\mathbf{V}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \mathrm{vech}(\boldsymbol{\Sigma})$, with

$$\mathrm{vech}(\mathbf{A}) = (a_{11}, \ldots, a_{k1}, a_{22}, \ldots, a_{kk})^\top, \tag{9}$$

is the unique $k(k+1)/2$-vector that stacks the columns of the lower triangle elements of a symmetric matrix $\mathbf{A}$. The matrix $\mathbf{L} = \partial\mathrm{vec}(\mathbf{V})/\partial\boldsymbol{\theta}^\top$ 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 $\mathcal{D}_k\mathrm{vech}(\mathbf{A}) = \mathrm{vec}(\mathbf{A})$ and $(\mathcal{D}_k^\top\mathcal{D}_k)^{-1}\mathcal{D}_k^\top\mathrm{vec}(\mathbf{A}) = \mathrm{vech}(\mathbf{A})$. Moreover, from the properties of $\mathcal{D}_k$ (see, e.g., Magnus and Neudecker [19, Ch. 3, Sec. 8]), it follows that

$$\mathcal{D}_k\left(\mathcal{D}_k^\top\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathcal{D}_k\right)^{-1}\mathcal{D}_k^\top = \frac{1}{2}\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}). \tag{10}$$

In this case, the expression (3) with $\mathbf{L} = \mathcal{D}_k$ coincides with the expression (1).

3. Projections of estimators of radial type

A sequence $\{\mathbf{N}_n\}$ of $k \times k$ symmetric estimators for $\boldsymbol{\Sigma}$ is said to be *asymptotically of radial type* if there exists a sequence of real numbers a_n increasing to infinity, such that $a_n(\mathbf{N}_n - \boldsymbol{\Sigma}) \to \mathbf{N}$ in distribution with $\mathbf{N}$ being of radial type with respect to $\boldsymbol{\Sigma}$, see Tyler [27]. In a large class of multivariate statistical models for estimators $\mathbf{V}_n$ of a linearly structured covariance matrix, it turns out that the limiting behavior of $\text{vec}(\mathbf{V}_n)$ is the same as that of the projection $\boldsymbol{\Pi}_L \text{vec}(\mathbf{N}_n)$ of a random matrix $\mathbf{N}_n$ that is asymptotically of radial type with respect to $\boldsymbol{\Sigma}$, where $\boldsymbol{\Pi}_L$ is defined in (6). When $\mathbf{V}$ is nonlinear, the behavior is similar, but slightly different. We illustrate things in the following linear model with a structured covariance.

Consider independent observations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in \mathbb{R}^k \times \mathbb{R}^{kq}$ with distribution P, where $\mathbf{s}_i = (\mathbf{y}_i, \mathbf{X}_i)$, $i \in \{1, \ldots, n\}$, for which we assume the following model

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i, \quad i \in \{1, \ldots, n\}, \tag{11}$$

where $\mathbf{y}_i \in \mathbb{R}^k$, $\boldsymbol{\beta} \in \mathbb{R}^q$ is an unknown parameter vector, $\mathbf{X}_i \in \mathbb{R}^{k \times q}$ is a known design matrix, and $\mathbf{u}_i \in \mathbb{R}^k$ 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. Suppose that the distribution P for random variable $\mathbf{s} = (\mathbf{y}, \mathbf{X})$ is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} g\left((\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right), \tag{12}$$

where $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta}_0)$, for some vector $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$ of variance components. We assume that

A1: $\mathbf{V}$ is identifiable in the sense that, if $\mathbf{V}(\boldsymbol{\theta}_1) = \mathbf{V}(\boldsymbol{\theta}_2)$, then $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$;
A2: $\mathbf{V}$ is twice continuously differentiable.

This setup includes several multivariate statistical models of interest. An important case of interest is the (balanced) linear mixed effects model. An example of such a model is

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

where the $\mathbf{Z}_j$'s are known $k \times g_j$ design matrices and the $\gamma_{ij} \in \mathbb{R}^{g_j}$ are independent mean zero random variables with covariance matrix $\sigma_j^2 \mathbf{I}_{g_j}$, for $j \in \{1, \ldots, r\}$, and $\boldsymbol{\epsilon}_i$ has covariance matrix $\sigma_0^2 \mathbf{I}_k$. This leads to a linear covariance structure $\mathbf{V}(\boldsymbol{\theta}) = \sum_{j=1}^r \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j^\top + \sigma_0^2 \mathbf{I}_k$ and $\boldsymbol{\theta} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_r^2)^\top$.

Structured covariance matrices may also arise in linear models (11) with $\mathbf{u}_1, \ldots, \mathbf{u}_n$ generated from a time series. One example is an autoregressive process of order one, which leads to $\mathbf{V}$ with elements

$$v_{st} = \sigma^2 \rho^{|s-t|}, \quad s, t \in \{1, \ldots, k\}.$$

Another basic example is the equicorrelated model, which corresponds to

$$v_{st} = \begin{cases} \sigma^2, & s = t; \\ \sigma^2 \rho, & \text{otherwise}, \end{cases}$$

for $s, t \in \{1, \ldots, k\}$. Both models are an example of a nonlinear covariance structure, where the vector of unknown covariance parameters is $\boldsymbol{\theta} = (\sigma^2, \rho)^\top \in (0, \infty) \times [-1, 1]$. A general stationary process leads to

$$v_{st} = \theta_{|s-t|+1}, \quad s, t \in \{1, \ldots, k\},$$

which is a linear covariance structure with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top \in \mathbb{R}^k$, where $\theta_{|s-t|+1}$ represents the autocovariance over lag $|s - t|$.

Note that current setup also allows models with an unstructured covariance matrix, such as the multivariate location-scale model or the multivariate regression model. See, e.g., Jennrich and Schluchter [11] or Fitzmaurice et al. [6], for different possible covariance structures, and Lopuhaä et al. [17], who provide a uniform treatment of S-estimators in these models.

Estimators $\boldsymbol{\xi}_n = (\boldsymbol{\beta}_n, \boldsymbol{\theta}_n)$ for $\boldsymbol{\xi}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ are typically solutions of estimating equations of the following type

$$\int \boldsymbol{\Psi}(\mathbf{s}, \boldsymbol{\xi}) \, \mathrm{d}\mathbb{P}_n(\mathbf{s}) = \mathbf{0}, \tag{13}$$

where $\mathbb{P}_n$ denotes the empirical measure corresponding to $\mathbf{s}_1, \ldots, \mathbf{s}_n$, and where $\boldsymbol{\Psi} = (\boldsymbol{\Psi}_{\boldsymbol{\beta}}, \boldsymbol{\Psi}_{\boldsymbol{\theta}})$, with

$$\begin{aligned} \boldsymbol{\Psi}_{\boldsymbol{\beta}}(\mathbf{s}, \boldsymbol{\xi}) &= w_1(d) \mathbf{X}^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \boldsymbol{\Psi}_{\boldsymbol{\theta}}(\mathbf{s}, \boldsymbol{\xi}) &= \mathbf{L}^\top (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \text{vec}\left\{ w_2(d) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top - w_3(d) \mathbf{V} \right\}, \end{aligned} \tag{14}$$

where $d^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$, $\mathbf{L}$ is defined in (4), and where we write $\mathbf{V}$ for $\mathbf{V}(\boldsymbol{\theta})$. We give some examples below. Furthermore, typically $\boldsymbol{\xi}_n$ will then converge to a solution of the corresponding population equation

$$\int \boldsymbol{\Psi}(\mathbf{s}, \boldsymbol{\xi}) \, \mathrm{d}P(\mathbf{s}) = \mathbf{0}. \tag{15}$$

Let $\mathbf{V}_n = \mathbf{V}(\boldsymbol{\theta}_n)$. From the estimating Eqs. (13) for $\boldsymbol{\xi}_n$, we will establish that $\text{vec}(\mathbf{V}_n)$ is asymptotically normal. To this end, we require the following conditions:

C1: $w_i(s)$ is of bounded variation and continuously differentiable, for $i \in \{1,2,3\}$;
C2: $w_1'(s)s^2$, $w_2'(s)s^3$, and $w_3'(s)s^2$ are bounded.

Furthermore, the derivation of the limiting distribution uses a Taylor expansion of the left-hand side of (15) around $\boldsymbol{\xi}_0$. To guarantee the existence of the inverse of the derivative involved, we need conditions on the following constants

$$\gamma_1 = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3 + k(k+2)w_3(\|\mathbf{z}\|)\right]}{k(k+2)}, \quad \gamma_2 = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[(k+2)w_3'(\|\mathbf{z}\|)\|\mathbf{z}\| - w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3\right]}{2k(k+2)} \tag{16}$$
$$\pi_L = \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^\top \boldsymbol{\Pi}_L \mathrm{vec}(\boldsymbol{\Sigma}),$$

where $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}$ denotes the expectation with respect to density (12) with parameters $(\boldsymbol{\mu},\boldsymbol{\Sigma}) = (\mathbf{0},\mathbf{I}_k)$ and $\boldsymbol{\Pi}_L$ is defined in (6). To ensure the existence of the scalars σ_1 and σ_2 in Theorem 2, we require

C3: $\gamma_1 \neq 0$ and $\gamma_1 - \gamma_2\pi_L \neq 0$.

Maronna [22] and Tyler [27] consider M-estimators for multivariate location and covariance. Estimating equations for these estimators would correspond to Ψ_θ without the factor $\mathbf{L}^\top(\mathbf{V}^{-1} \otimes \mathbf{V}^{-1})$ (see Example 2 below) and $w_3 = 1$. Moreover, they assume that w_2' is non-negative, which obviously implies (C3).

Theorem 2. *Let P be a distribution for random variable $\mathbf{s} = (\mathbf{y},\mathbf{X})$, such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density (12), with parameters $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma} = \mathbf{V}(\theta_0)$, for a covariance structure $\mathbf{V}$ that satisfies (A1)–(A2), such that $\mathbf{L}$, as defined in (4), has full column rank. Let $\boldsymbol{\xi}_n$ and $\boldsymbol{\xi}_0$ be solutions of (13) and (15), respectively, and suppose that $\boldsymbol{\xi}_n \to \boldsymbol{\xi}_0$ in probability. Suppose that $\mathbb{E}\|\mathbf{s}\|^4 < \infty$ and that $\mathbf{X}$ has full rank with probability one. If w_1, w_2, and w_3 satisfy (C1)–(C3), then $\sqrt{n}\left\{\mathrm{vec}(\mathbf{V}_n) - \mathrm{vec}(\boldsymbol{\Sigma})\right\}$ is asymptotically normal with mean zero and variance (7), and $\sqrt{n}(\theta_n - \theta_0)$ is asymptotically normal with mean zero and variance (8), where*

$$\sigma_1 = \frac{k(k+2)\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{\left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3 + k(k+2)w_3(\|\mathbf{z}\|)\right]\right)^2}, \quad \sigma_2 = -\frac{2}{k}\sigma_1 + \frac{2\gamma_3\gamma_2(2\gamma_1 - \gamma_2\pi_L)}{\gamma_1^2(\gamma_1 - \gamma_2\pi_L)^2}\left(1 - \frac{\pi_L}{k}\right) + \frac{\gamma_4}{(\gamma_1 - \gamma_2\pi_L)^2}, \tag{17}$$

with γ_1, γ_2 and π_L defined in (16) and

$$\gamma_3 = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{k(k+2)}, \quad \gamma_4 = \frac{1}{k^2}\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(w_2(\|\mathbf{z}\|)\|\mathbf{z}\|^2 - kw_3(\|\mathbf{z}\|)\right)^2\right]. \tag{18}$$

Note that in general the constant σ_2 in Theorem 2 does depend on the projection $\boldsymbol{\Pi}_L$ through the constant π_L defined in (16). This is no longer the case for linear covariance structures. In that case, $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}) = \mathrm{vec}(\boldsymbol{\Sigma})$ so that $\pi_L = k$. This means we can relax condition (C3) and have the following corollary.

Corollary 2. *Under the assumptions of Theorem 2, where (C3) holds with $\pi_L = k$, suppose that $\mathbf{V}$ satisfies (5). Then $\sqrt{n}\left\{\mathrm{vec}(\mathbf{V}_n) - \mathrm{vec}(\boldsymbol{\Sigma})\right\}$ is asymptotically normal with mean zero and variance (3), and $\sqrt{n}(\theta_n - \theta_0)$ is asymptotically normal with mean zero and variance (2), with σ_1 defined in (17) and*

$$\sigma_2 = -\frac{2}{k}\sigma_1 + \frac{4\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(w_2(\|\mathbf{z}\|)\|\mathbf{z}\|^2 - kw_3(\|\mathbf{z}\|)\right)^2\right]}{\left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3 + 2kw_3(\|\mathbf{z}\|) - kw_3'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\right)^2}.$$

Remark 1. From the proof of Theorem 2 one can obtain that if $\mathbf{V}$ is linear, then

$$\sqrt{n}\left\{\mathrm{vec}(\mathbf{V}_n) - \mathrm{vec}(\boldsymbol{\Sigma})\right\} = -\boldsymbol{\Pi}_L\mathrm{vec}\left\{\sqrt{n}(\mathbf{N}_n - \mathbb{E}[\mathbf{N}_n])\right\} + o_P(1),$$

where $\boldsymbol{\Pi}_L$ is defined in (6) and

$$\mathbf{N}_n = \frac{1}{n}\sum_{i=1}^n\left\{v_1(d_i)(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_0)(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_0)^\top - v_2(d_i)\boldsymbol{\Sigma}\right\},$$

where $d_i^2 = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_0)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_0)$, and

$$v_1(s) = \frac{w_2(s)}{\gamma_1}, \quad v_2(s) = \frac{-\gamma_2 w_2(s)s^2 + \gamma_1 w_3(s)}{\gamma_1(\gamma_1 - k\gamma_2)},$$

where γ_1 and γ_2 are defined in (16). Moreover, $\sqrt{n}(\mathbf{N}_n - \mathbb{E}[\mathbf{N}_n]) \to \mathbf{N}$ in distribution, where $\mathbf{N}$ is a random matrix that has a multivariate distribution with mean zero and variance of the form (1).

The random matrix $\mathbf{N}$ in Remark 1 is of radial type with respect to $\boldsymbol{\Sigma}$. This follows from the fact that $\mathbf{R} = \boldsymbol{\Sigma}^{-1/2}\mathbf{N}\boldsymbol{\Sigma}^{-1/2}$ is multivariate normal with mean zero and variance

$$\mathrm{var}\{\mathrm{vec}(\mathbf{R})\} = (\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2})\mathrm{var}\{\mathrm{vec}(\mathbf{N})\}(\boldsymbol{\Sigma}^{-1/2} \otimes \boldsymbol{\Sigma}^{-1/2}) = \sigma_1(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}) + \sigma_2\mathrm{vec}(\mathbf{I}_k)\mathrm{vec}(\mathbf{I}_k)^\top.$$

This immediately gives that for any orthogonal matrix $\mathbf{O}$, the matrix $\mathbf{ORO}^\top$ is multivariate normal with mean zero and the same variance. From Corollary 2 it follows that $\sqrt{n}\left\{\text{vec}(\mathbf{V}_n) - \text{vec}(\boldsymbol{\Sigma})\right\}$ is asymptotically normal with mean zero and a variance that is the same as the variance of $\text{vec}(\mathbf{M}) = \boldsymbol{\Pi}_L \text{vec}(\mathbf{N})$. According to Corollary 1 this variance is of the type given by (3). Furthermore, if we write $\text{vec}(\mathbf{M}) = \mathbf{LT}$, then

$$\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\sqrt{n}\left\{\text{vec}(\mathbf{V}_n) - \text{vec}(\boldsymbol{\Sigma})\right\} \to \mathbf{T},$$

in distribution, where $\mathbf{T}$ is multivariate normal with mean zero and variance given by (2).

3.1. Examples

We discuss some examples of multivariate statistical models that are covered by the setup in (11), in which the estimators $(\boldsymbol{\beta}_n, \boldsymbol{\theta}_n)$ are solutions of estimating equation (13) for particular functions w_1, w_2, and w_3. In the Appendix A we provide a detailed derivation of σ_1 and σ_2 for specific special cases and show that their expressions coincide with the ones in Tyler [27] and Lopuhaä et al. [16].

Example 1 (*Maximum Likelihood for Multivariate Normal*)**.** Suppose $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$ are independent, such that $\mathbf{y}_i \mid \mathbf{X}_i \sim N_k(\mathbf{X}_i\boldsymbol{\beta}_0, \mathbf{V}(\boldsymbol{\theta}_0))$. The loglikelihood is then given by

$$\mathcal{L} = -\frac{nk}{2}\log(2\pi) - \frac{n}{2}\log|\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2}\sum_{i=1}^n(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top\mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}).$$

Setting the partial derivatives $\partial\mathcal{L}/\partial\boldsymbol{\beta}$ and $\partial\mathcal{L}/\partial\theta_j$ equal to zero gives the following estimating equations

$$\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^\top\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) = \mathbf{0}, \quad \frac{1}{n}\sum_{i=1}^n\left\{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top\mathbf{V}^{-1}\mathbf{L}_j\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - \text{tr}(\mathbf{V}^{-1}\mathbf{L}_j)\right\} = 0, \tag{19}$$

for $j \in \{1, \ldots, \ell\}$, where we write $\mathbf{V}$ for $\mathbf{V}(\boldsymbol{\theta})$. By using the vec-notation and $\mathbf{L}$ as defined in (4), we can combine the partial derivatives with respect to θ_j in the second line of (19) as follows

$$\mathbf{L}^\top(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1})\text{vec}\left\{\frac{1}{n}\sum_{i=1}^n(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top - \mathbf{V}\right\} = \mathbf{0}. \tag{20}$$

It follows that the maximum likelihood estimator $(\boldsymbol{\beta}_n, \boldsymbol{\theta}_n)$ satisfies (13) and $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ satisfies (15), where Ψ is defined in (14) with $w_1(s) = w_2(s) = w_3(s) = 1$. Theorem 2 applies and one finds $\sigma_1 = 1$ and $\sigma_2 = 0$.

When each $\mathbf{X}_i = \mathbf{I}_k$, for $i \in \{1, \ldots, n\}$, then the model (11) reduces to the multivariate location-scale model. If $\boldsymbol{\Sigma}$ is unstructured, then $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta}_0)$, with $\boldsymbol{\theta}_0 = \text{vech}(\boldsymbol{\Sigma})$ and $\mathbf{L} = \partial\text{vec}(\mathbf{V}(\boldsymbol{\theta}_0))/\partial\boldsymbol{\theta}^\top$ is equal to the duplication matrix $\mathcal{D}_k$. In this case, we can remove the factor $\mathbf{L}^\top(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1})$ from (20), and $\mathbf{V}_n$ is simply the sample covariance of $\mathbf{y}_1, \ldots, \mathbf{y}_n$. This example then coincides with Example 1 in Tyler [27].

Example 2 (*M-estimators*)**.** As mentioned in Example 1, when each $\mathbf{X}_i = \mathbf{I}_k$, for $i \in \{1, \ldots, n\}$, and $\boldsymbol{\Sigma}$ is unstructured, then the model (11) reduces to the multivariate location-scale model and we can remove the factor $\mathbf{L}^\top(\mathbf{V}^{-1}\otimes\mathbf{V}^{-1})$ from Ψ_θ in (13). In that case, estimating Eqs. (13) are equivalent to equations (1.1)–(1.2) in Maronna [22] or equations (4.11)–(4.12) in Huber [10] for M-estimators of multivariate location and covariance. In view of this, solutions $(\boldsymbol{\beta}_n, \boldsymbol{\theta}_n)$ of estimating Eqs. (13) are called M-estimators for $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$. The expression for σ_1 in Theorem 2 then coincides with the one in Example 3 in Tyler [27]. For linear covariance structures, the expression for σ_2 in Corollary 2 coincides with the one in Example 3 in Tyler [27].

As a special case, this includes the estimating equations that correspond to maximum likelihood estimators based on independent observations $(\mathbf{y}_1, \mathbf{X}_1), \ldots, (\mathbf{y}_n, \mathbf{X}_n)$ from an elliptical density (12). The maximum likelihood estimators $(\boldsymbol{\beta}_n, \boldsymbol{\theta}_n)$ then satisfy estimating Eqs. (13), for $w_1(s) = w_2(s) = -2g'(s^2)/g(s^2)$ and $w_3(s) = 1$. The expression for σ_1 in Theorem 2 then coincides with the one in Example 2 in Tyler [27]. For linear covariance structures, the expression for σ_2 in Corollary 2 coincides with the one in Example 2 in Tyler [27].

Example 3 (*S-estimators*)**.** S-estimators for $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ are defined by means of a function $\rho : \mathbb{R} \to [0, \infty)$, as the solution to minimizing $|\mathbf{V}(\boldsymbol{\theta})|$, subject to

$$\frac{1}{n}\sum_{i=1}^n \rho\left(\sqrt{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top\mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}\right) \le b_0,$$

where the minimum is taken over all $\boldsymbol{\beta} \in \mathbb{R}^q$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$, such that $\mathbf{V}(\boldsymbol{\theta}) \in \text{PDS}(k)$. These estimators have been studied for linear mixed effects models in Copt and Victoria-Feser [4], Chervoneva and Vishnyakov [1,2] and for general linear models with a structured covariance in Lopuhaä et al. [16]. When $\mathbf{V}$ is linear, then according to Section 7.2 in [16], S-estimators $(\boldsymbol{\beta}_n, \boldsymbol{\theta}_n)$ satisfy estimating Eqs. (13), with $w_1(d) = \rho'(d)/d$, $w_2(s) = k\rho'(s)/s$ and $w_3(s) = \rho'(s)s - \rho(s) + b_0$. The expressions for σ_1 and σ_2 in Corollary 2 coincide with the ones in Corollary 9.2 in Lopuhaä et al. [16].

4. Homogeneous mappings of order zero

Let $H(\mathbf{v})$ be a mapping from $\mathbb{R}^l$ to $\mathbb{R}^m$ that is *homogeneous of order zero*, that is

$$H(\mathbf{v}) = H(\alpha\mathbf{v}), \quad \alpha > 0. \tag{21}$$

These mappings have several applications to affine equivariant covariance estimators that have limiting variance (1). Tyler [28] uses such a mapping to show that the likelihood ratio criterion is asymptotically robust over the class of elliptical distributions. Kent and Tyler [12] consider the shape component of covariance CM-estimators and show that the limiting variance of CM-estimators of shape depends on σ_1 only, which may then serve as an index for the asymptotic relative efficiency. Salibián-Barrera et al. [25] derive the influence function of the shape component of covariance MM-functionals and use this to obtain that the limiting variance of MM-estimators of shape only depends on a single scalar. This property of the shape component is a special case of a general result in Tyler [28] for multivariate functionals of affine equivariant covariance estimators that are asymptotically normal with limiting variance (1).

Estimators for a structured covariance matrix are typically not affine equivariant and have limiting variance (7) or (3) instead of (1), so that the previous results do not directly apply. The objective of this section is to extend Theorem 1 in Tyler [28] to estimators for a linearly structured covariance, and discuss its consequences for corresponding estimators of shape and scale. Moreover, we establish a similar result for estimators of the vector of variance components and apply this to its normalized version. We then have the following theorem.

Theorem 3. *Consider $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta}_0) \in PDS(k)$, for some vector $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$, where $\mathbf{V}$ satisfies (5), such that $\mathbf{L}$, as defined in (4), is of full column rank. Let $\{\mathbf{V}_n : n \geq 1\}$ be a sequence of estimators for $\boldsymbol{\Sigma}$ and let $\{\boldsymbol{\theta}_n : n \geq 1\}$ be a sequence of estimators for the vector $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$ of variance components.*

(i) *For $\mathbf{V} \in PDS(k)$, let $H(\mathbf{V})$ be continuously differentiable satisfying (21). When $\sqrt{n}(\mathbf{V}_n - \boldsymbol{\Sigma})$ converges in distribution to a random matrix $\mathbf{M}$ that has a multivariate normal distribution with mean zero and variance given by (3), then $\sqrt{n}(H(\mathbf{V}_n) - H(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and variance*

$$2\sigma_1 H'(\boldsymbol{\Sigma})\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top H'(\boldsymbol{\Sigma})^\top.$$

(i) *For $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$, let $H(\boldsymbol{\theta})$ be continuously differentiable satisfying (21). When $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ is asymptotically normal with mean zero and variance (2), then $\sqrt{n}(H(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}_0))$ is asymptotically normal with mean zero and variance*

$$2\sigma_1 H'(\boldsymbol{\theta}_0)\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}H'(\boldsymbol{\theta}_0)^\top.$$

Remark 2. The restriction to linear $\mathbf{V}$ in Theorem 3 is essential. It can be seen from the proof that Theorem 3(i) is a direct consequence of the property $H'(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{0}$. In view of (7), a similar result for nonlinear $\mathbf{V}$ would require $H'(\boldsymbol{\Sigma})\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{0}$, which is not necessarily the case. Similarly, Theorem 3(ii) is a direct consequence of $H'(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 = \mathbf{0}$. In view of (8), a similar result for nonlinear $\mathbf{V}$ would require $H'(\boldsymbol{\theta}_0)\boldsymbol{\theta}_L = \mathbf{0}$, where $\boldsymbol{\theta}_L \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$, such that $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_L$, which may also not be true.

When $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta}_0)$ is unstructured, then $\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_0$, with $\boldsymbol{\theta}_0 = \mathrm{vech}(\boldsymbol{\Sigma})$, as defined in (9), and $\mathbf{L}$ is the duplication matrix $\mathcal{D}_k$. Because $\mathbf{K}_{k,k}H'(\mathbf{V})^\top = H'(\mathbf{V})^\top$, for symmetric $\mathbf{V}$, from (10) it follows that Theorem 3(i) with $\mathbf{L} = \mathcal{D}_k$ recovers Theorem 1 in Tyler [28].

From Theorem 3 it follows immediately that the asymptotic relative efficiency of different estimators $H(\mathbf{V}_n)$ for $H(\boldsymbol{\Sigma})$ can be compared by simply comparing the values of the corresponding scalar σ_1. Similarly, the scalar σ_1 can also be used as an index for the asymptotic relative efficiency of different estimators $H(\boldsymbol{\theta}_n)$ for $H(\boldsymbol{\theta}_0)$. We discuss some examples below.

The results may also be relevant when studying the behavior of robust inference procedures. Robust tests are meant to have a stable level under small arbitrary departures from the null hypothesis, and to have good power under small arbitrary departures from specified alternatives. Tyler [27] considers a robust likelihood ratio test, which is a standardized version of $H(\mathbf{V}_n)$, where $\mathbf{V}_n$ is a covariance M-estimator and H is the scale invariant mapping defined by $H(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{1/2}/(\mathrm{tr}(\boldsymbol{\Sigma}))^{k/2}$. Under appropriate scaling $H(\mathbf{V}_n)$ has a limiting distribution that only depends on the scalar σ_1 from (1). Related robust inference procedures can be found in [9] for a general parametric setup, in [3] for linear mixed effects models, and in [30] for a one-way multivariate ANOVA, among others.

Example 4 (*Shape and Scale of a Linearly Structured Covariance*)**.** Suppose that $\sqrt{n}(\mathbf{V}_n - \boldsymbol{\Sigma})$ is asymptotically normal with mean zero and variance given by (3). Consider the shape component $H(\mathbf{C}) = \mathrm{vec}(\mathbf{C})/|\mathbf{C}|^{1/k}$, where $\mathbf{C} \in \mathrm{PDS}(k)$. We have that

$$H'(\mathbf{C}) = \frac{\partial H(\mathbf{C})}{\partial \mathrm{vec}(\mathbf{C})^\top} = -\frac{1}{k}|\mathbf{C}|^{-1/k}\mathrm{vec}(\mathbf{C})\mathrm{vec}(\mathbf{C}^{-1})^\top + |\mathbf{C}|^{-1/k}\mathbf{I}_{k^2}. \tag{22}$$

Then, according to Theorem 3(i), for the shape component it follows that $\sqrt{n}(H(\mathbf{V}_n) - H(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and variance (see Appendix A for details)

$$\frac{2\sigma_1}{|\boldsymbol{\Sigma}|^{2/k}}\left\{\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top - \frac{1}{k}\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top\right\}. \tag{23}$$

When $\boldsymbol{\Sigma}$ is unstructured, then $\text{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_0$, with $\boldsymbol{\theta}_0 = \text{vech}(\boldsymbol{\Sigma})$ and $\mathbf{L}$ is the duplication matrix D_k. In that case, from (10) it follows that (23) with $\mathbf{L} = D_k$ reduces to

$$\frac{\sigma_1}{|\boldsymbol{\Sigma}|^{2/k}}\left\{\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) - \frac{2}{k}\text{vec}(\boldsymbol{\Sigma})\text{vec}(\boldsymbol{\Sigma})^\top\right\}.$$

This coincides with expression (9) found in [25]. For completeness, consider the scale component $\sigma(\mathbf{C}) = |\mathbf{C}|^{1/(2k)}$. It can be seen that

$$\sigma'(\mathbf{C}) = \frac{1}{2k}|\mathbf{C}|^{1/(2k)}\text{vec}(\mathbf{C}^{-1})^\top. \tag{24}$$

Application of the delta method then yields that $\sqrt{n}(\sigma(\mathbf{V}_n) - \sigma(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and variance

$$\frac{1}{4}\left(\frac{2\sigma_1}{k} + \sigma_2\right)|\boldsymbol{\Sigma}|^{1/k}.$$

Example 5 (*Direction of the Vector of Variance Components of a Linear Covariance Structure*)**.** Suppose that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ is asymptotically normal with mean zero and variance given by (2). In order to create a single scalar as an index of the asymptotic efficiency for estimators $\boldsymbol{\theta}_n$ for the vector $\boldsymbol{\theta}_0$ of variance components of a linear covariance structure, it is helpful to separate $\boldsymbol{\theta}_0$ into its direction and length. The direction component $H(\boldsymbol{\theta}) = \boldsymbol{\theta}/\|\boldsymbol{\theta}\|$ satisfies (21). Its derivative is given by

$$H'(\boldsymbol{\theta}) = \frac{\partial H(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = \frac{1}{\|\boldsymbol{\theta}\|}\left(\mathbf{I}_\ell - \frac{\boldsymbol{\theta}\boldsymbol{\theta}^\top}{\|\boldsymbol{\theta}\|^2}\right). \tag{25}$$

Then, according to Theorem 3(ii), for the direction estimator it follows that $\sqrt{n}(H(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}))$ is asymptotically normal with mean zero and variance

$$\frac{2\sigma_1}{\|\boldsymbol{\theta}_0\|^2}\left(\mathbf{I}_\ell - \frac{\boldsymbol{\theta}_0\boldsymbol{\theta}_0^\top}{\|\boldsymbol{\theta}_0\|^2}\right)\left(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1}\left(\mathbf{I}_\ell - \frac{\boldsymbol{\theta}_0\boldsymbol{\theta}_0^\top}{\|\boldsymbol{\theta}_0\|^2}\right).$$

It does not seem possible to simplify this expression any further, but it illustrates that one can use the scalar σ_1 as an index for the asymptotic relative efficiency of estimators $H(\boldsymbol{\theta}_n)$ for $H(\boldsymbol{\theta}_0)$.

An alternative is the mapping $H(\boldsymbol{\theta}) = \boldsymbol{\theta}/|\mathbf{V}(\boldsymbol{\theta})|^{1/k}$. When $\mathbf{V}$ is linear, this H also satisfies (21). For $\mathbf{V}_n = \mathbf{V}(\boldsymbol{\theta}_n)$, it holds that $\boldsymbol{\theta}_n = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\text{vec}(\mathbf{V}_n)$, so that

$$H(\boldsymbol{\theta}_n) = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\text{vec}\left(\mathbf{V}_n/|\mathbf{V}_n|^{1/k}\right).$$

From Example 4, it follows that $\sqrt{n}(H(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}))$ is asymptotically normal with mean zero and variance

$$\frac{2\sigma_1}{|\boldsymbol{\Sigma}|^{2/k}}\left\{\left(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1} - \frac{1}{k}\boldsymbol{\theta}_0\boldsymbol{\theta}_0^\top\right\}.$$

This component H leads to a simpler expression for the limiting variance and the scalar σ_1 can again be used as an index for the asymptotic relative efficiency of estimators $H(\boldsymbol{\theta}_n)$ for $H(\boldsymbol{\theta}_0)$.

5. Influence function of structured covariance functionals

The influence function measures the local robustness of an estimator. It describes the effect of an infinitesimal contamination at a single point on the corresponding functional (see Hampel [7]). Good local robustness is therefore illustrated by a bounded influence function. It is defined as follows. Let P be a distribution on $\mathbb{R}^k$. For $0 < h < 1$ and $\mathbf{y} \in \mathbb{R}^k$ fixed, define the perturbed probability measure $P_{h,\mathbf{y}} = (1-h)P + h\delta_{\mathbf{y}}$, where $\delta_{\mathbf{y}}$ denotes the Dirac measure at $\mathbf{y} \in \mathbb{R}^k$. The *influence function* of a $k\times k$ covariance functional $\mathbf{C}(\cdot)$ at probability measure P, is defined as

$$\text{IF}(\mathbf{y}; \mathbf{C}, P) = \lim_{h\downarrow 0}\frac{\mathbf{C}((1-h)P + h\delta_{\mathbf{y}}) - \mathbf{C}(P)}{h},$$

if this limit exists.

Let P be a distribution on $\mathbb{R}^k$ with density $|\boldsymbol{\Sigma}|^{-1/2}g\left((\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$, where $\boldsymbol{\mu} \in \mathbb{R}^k$ and $\boldsymbol{\Sigma} \in \text{PDS}(k)$, and let $\mathbf{C}$ be *Fisher consistent* for $\boldsymbol{\Sigma}$, that is $\mathbf{C}(P) = \boldsymbol{\Sigma}$, and *affine equivariant*, meaning $\mathbf{C}(P_{\mathbf{Ay}+\mathbf{b}}) = \mathbf{A}\mathbf{C}(P_{\mathbf{y}})\mathbf{A}^\top$, for any nonsingular $k\times k$ matrix $\mathbf{A}$ and $\mathbf{b} \in \mathbb{R}^k$, where $P_{\mathbf{y}}$ denotes the distribution of a random vector $\mathbf{y}$. Croux and Haesbroeck [5] show that the influence function of such covariance functionals at the $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution is given by

$$\text{IF}(\mathbf{y}; \mathbf{C}, P) = \alpha_C(d(\mathbf{y}))(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top - \beta_C(d(\mathbf{y}))\boldsymbol{\Sigma}, \tag{26}$$

for some real valued functions α_C and β_C and where $d(\mathbf{y})^2 = (\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$. For more details on α_C and β_C for different covariance functionals, see Croux and Haesbroeck [5].

Structured covariance functionals $\mathbf{M}(\cdot) = \mathbf{V}(\boldsymbol{\theta}(\cdot))$ are not necessarily affine equivariant, so that the above characterizations do not directly apply. However, Lopuhaä et al. [17] find similar expressions for the influence function of the covariance S-functionals $\mathbf{M}(\cdot)$ and $\boldsymbol{\theta}(\cdot)$ in a linear model with a linearly structured covariance $\mathbf{V}$, see Corollary 8.4 in [17]. The next lemma shows that these type of expressions will always appear at elliptical distributions for covariance functionals that are a projection of some affine equivariant covariance functional.

Lemma 1. *Let P be a distribution on $\mathbb{R}^k$ with density $|\boldsymbol{\Sigma}|^{-1/2}g\left((\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$, where $\boldsymbol{\mu}\in\mathbb{R}^k$ and $\boldsymbol{\Sigma}\in PDS(k)$. Let $\mathbf{C}$ be an affine equivariant covariance functional which possesses an influence function. Suppose that $\boldsymbol{\Sigma}=\mathbf{V}(\boldsymbol{\theta}_0)$, for some $\boldsymbol{\theta}_0\in\boldsymbol{\Theta}\subset\mathbb{R}^\ell$, such that $\mathbf{L}$, as defined in (4), is of full column rank. Let $\boldsymbol{\Pi}_L$ be the projection matrix defined in (6) and define the covariance functional $\mathbf{M}$ by $\mathrm{vec}(\mathbf{M})=\boldsymbol{\Pi}_L\mathrm{vec}(\mathbf{C})$. Then the following holds.*

(i) *There exist functions $\alpha_C,\beta_C:[0,\infty)\to\mathbb{R}$, such that IF$(\mathbf{y};\mathrm{vec}(\mathbf{M}),P)$ is given by*

$$\alpha_C(d(\mathbf{y}))\mathbf{L}\Big(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{L}\Big)^{-1}\mathbf{L}^\top\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}\right)-\beta_C(d(\mathbf{y}))\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}),$$

where $d(\mathbf{y})^2=(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$.

(i) *If $\boldsymbol{\theta}(P)\in\boldsymbol{\Theta}\subset\mathbb{R}^\ell$ is the functional, such that $\mathrm{vec}(\mathbf{M}(\cdot))=\mathbf{L}\boldsymbol{\theta}(\cdot)$, then IF$(\mathbf{y};\boldsymbol{\theta},P)$ is given by*

$$\alpha_C(d(\mathbf{y}))\Big(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{L}\Big)^{-1}\mathbf{L}^\top\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}\right)-\beta_C(d(\mathbf{y}))\boldsymbol{\theta}_L,$$

where $\boldsymbol{\theta}_L\in\boldsymbol{\Theta}\subset\mathbb{R}^\ell$ is such that $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma})=\mathbf{L}\boldsymbol{\theta}_L$.

Note that the functions α_C and β_C have nothing to do with the projection $\boldsymbol{\Pi}_L$, but are inherited from the influence function (26) of the affine equivariant covariance functional $\mathbf{C}$. At a distribution P that has an elliptical density (12) with a linearly structured covariance, one has $\mathrm{vec}(\boldsymbol{\Sigma})=\mathbf{L}\boldsymbol{\theta}_0$, so that $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma})=\mathrm{vec}(\boldsymbol{\Sigma})$ in part (i) and $\boldsymbol{\theta}_L=\boldsymbol{\theta}_0$ in part (ii) of Lemma 1. For this case, Lopuhaä et al. [17] find expressions similar to the ones in Lemma 1 for the covariance S-functionals. If the S-functional is defined by some function ρ and constant b_0 (see Example 3), then

$$\alpha_C(s)=\frac{k\rho'(s)}{s\delta_1},\quad \beta_C(s)=\frac{\rho'(s)s}{\delta_1}-\frac{2(\rho(s)-b_0)}{\delta_2}, \tag{27}$$

where

$$\delta_1=\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}}\left[\rho''(\|\mathbf{z}\|)\|\mathbf{z}\|^2+(k+1)\rho'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]}{k+2},\quad \delta_2=\mathbb{E}_{\mathbf{0},\mathbf{I}}\left[\rho'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]. \tag{28}$$

These α_C and β_C are the same as the ones that appear in the expression for the influence function of the affine equivariant covariance S-functional $\mathbf{C}$ in the multivariate location-scale model, see Lopuhaä [14] or Salibián-Barrera et al. [25], or in the multivariate regression model, see Van Aelst and Willems [29]. Indeed, the influence function IF$(\mathbf{y},\mathrm{vec}(\mathbf{V}(\boldsymbol{\theta})),P)$ of the linearly structured covariance functional in Lopuhaä et al. [17] is precisely the projection $\boldsymbol{\Pi}_L$ of IF$(\mathbf{y},\mathrm{vec}(\mathbf{C}),P)$ as obtained in [14,25,29].

When $\boldsymbol{\Sigma}=\mathbf{V}(\boldsymbol{\theta}_0)$ is unstructured, then $\mathrm{vec}(\boldsymbol{\Sigma})=\mathbf{L}\boldsymbol{\theta}_0$ with $\boldsymbol{\theta}_0=\mathrm{vech}(\boldsymbol{\Sigma})$ and $\mathbf{L}$ is the duplication matrix $\mathcal{D}_k$. In that case, from (10) it follows that the expression for IF$(\mathbf{y};\mathrm{vec}(\mathbf{M}),P)$ in Lemma 1(i) with $\mathbf{L}=\mathcal{D}_k$ reduces to

$$\mathrm{IF}(\mathbf{y};\mathrm{vec}(\mathbf{M}),P)=\mathrm{vec}\left\{\alpha_C(d(\mathbf{y}))(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top-\beta_C(d(\mathbf{y}))\boldsymbol{\Sigma}\right\}.$$

This coincides with the expression found in Lemma 1 in Croux and Haesbroeck [5].

Mappings H that satisfy (21) also have useful applications to influence functions of affine equivariant covariance functionals $\mathbf{C}$ and their the gross-error-sensitivity (GES). Kent and Tyler [12] consider functionals $\mathbf{C}/|\mathbf{C}|^{1/k}$ and $\mathbf{C}/\mathrm{tr}(\mathbf{C})$ to obtain that the GES of different CM-functionals is proportional to a single scalar. Salibián-Barrera et al. [25] derive the influence function of the shape component of covariance MM-functionals and show that it is proportional to a single function α_C, which no longer depends on the scale-functional used in the first step. In fact, these properties hold more general for functionals H satisfying (21) applied to affine equivariant covariance functionals. The next lemma establishes similar results for linearly structured covariance functionals. Similar to Remark 2, one can only obtain this result for linear covariance functionals.

Lemma 2. *Let P be a distribution on $\mathbb{R}^k$ with an elliptical contoured density (12). Suppose that $\boldsymbol{\Sigma}=\mathbf{V}(\boldsymbol{\theta}_0)\in PDS(k)$, for some vector $\boldsymbol{\theta}_0\in\boldsymbol{\Theta}\subset\mathbb{R}^\ell$, where $\mathbf{V}$ satisfies (5), such that $\mathbf{L}$, as defined in (4) is of full column rank.*

(i) *Let $\mathbf{M}\in PDS(k)$ be a covariance functional that is Fisher consistent for $\boldsymbol{\Sigma}$ and which possesses an influence function given by Lemma 1(i). Let $H(\mathbf{M})$ be continuously differentiable in a neighborhood of $\mathbf{M}(P)$ satisfying (21). Then IF$(\mathbf{y};H(\mathbf{M}),P)$ is given by*

$$\alpha_C(d(\mathbf{y}))H'(\boldsymbol{\Sigma})\mathbf{L}\Big(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{L}\Big)^{-1}\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\Big((\mathbf{y}-\boldsymbol{\mu})\otimes(\mathbf{y}-\boldsymbol{\mu})\Big),$$

where $d^2(\mathbf{y})=(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$.

(i) *Let $\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset\mathbb{R}^\ell$ be a functional that is Fisher consistent for $\boldsymbol{\theta}_0$ and which possesses an influence function given by Lemma 1(ii). Let $H(\boldsymbol{\theta})$ be continuously differentiable in a neighborhood of $\boldsymbol{\theta}(P)$ satisfying (21). Then IF$(\mathbf{y};H(\boldsymbol{\theta}),P)$ is given by*

$$\alpha_C(d(\mathbf{y}))H'(\boldsymbol{\theta}_0)\Big(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{L}\Big)^{-1}\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\Big((\mathbf{y}-\boldsymbol{\mu})\otimes(\mathbf{y}-\boldsymbol{\mu})\Big).$$

Consider the GES defined by $\sup_{\mathbf{y}\in\mathbb{R}^k}\|\mathrm{IF}(\mathbf{y};\cdot)\|$, for some norm $\|\cdot\|$. From Lemma 2 it follows immediately that regardless of the choice of the norm, the value $\|\mathrm{IF}(\mathbf{y};H(\mathbf{M}),P)\|$ for different functionals $H(\mathbf{M}(P))$ is proportional to $|\alpha_C(d(\mathbf{y}))|$ and similarly for functionals $H(\boldsymbol{\theta}(P))$. We discuss some examples below.

Example 6 (*Shape and Scale of a Linearly Structured Covariance*)**.** For the shape functional $H(\mathbf{M}) = \text{vec}(\mathbf{M})/|\mathbf{M}|^{1/k}$, from Lemma 2(i) together with (22) we find

$$\text{IF}(\mathbf{y}; H(\mathbf{M}), P) = -\frac{1}{k}|\boldsymbol{\Sigma}|^{-1/k}\text{tr}\left(\boldsymbol{\Sigma}^{-1}\text{IF}(\mathbf{y}; \mathbf{M}, P)\right)\cdot\text{vec}(\boldsymbol{\Sigma}) + |\boldsymbol{\Sigma}|^{-1/k}\text{IF}(\mathbf{y}; \text{vec}(\mathbf{M}), P).$$

See also Salibián et al. [25]. In particular, at a distribution P with an elliptically contoured density with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta}_0)$ one finds that $\text{IF}(\mathbf{y}; H(\mathbf{M}), P)$ is given by

$$\frac{\alpha_C(d(\mathbf{y}))}{|\boldsymbol{\Sigma}|^{1/k}}\left\{\mathbf{L}\Big(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{L}\Big)^{-1}\mathbf{L}^\top\text{vec}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}\right) - \frac{d(\mathbf{y})^2}{k}\text{vec}(\boldsymbol{\Sigma})\right\}, \tag{29}$$

where $d(\mathbf{y})^2 = (\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$. It follows that $\|\text{IF}(\mathbf{y}; H(\boldsymbol{\theta}), P)\|$ will be proportional to $|\alpha_C(d(\mathbf{y}))d(\mathbf{y})^2|$. When $\boldsymbol{\Sigma}$ is unstructured, then $\text{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0 = \text{vech}(\boldsymbol{\Sigma})$, as defined in (9), and $\mathbf{L}$ is the duplication matrix $\mathcal{D}_k$. In that case, from (10) it follows that (29) with $\mathbf{L} = \mathcal{D}_k$ reduces to

$$\frac{\alpha_C(d(\mathbf{y}))}{|\boldsymbol{\Sigma}|^{1/k}}\text{vec}\left\{(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top - \frac{d(\mathbf{y})^2}{k}\boldsymbol{\Sigma}\right\},$$

which coincides with formula (3) in [25]. For completeness, consider the scale component $\sigma(\mathbf{M}) = |\mathbf{M}|^{1/(2k)}$. From (24), it follows that

$$\text{IF}(\mathbf{y}; \sigma, P) = \frac{1}{2}|\boldsymbol{\Sigma}|^{-1/(2k)}\gamma_C(d(\mathbf{y})),$$

where $\gamma_C(s) = \alpha_C(s)s^2/k - \beta_C(s)$, which matches with equation (4) in [25].

Example 7 (*Direction of the Vector of Variance Components of a Linear Covariance Structure*)**.** For the direction functional $H(\boldsymbol{\theta}) = \boldsymbol{\theta}/\|\boldsymbol{\theta}\|$, from Lemma 2(ii) together with (25) we find that, at a distribution P with an elliptically contoured distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta}_0)$, $\text{IF}(\mathbf{y}; H(\boldsymbol{\theta}), P)$ is given by

$$\alpha_C(d(\mathbf{y}))\left(\frac{1}{\|\boldsymbol{\theta}_0\|}\mathbf{I}_\ell - \frac{\boldsymbol{\theta}_0\boldsymbol{\theta}_0^\top}{\|\boldsymbol{\theta}_0\|^3}\right)\Big(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{L}\Big)^{-1}\mathbf{L}^\top\text{vec}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}\right).$$

It follows that $\|\text{IF}(\mathbf{y}; H(\boldsymbol{\theta}), P)\|$ will be proportional to $|\alpha_C(d(\mathbf{y}))d(\mathbf{y})^2|$. An alternative is the mapping $H(\boldsymbol{\theta}) = \boldsymbol{\theta}/|\mathbf{V}(\boldsymbol{\theta})|^{1/k}$. Since $\mathbf{V}$ is linear, H satisfies (21). For $\mathbf{M}(P) = \mathbf{V}(\boldsymbol{\theta}(P))$, it holds that $\boldsymbol{\theta}(P) = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\text{vec}(\mathbf{M}(P))$, so that

$$H(\boldsymbol{\theta}(P)) = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\text{vec}(\mathbf{M}(P))/|\mathbf{M}(P)|^{1/k}.$$

From Example 6 it follows that $\text{IF}(\mathbf{y}; H(\boldsymbol{\theta}), P)$ is given by

$$\frac{\alpha_C(d(\mathbf{y}))}{|\boldsymbol{\Sigma}|^{1/k}}\left\{\Big(\mathbf{L}^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{L}\Big)^{-1}\mathbf{L}^\top\text{vec}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}\right) - \frac{d(\mathbf{y})^2}{k}\boldsymbol{\theta}_0\right\},$$

using that $\text{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_0$. Again we find that $\|\text{IF}(\mathbf{y}; H(\boldsymbol{\theta}), P)\|$ is proportional to $|\alpha_C(d(\mathbf{y}))d(\mathbf{y})^2|$.

6. Application

We apply our results to S-estimators and S-functionals in the linear model (11). Let P be the distribution for the random variable $\mathbf{s} = (\mathbf{y}, \mathbf{X})$, which is such that $\mathbf{y} \mid \mathbf{X}$ has an elliptically contoured density (12) with parameters $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta}_0) = \theta_{01}\mathbf{L}_1 + \cdots + \theta_{0\ell}\mathbf{L}_\ell$. Consider the S-estimator for $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ defined as the solution to minimizing $|\mathbf{V}(\boldsymbol{\theta})|$, subject to

$$\frac{1}{n}\sum_{i=1}^n \rho\left(\sqrt{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^\top\mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}\right) = b_0,$$

where the minimum is taken over all $\boldsymbol{\beta} \in \mathbb{R}^q$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^\ell$, such that $\mathbf{V}(\boldsymbol{\theta}) \in \text{PDS}(k)$. For the function ρ we take Tukey's bi-weight

$$\rho_{\text{B}}(s; c) = \begin{cases} s^2/2 - s^4/(2c^2) + s^6/(6c^4), & |s| \le c; \\ c^2/6 & |s| > c, \end{cases}$$

and $b_0 = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}[\rho_B(\|\mathbf{z}\|; c)]$. From Theorem 6.1 in Lopuhaä et al. [16] it is known that the breakdown point of the S-estimator depends on the cut-off constant c and is at least $\lceil nb_0/(c^2/6)\rceil/n$, or asymptotically $\epsilon^* = b_0/(c^2/6)$. Table 1 gives the cut-off values of ρ_B for given asymptotic lower bounds $\epsilon^* \in \{0.05, 0.10, \ldots, 0.50\}$ on the breakdown point in dimensions $k \in \{1, 2, 5, 10\}$. This table partly overlaps with Table 3 in Rousseeuw and Yohai [24].

According to Corollary 9.2 in Lopuhaä et al. [17], the scalar $\lambda = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'_B(\|\mathbf{z}\|; c)^2\right]/(k\alpha^2)$ represents the asymptotic efficiency of the regression S-estimator $\boldsymbol{\beta}_n$ relative to the least squares estimator (for which $\lambda = 1$), where

$$\alpha = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(1 - \frac{1}{k}\right)\frac{\rho'_B(\|\mathbf{z}\|; c)}{\|\mathbf{z}\|} + \frac{1}{k}\rho''_B(\|\mathbf{z}\|; c)\right]. \tag{30}$$

Table 1
Cut-off values of ρ_B for different breakdown points $\epsilon^* \in \{0.05, 0.10, \ldots, 0.50\}$ and dimensions $k \in \{1, 2, 5, 10\}$.

	Breakdown point									
k	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
1	7.545	5.182	4.096	3.421	2.937	2.561	2.252	1.988	1.756	1.548
2	10.767	7.474	5.981	5.069	4.427	3.938	3.542	3.209	2.920	2.661
5	17.114	11.950	9.628	8.220	7.242	6.505	5.918	5.432	5.017	4.652
10	24.246	16.961	13.694	11.719	10.351	9.324	8.510	7.840	7.271	6.776

From Examples 4 and 5, together with Theorem 2 and Example 3, it follows that the scalar

$$\sigma_1 = \frac{k\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'_B(\|\mathbf{z}\|;c)^2\|\mathbf{z}\|^2\right]}{(k+2)\delta_1^2},$$

where δ_1 is defined in (28), serves as an index for the asymptotic efficiency of both the S-estimator of shape as well as the S-estimator for the direction of the vector of variance components, relative to the least squares estimators of shape and direction, respectively (for which $\sigma_1 = 1$). Finally, from Example 4, together with Example 3, it follows that

$$\sigma_3 = \frac{1}{4}\left(\frac{2\sigma_1}{k} + \sigma_2\right) = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(\rho_B(\|\mathbf{z}\|;c) - b_0\right)^2\right]}{\delta_2^2},$$

where δ_2 is defined in (28), serves as an index for the asymptotic efficiency of the S-estimator of scale relative the least squares (for which $\sigma_3 = 1/(2k)$). As a consequence, the cutoff constant c of ρ_B can be tuned in such a way that the asymptotic efficiency $1/\lambda$ relative to the least squares estimator is high at the normal distribution and similarly for $1/\sigma_1$ and $1/(2k\sigma_3)$. Since c also determines the breakdown point, this forces a trade-off between efficiency and breakdown point. Typically, large values of c correspond to high efficiency and low breakdown point, and vice-versa for moderate values of c.

We further investigate how this trade-off relates to the gross error sensitivity (GES) of the corresponding S-functionals. For simplicity we only consider perturbations in $\mathbf{y}$ and leave $\mathbf{X}$ unchanged. From Corollary 8.4 in Lopuhaä et al. [17], for the regression S-functional it then follows that $\|\mathrm{IF}(\mathbf{y};\boldsymbol{\beta},P)\|$ is proportional to $\alpha^{-1}\left|\rho'_B(d(\mathbf{y});c)\right|$, where α is defined in (30) and $d(\mathbf{y})^2 = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)$. Therefore, we propose the scalar

$$G_1 = \frac{1}{\alpha}\sup_{s>0}\left|\rho'_B(s;c)\right|,$$

as an index for the GES of regression S-functionals. This coincides with the GES index for location CM-functionals in Kent and Tyler [12]. From Examples 6 and 7, together with Lemma 2 and (27), for both the shape and direction S-functional, it follows that $\|\mathrm{IF}(\mathbf{y})\|$ is proportional to $\delta_1^{-1}|\rho'_B(d(\mathbf{y});c)d(\mathbf{y})|$, where δ_1 is defined in (28). We propose the scalar

$$G_2 = \frac{k}{(k+2)\delta_1}\sup_{s>0}\left|\rho'_B(s;c)s\right|,$$

as an index for the GES of shape and direction S-functionals. In this way, G_2 coincides with the GES index for CM-functionals of shape in Kent and Tyler [12]. Finally, from Example 6 and (27), if follows that for the scale functional $\|\mathrm{IF}(\mathbf{y})\|$ is proportional to $\delta_2^{-1}|\rho_B(d(\mathbf{y});c) - b_0|$, where δ_2 is defined in (28). We propose

$$G_3 = \frac{1}{\delta_2}\sup_{s>0}\left|\rho_B(s;c) - b_0\right|,$$

as an index for the GES of the S-functional of scale.

We investigate how the asymptotic efficiency at the normal distribution of the S-estimators, and the GES of the corresponding S-functionals behave as we vary the breakdown point of the S-estimator between 0 and 0.5. Given a value ϵ^* of the breakdown point, we determine the corresponding cut-off constant c by solving $\epsilon^* = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}[\rho_B(\|\mathbf{z}\|;c)]/(c^2/6)$. With this value of c, we compute the values of λ, σ_1 and σ_3 and the GES indices G_1, G_2 and G_3. In Fig. 1, on the top row we have plotted the asymptotic relative efficiencies $1/\lambda$, $1/\sigma_1$ and $1/(2k\sigma_3)$ as a function of the breakdown point for dimensions $k \in \{2, 5, 10\}$, and the bottom row contains plots of the GES indices G_1, G_2 and G_3 for the same dimensions. The efficiency of high-breakdown S-estimators quickly increases with dimension k, especially for the regression and shape/direction estimators. At the same time the effect on the GES is much smaller for high breakdown S-estimators. Furthermore, as expected, the efficiency decreases with increasing breakdown point, but the loss of efficiency is less severe for the S-estimator of scale compared to the S-estimator for regression and the S-estimators for shape and direction.

In dimension $k = 2$ (solid lines), the 50% breakdown S-estimators have asymptotic efficiencies $1/\lambda = 0.580$, $1/\sigma_1 = 0.376$, and $1/(4\sigma_3) = 0.755$. However, one can gain both efficiency and lower the GES at the cost of a lower breakdown point. For example, the GES index of the regression functional attains its minimal value $G_1 = 1.927$ at breakdown point 28%, which corresponds to cut-off value $c = 4.115$. For this cut-off value the GES index of the shape and direction functional is $G_2 = 1.368$, which is not far off from its minimal value 1.344, and the GES index for scale is $G_3 = 3.323$. Furthermore, the asymptotic efficiencies then become $1/\lambda = 0.884$, $1/\sigma_1 = 0.803$, and $1/(4\sigma_3) = 0.939$, for the regression estimator, the estimators of shape and direction, and the scale

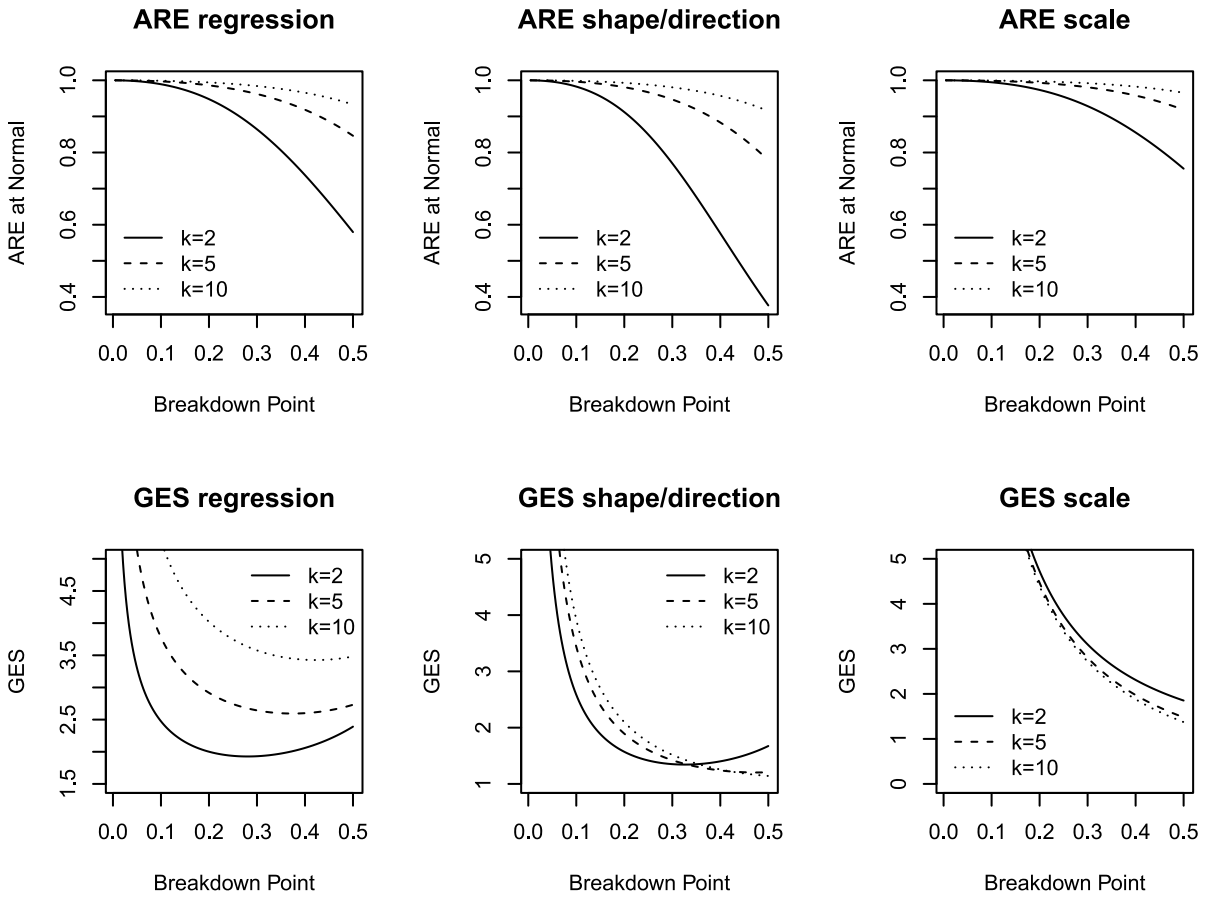

Fig. 1. Asymptotic efficiencies relative to the least squares estimator (ARE) and the gross error sensitivities (GES) for the S-estimators of the regression parameter β_0 (left), shape $\boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/k}$ and direction $\theta_0/\|\theta_0\|$ (middle), and scale $|\boldsymbol{\Sigma}|^{1/(2k)}$ (right) as functions of the breakdown point $\epsilon^* \in (0, 0.5)$ at the multivariate normal in dimensions $k = 2$ (solid), $k = 5$ (dashed) and $k = 10$ (dotted).

estimator, respectively. Similarly, the GES index of the shape and direction functionals attains its minimal value $G_2 = 1.344$ for $c = 3.722$. This would yield $G_1 = 1.947$, $G_3 = 2.844$, $1/\lambda = 0.835$, $1/\sigma_1 = 0.723$, $1/(4\sigma_3) = 0.912$ and breakdown point 33%. The GES index of the scale functional attains its minimum value $G_3 = 1.852$ at 50% breakdown point, so no simultaneous gain in efficiency and smaller GES values G_1 and G_2 can be achieved at the cost of a smaller breakdown point.

In dimension $k = 5$ (dashed lines), the 50% breakdown S-estimators have asymptotic efficiencies $1/\lambda = 0.864$, $1/\sigma_1 = 0.778$, and $1/(10\sigma_3) = 0.918$. The GES index of the regression functional attains its minimal value $G_1 = 2.595$ at breakdown point 37%. The corresponding GES index for shape and direction functionals is $G_2 = 1.271$ and $G_3 = 1.480$ for the scale functionals. Corresponding to this smaller regression GES index we observe a gain in the asymptotic efficiencies: $1/\lambda = 0.932$, $1/\sigma_1 = 0.903$, and $1/(10\sigma_3) = 0.965$, for the regression estimator, the estimators of shape and direction, and the scale estimator, respectively. The GES index of the shape and direction functionals attains its minimal value at breakdown point 47%, so the gain in both efficiency and a smaller G_2 value is negligible. The situation for the GES index for scale is the same as in dimension $k = 2$, where no simultaneous gain in efficiency and smaller GES values G_1 and G_2 can be achieved at the cost of a smaller breakdown point.

Finally, in dimension $k = 10$ (dotted lines), the 50% breakdown S-estimators have asymptotic efficiencies $1/\lambda = 0.933$, $1/\sigma_1 = 0.915$, and $1/(20\sigma_3) = 0.965$. The GES index of the regression functional attains its minimal value $G_1 = 3.426$ at breakdown point 42%. The corresponding GES index for shape and direction functionals is $G_2 = 1.221$ and $G_3 = 1.744$ for the scale functionals. Corresponding to this smaller regression GES index we observe a gain in the asymptotic efficiencies: $1/\lambda = 0.960$, $1/\sigma_1 = 0.949$, and $1/(20\sigma_3) = 0.979$, for the regression estimator, the estimators of shape and direction, and the scale estimator, respectively. Both GES indices G_2 and G_3 attain their minimal values at 50% breakdown, so no simultaneous gain in efficiency and smaller GES value G_1 can be achieved at the cost of a smaller breakdown point.

We conclude that at a moderate loss of breakdown point, from 50% to about 30%–40%, one can gain efficiency of the S-estimators and at the same time reduce the GES of the regression S-estimator. The improvements becomes less as the dimension increases.

In the top row of Fig. 1 we see that the efficiency becomes close to one when the dimension is large. This is a well know phenomenon observed when the efficiency is being computed under a multivariate normal setting. One of the referees raised the

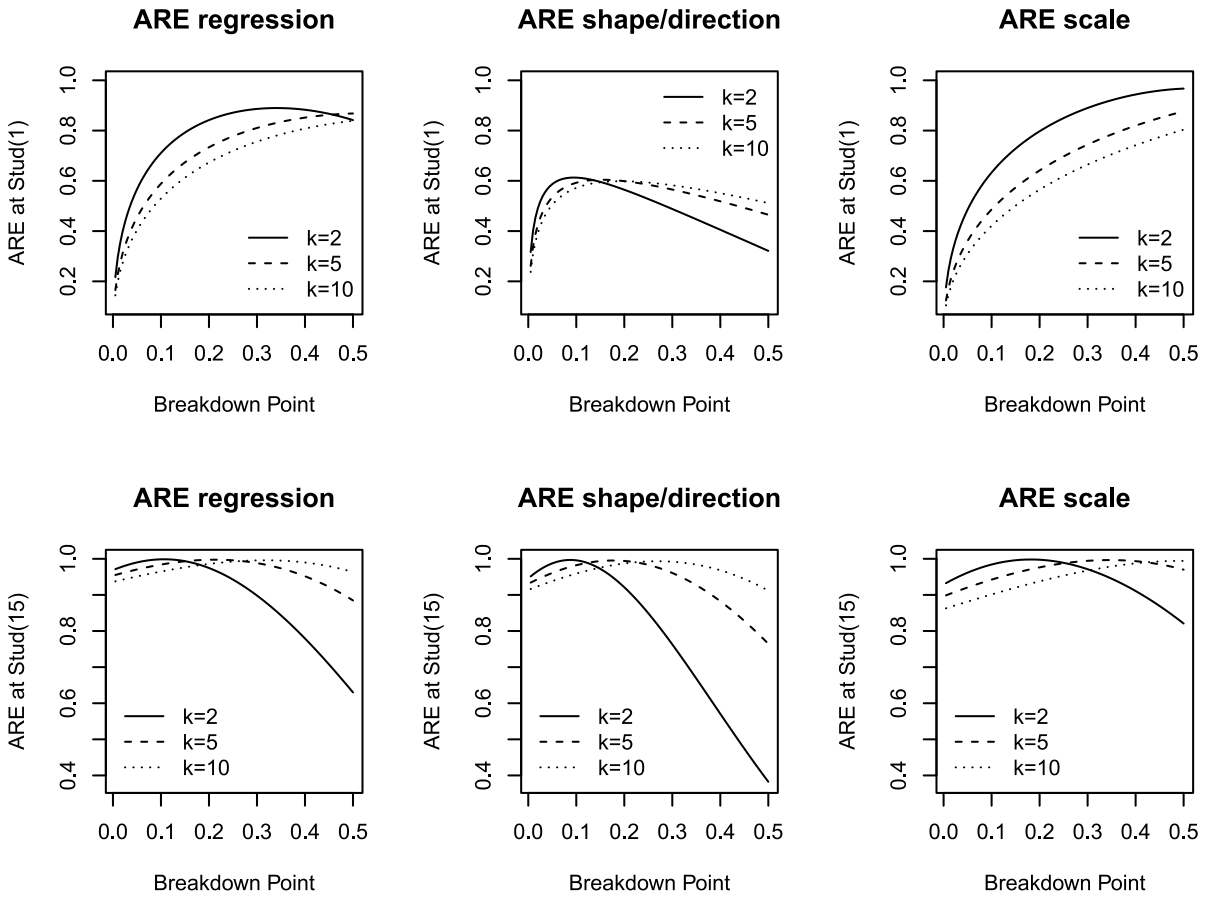

Fig. 2. Asymptotic efficiencies relative to the maximum likelihood estimator (ARE) for the S-estimators of the regression parameter $\boldsymbol{\beta}_0$ (left), shape $\boldsymbol{\Sigma}/|\boldsymbol{\Sigma}|^{1/k}$ and direction $\boldsymbol{\theta}_0/\|\boldsymbol{\theta}_0\|$ (middle), and scale $|\boldsymbol{\Sigma}|^{1/(2k)}$ (right) as functions of the breakdown point $\epsilon^* \in (0, 0.5)$ at the multivariate Student with $\nu \in \{1, 15\}$ degrees of freedom in dimensions $k = 2$ (solid), $k = 5$ (dashed) and $k = 10$ (dotted).

question whether this behavior would also be observed in a neighborhood of the normal distribution. We have addressed this question by computing asymptotic relative efficiencies with respect to the maximum likelihood estimator at the k-variate Student distribution with degrees of freedom $\nu = 1$ and $\nu = 15$. In these settings, the functions w_i, for $i \in \{1, 2, 3\}$ in (14) are given by $w_1(s) = w_2(s) = (\nu + k)/(\nu + s^2)$ and $w_3(s) = 1$, see also Example 2. From Maronna [22] one may obtain that the limiting variance of the regression ML estimator can be represented by the scalar

$$\lambda_{\mathrm{ML}} = \frac{(1/k)\mathbb{E}\left[w_1(\|\mathbf{z}\|)^2\|\mathbf{z}\|^2\right]}{\left(\mathbb{E}\left[w_1(\|\mathbf{z}\|) + \frac{1}{k}w_1'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\right)^2}.$$

The limiting variances of the ML estimators of shape and scale are represented by the scalars σ_1 and $\sigma_3 = (2\sigma_1/k + \sigma_2)/4$, respectively, where σ_1 and σ_2 are defined in (17).

The asymptotic efficiencies relative to the maximum likelihood estimator at the k-variate Student distribution with $\nu \in \{1, 15\}$ degrees of freedom are visible in Fig. 2. The graphs in the top row correspond to $\nu = 1$ and differ quite a lot from the ones the top row in Fig. 1. As expected the S-estimators with small breakdown points have poor efficiencies, because they behave similar to the least squares estimators. For higher breakdown points the efficiencies become higher, but in general remain far below one and do not necessarily reach their maximum value at 50% breakdown point, especially for the S-estimator of shape. The maximum efficiency 0.613 is obtained by the S-estimator of shape with breakdown point 9.5%. For the high breakdown S-estimator of shape the efficiency increases with larger dimension, but does not improve on the maximum value. For the regression and scale S-estimators the maximum efficiencies 0.890 and 0.967, respectively, are obtained by the S-estimators with breakdown points 34% and 5%, respectively. In contrast with the behavior at the normal distribution, in general the efficiencies of the regression and scale estimators decrease with larger dimension. This behavior can also be observed for some of the robust location M-estimators considered in Maronna [22].

For the Student distribution with $\nu = 15$ degrees of freedom, the behavior of the efficiency is somewhat in between the one at the normal and the one at the Student distribution with $\nu = 1$ degree of freedom. The graphs in the bottom row of Fig. 2 are more similar

to the ones in the top row of Fig. 1, although the maximum efficiencies are not attained at 0% breakdown point. For the regression, shape and scale S-estimators the maximal efficiencies 0.998, 0.996, and 0.998, respectively, are obtained by the S-estimators with breakdown points 10.5%, 9%, and 18%, respectively. Furthermore, only for higher breakdown points the efficiency increases with larger dimension. For the low breakdown point S-estimators the efficiency decreases with larger dimension. This behavior can also be observed for the 25% regression S-estimator in Van Aelst and Willems [29] at the Student distribution with 3 degrees of freedom.

Acknowledgments

I thank the two referees for their comments and suggestions, which led to a considerable improvement of the original version of the manuscript.

Appendix A. Proofs

Proof of Theorem 1. It can be seen that the projection matrix, as defined in (6), is given by

$$\mathbf{\Pi}_L = \mathbf{L}\left(\mathbf{L}^\top\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1}\mathbf{L}^\top\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right). \tag{A.1}$$

Since $\mathbf{N}$ is of radial type with respect to $\mathbf{\Sigma}$, it follows from Corollary 1 in Tyler [27] that there exist constants η, σ_1 and σ_2 with $\sigma_1 \geq 0$ and $\sigma_2 \geq -2\sigma_1/k$, such that $\mathbb{E}[\mathbf{N}] = \eta\mathbf{\Sigma}$ and

$$\mathrm{var}\{\mathrm{vec}(\mathbf{N})\} = \sigma_1(\mathbf{I}_{k^2} + \mathbf{K}_{k,k})(\mathbf{\Sigma}\otimes\mathbf{\Sigma}) + \sigma_2\mathrm{vec}(\mathbf{\Sigma})\mathrm{vec}(\mathbf{\Sigma})^\top.$$

It follows that $\mathbf{M}$ has expectation $\mathbb{E}[\mathrm{vec}(\mathbf{M})] = \mathbf{\Pi}_L\mathrm{vec}(\mathbb{E}[\mathbf{N}]) = \eta\mathbf{\Pi}_L\mathrm{vec}(\mathbf{\Sigma})$, and variance

$$\mathrm{var}(\mathrm{vec}(\mathbf{M})) = \mathbf{\Pi}_L\mathrm{var}(\mathrm{vec}(\mathbf{N}))\mathbf{\Pi}_L^\top = \sigma_1\mathbf{\Pi}_L(\mathbf{I}_{k^2} + \mathbf{K}_{k,k})(\mathbf{\Sigma}\otimes\mathbf{\Sigma})\mathbf{\Pi}_L^\top + \sigma_2\mathbf{\Pi}_L\mathrm{vec}(\mathbf{\Sigma})\mathrm{vec}(\mathbf{\Sigma})^\top\mathbf{\Pi}_L^\top.$$

Note that

$$\mathbf{K}_{k,k}(\mathbf{A}\otimes\mathbf{B}) = (\mathbf{B}\otimes\mathbf{A})\mathbf{K}_{k,k}, \quad \mathbf{K}_{k,k}\mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\mathbf{A}^\top),$$

see, e.g., [19, Chapter 3, Section 7]. Since $\mathbf{\Sigma} = \mathbf{V}(\boldsymbol{\theta})$ is symmetric, also $\mathbf{L}_j = \partial\mathbf{V}/\partial\theta_j$ is symmetric, for $j \in \{1,\ldots,\ell\}$. This means that $\mathbf{K}_{k,k}\mathbf{L} = \mathbf{L}$ and it follows that

$$\begin{aligned}\mathbf{\Pi}_L(\mathbf{I}_{k^2} + \mathbf{K}_{k,k})(\mathbf{\Sigma}\otimes\mathbf{\Sigma})\mathbf{\Pi}_L^\top &= \mathbf{L}\left(\mathbf{L}^\top\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1}\mathbf{L}^\top(\mathbf{I}_{k^2} + \mathbf{K}_{k,k})\mathbf{\Pi}_L^\top \\ &= 2\mathbf{L}\left(\mathbf{L}^\top\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1}\mathbf{L}^\top\mathbf{\Pi}_L^\top = 2\mathbf{L}\left(\mathbf{L}^\top\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1}\mathbf{L}^\top.\end{aligned} \tag{A.2}$$

This finishes the proof of part (i). Since $\mathbf{L}$ has full rank, it holds that $(\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\mathrm{vec}(\mathbf{M}) = \mathbf{T}$. This immediately gives

$$\mathbb{E}[\mathbf{T}] = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\mathbb{E}[\mathrm{vec}(\mathbf{M})] = \eta(\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{\Pi}_L\mathrm{vec}(\mathbf{\Sigma}) = \eta\boldsymbol{\theta}_L,$$

and

$$\mathrm{var}(\mathbf{T}) = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\mathrm{var}\{\mathrm{vec}(\mathbf{M})\}\mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1}.$$

When we insert the expression for $\mathrm{var}\{\mathrm{vec}(\mathbf{M})\}$ from part (i), the theorem follows. □

Proof of Corollary 1. When $\mathbf{V}$ is linear, then $\mathrm{vec}(\mathbf{\Sigma}) = \mathbf{L}\boldsymbol{\theta}$. This means that $\mathbf{\Pi}_L\mathrm{vec}(\mathbf{\Sigma}) = \mathrm{vec}(\mathbf{\Sigma})$ and $\boldsymbol{\theta}_L = \boldsymbol{\theta}$. The corollary then follows directly from Theorem 1. □

Proof of Theorem 2. The proof follows the line of reasoning used in the proofs of Theorem 9.1 and Corollary 9.2 in Lopuhaä et al. [17] for S-estimators. These proofs are based on estimating Eqs. (13) with $w_1(d) = \rho'(d)/d$, $w_2(d) = k\rho'(d)/d$ and $w_3(d) = \rho'(d)d - \rho(d) + b_0$, and require conditions (R1)–(R5) in [17] on the function ρ. For the proof of Theorem 2 these conditions have been reformulated into similar conditions (C1)–(C3) for general w_1, w_2, and w_3. Furthermore, in order to incorporate the case $w_1 = w_2 = w_3 = 1$ of Example 1, we have slightly adapted some of the boundedness conditions and use that

$$d^2 = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \leq \frac{\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2}{\lambda_k(\mathbf{V})} \leq \frac{(\|\mathbf{y}\| + \|\mathbf{X}\|\cdot\|\boldsymbol{\beta}\|)^2}{\lambda_k(\mathbf{V})} \leq \frac{\|\mathbf{s}\|^2(1+\|\boldsymbol{\beta}\|)^2}{\lambda_k(\mathbf{V})}.$$

This will ensure that d^2 is bounded by a multiple of $\|\mathbf{s}\|^2$ on a neighborhood of $\boldsymbol{\xi}_0$. In order to apply dominated convergence, we then require $\mathbb{E}\|\mathbf{s}\|^4 < \infty$ in Theorem 2 instead of $\mathbb{E}\|\mathbf{X}\|^2 < \infty$, which was sufficient for Corollary 9.2 in [17].

Define

$$\Lambda(\boldsymbol{\xi}) = \int \Psi(\mathbf{s},\boldsymbol{\xi})\,\mathrm{d}P(\mathbf{s}).$$

Since ξ_0 is a solution of (15), we have that $\Lambda(\xi_0) = \mathbf{0}$. Conditions (C1)–(C3) and (A2) yield that Λ is continuously differentiable in a neighborhood of ξ_0 and by application of empirical process theory (see, e.g., Lemma 11.8 in [18] for the special case of S-estimators) one finds

$$\begin{aligned}\mathbf{0} &= \int \Psi(\mathbf{s}, \xi_n)\, \mathrm{d}P(\mathbf{s}) + \int \Psi(\mathbf{s}, \xi_0)\, \mathrm{d}(\mathbb{P}_n - P)(\mathbf{s}) + o_P(n^{-1/2}) \\ &= \Lambda'(\xi_0)(\xi_n - \xi_0) + \frac{1}{n}\sum_{i=1}^{n}\Big\{\Psi(\mathbf{s}_i, \xi_0) - \mathbb{E}[\Psi(\mathbf{s}_i, \xi_0)]\Big\} + o_P(n^{-1/2}).\end{aligned} \tag{A.3}$$

Similar to Lemma 8.3 in Lopuhaä et al. [17], we find that $\Lambda'(\xi_0)$ is a block matrix, with blocks $\Lambda'_{\beta}(\xi_0)$ and $\Lambda'_{\theta}(\xi_0)$ on the main diagonal, where

$$\Lambda'_{\theta}(\xi_0) = \int \frac{\partial \Psi_{\theta}(\mathbf{s}, \xi_0)}{\partial \boldsymbol{\theta}}\, \mathrm{d}P(\mathbf{s}) = \gamma_1 \mathbf{L}^{\top}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{L} - \gamma_2 \mathbf{L}^{\top}\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^{\top}\mathbf{L},$$

where Ψ_{θ} is defined in (14), and γ_1, γ_2 defined in (16). This implies that $\sqrt{n}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_0)$ and $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ are asymptotically independent and from (A.3) we obtain

$$\mathbf{0} = \Lambda'_{\theta}(\xi_0)(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) + \frac{1}{n}\sum_{i=1}^{n}\Big\{\Psi_{\theta}(\mathbf{s}_i, \xi_0) - \mathbb{E}[\Psi_{\theta}(\mathbf{s}_i, \xi_0)]\Big\} + o_P(n^{-1/2}),$$

where Ψ_{θ} is defined in (14). Due to condition (C1) and $\mathbb{E}\|\mathbf{s}\|^4 < \infty$, the second term on the right hand side behaves according to the central limit theorem. This yields that $\boldsymbol{\theta}_n - \boldsymbol{\theta}_0 = O_P(n^{-1/2})$ and it follows that

$$\sqrt{n}(\mathrm{vec}(\mathbf{V}_n) - \mathrm{vec}(\boldsymbol{\Sigma})) = \sqrt{n}\,\mathbf{L}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) + O_P(n^{-1/2}) = -\mathbf{L}\Lambda'_{\theta}(\xi_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_{\theta}(\mathbf{s}_i, \xi_0) + o_P(1), \tag{A.4}$$

where we also use that $\mathbb{E}[\Psi_{\theta}(\mathbf{s}_i, \xi_0)] = \Lambda(\xi_0) = \mathbf{0}$. Furthermore, we can write

$$\Psi_{\theta}(\mathbf{s}, \xi_0) = \mathbf{L}^{\top}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathrm{vec}\left\{\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0)\right\}, \tag{A.5}$$

where

$$\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0) = w_2(d_0)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)^{\top} - w_3(d_0)\boldsymbol{\Sigma}, \tag{A.6}$$

with $d_0^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)$. To determine an expression for the inverse of $\Lambda'_{\theta}(\xi_0)$, note that we can write $\Lambda'_{\theta}(\xi_0) = \gamma_1\mathbf{B} - \gamma_2\mathbf{v}\mathbf{v}^{\top}$, where

$$\mathbf{B} = \mathbf{L}^{\top}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{L}, \quad \mathbf{v} = \mathbf{L}^{\top}\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) = \mathbf{L}^{\top}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{\Sigma}),$$

where we also use

$$\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^{\top} \otimes \mathbf{A})\mathrm{vec}(\mathbf{B}), \tag{A.7}$$

see, e.g., [19, Chapter 2, Section 4]. According to the Sherman–Morisson formula, this means that

$$\Lambda'_{\theta}(\xi_0)^{-1} = \frac{1}{\gamma_1}\mathbf{B}^{-1} - \frac{\gamma_2}{\gamma_1^2}\frac{\mathbf{B}^{-1}\mathbf{v}\mathbf{v}^{\top}\mathbf{B}^{-1}}{1 - (\gamma_2/\gamma_1)\mathbf{v}^{\top}\mathbf{B}^{-1}\mathbf{v}} = \frac{1}{\gamma_1}\mathbf{B}^{-1} - \frac{\gamma_2}{\gamma_1}\frac{\mathbf{B}^{-1}\mathbf{v}\mathbf{v}^{\top}\mathbf{B}^{-1}}{\gamma_1 - \gamma_2\mathbf{v}^{\top}\mathbf{B}^{-1}\mathbf{v}}.$$

Note that $\mathbf{v}^{\top}\mathbf{B}^{-1}\mathbf{v} = \pi_L$, where π_L is defined in (16). We find that $\mathbf{L}\Lambda'_{\theta}(\xi_0)^{-1}\Psi_{\theta}(\mathbf{s}, \xi_0) = \mathbf{L}\Lambda'_{\theta}(\xi_0)^{-1}\mathbf{L}^{\top}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathrm{vec}\left\{\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0)\right\}$, where

$$\begin{aligned}\mathbf{L}\Lambda'_{\theta}(\xi_0)^{-1}\mathbf{L}^{\top}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) &= \frac{1}{\gamma_1}\boldsymbol{\Pi}_L + \frac{\gamma_2}{\gamma_1(\gamma_1 - \gamma_2\pi_L)}\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma})(\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}))^{\top}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \\ &= \boldsymbol{\Pi}_L\left\{\frac{1}{\gamma_1}\mathbf{I}_{k^2} + \frac{\gamma_2}{\gamma_1(\gamma_1 - \gamma_2\pi_L)}\mathrm{vec}(\boldsymbol{\Sigma})(\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}))^{\top}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\right\}.\end{aligned}$$

It follows that

$$\sqrt{n}(\mathrm{vec}(\mathbf{V}_n) - \mathrm{vec}(\boldsymbol{\Sigma})) = -\boldsymbol{\Pi}_L(a\mathbf{I}_{k^2} + b\mathbf{M}_L)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{vec}\left\{\Psi_{\mathbf{C}}(\mathbf{s}_i, \xi_0)\right\} + o_P(1), \tag{A.8}$$

where

$$a = \frac{1}{\gamma_1}, \quad b = \frac{\gamma_2}{\gamma_1(\gamma_1 - \gamma_2\pi_L)}, \tag{A.9}$$

with π_L defined in (16), and where

$$\mathbf{M}_L = \mathrm{vec}(\boldsymbol{\Sigma})(\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}))^{\top}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}), \tag{A.10}$$

with $d_{i,0}^2 = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_0)^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_0)$, $i \in \{1, \ldots, n\}$.

Next, note that $\Lambda(\xi_0) = \mathbf{0}$, together with (A.5) and the fact that $\mathbf{L}$ has full rank, yields that $\mathbb{E}\left[\mathrm{vec}\left\{\Psi_{\mathbf{C}}(\mathbf{s}_i, \xi_0)\right\}\right] = \mathbf{0}$. This means that $\sqrt{n}(\mathrm{vec}(\mathbf{V}_n) - \mathrm{vec}(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and variance

$$\mathbb{E}\left[\mathrm{vec}(\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0))\mathrm{vec}(\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0))^{\top}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathrm{vec}(\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0))\mathrm{vec}(\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0))^{\top}\middle|\mathbf{X}\right]\right].$$

The inner expectation on the right hand side is the conditional expectation of $\mathbf{y} \mid \mathbf{X}$, which has the same distribution as $\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}$, where $\mathbf{z}$ has a spherical density $f_{\mathbf{0},\mathbf{I}_k}(\mathbf{z}) = g(\|\mathbf{z}\|^2)$. This implies that

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\boldsymbol{\Sigma}^{-1/2}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\boldsymbol{\Sigma}^{-1/2}\right)^\top\right] \\
&\quad = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\operatorname{vec}\left(\mathbf{u}\mathbf{u}^\top\right)\operatorname{vec}\left(\mathbf{u}\mathbf{u}^\top\right)^\top\right] - \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)w_3(\|\mathbf{z}\|)\|\mathbf{z}\|^2\right]\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\operatorname{vec}\left(\mathbf{u}\mathbf{u}^\top\right)\operatorname{vec}\left(\mathbf{I}_k\right)^\top\right] \\
&\quad - \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)w_3(\|\mathbf{z}\|)\|\mathbf{z}\|^2\right]\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\operatorname{vec}\left(\mathbf{I}_k\right)\operatorname{vec}\left(\mathbf{u}\mathbf{u}^\top\right)^\top\right] + \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_3(\|\mathbf{z}\|)^2\right]\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\operatorname{vec}\left(\mathbf{I}_k\right)\operatorname{vec}\left(\mathbf{I}_k\right)^\top\right],
\end{aligned} \tag{A.11}$$

where $\mathbf{u} = \mathbf{z}/\|\mathbf{z}\|$. From Lemma 5.1 in [14], we have

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

where $\sigma_1 = \sigma_2 = (k(k+2))^{-1}$. Hence, the first term on the right hand side of (A.11) is equal to

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{k(k+2)}\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k} + \operatorname{vec}(\mathbf{I}_k)\operatorname{vec}(\mathbf{I}_k)^\top\right).$$

For the left hand side of (A.11), this leads to a term $\mathbf{I}_{k^2} + \mathbf{K}_{k,k}$ with coefficient

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{k(k+2)},$$

and using that, according to Lemma 11.4 in [18], $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\mathbf{u}\mathbf{u}^\top\right] = (1/k)\mathbf{I}_k$, for the left hand side of (A.11), we find a second term $\operatorname{vec}(\mathbf{I}_k)\operatorname{vec}(\mathbf{I}_k)^\top$ with coefficient

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{k(k+2)} - \frac{2\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)w_3(\|\mathbf{z}\|)\|\mathbf{z}\|^2\right]}{k} + \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_3(\|\mathbf{z}\|)^2\right].$$

This means that

$$\mathbb{E}\left[\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\boldsymbol{\Sigma}^{-1/2}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\boldsymbol{\Sigma}^{-1/2}\right)^\top\right] = \delta_1\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right) + \delta_2\operatorname{vec}(\mathbf{I}_k)\operatorname{vec}(\mathbf{I}_k)^\top,$$

or equivalently

$$\mathbb{E}\left[\operatorname{vec}(\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0))\operatorname{vec}(\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0))^\top\right] = \delta_1\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \delta_2\operatorname{vec}(\boldsymbol{\Sigma})\operatorname{vec}(\boldsymbol{\Sigma})^\top. \tag{A.12}$$

where

$$\delta_1 = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{k(k+2)}, \quad \delta_2 = \delta_1 - \frac{2\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)w_3(\|\mathbf{z}\|)\|\mathbf{z}\|^2\right]}{k} + \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_3(\|\mathbf{z}\|)^2\right]. \tag{A.13}$$

The limiting variance of $\sqrt{n}(\operatorname{vec}(\mathbf{V}_n) - \operatorname{vec}(\boldsymbol{\Sigma}))$ then becomes

$$\begin{aligned}
&\boldsymbol{\Pi}_L\left(a\mathbf{I}_{k^2} + b\mathbf{M}_L\right)\mathbb{E}\left[\operatorname{vec}(\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0))\operatorname{vec}(\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0))^\top\right]\left(a\mathbf{I}_{k^2} + b\mathbf{M}_L^\top\right)\boldsymbol{\Pi}_L^\top \\
&\quad = \delta_1\boldsymbol{\Pi}_L\left(a\mathbf{I}_{k^2} + b\mathbf{M}_L\right)\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\left(a\mathbf{I}_{k^2} + b\mathbf{M}_L^\top\right)\boldsymbol{\Pi}_L^\top \\
&\qquad + \delta_2\boldsymbol{\Pi}_L\left(a\mathbf{I}_{k^2} + b\mathbf{M}_L\right)\operatorname{vec}(\boldsymbol{\Sigma})\operatorname{vec}(\boldsymbol{\Sigma})^\top\left(a\mathbf{I}_{k^2} + b\mathbf{M}_L^\top\right)\boldsymbol{\Pi}_L^\top,
\end{aligned} \tag{A.14}$$

where a and b are defined in (A.9). First consider the second term on the right hand side of (A.14). Because $\mathbf{M}_L\operatorname{vec}(\boldsymbol{\Sigma}) = \pi_L\operatorname{vec}(\boldsymbol{\Sigma})$, where $\mathbf{M}_L$ and π_L are defined in (16) and (A.10), the second term in (A.14) is equal to

$$\delta_2(a + b\pi_L)^2\boldsymbol{\Pi}_L\operatorname{vec}(\boldsymbol{\Sigma})\operatorname{vec}(\boldsymbol{\Sigma})^\top\boldsymbol{\Pi}_L^\top.$$

Next consider the first term on the right hand side of (A.14). The matrix product after the factor δ_1 is equal to

$$\begin{aligned}
&a^2\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + ab\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{M}_L^\top \\
&\qquad + ba\mathbf{M}_L\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + b^2\mathbf{M}_L\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{M}_L^\top.
\end{aligned}$$

We have that $\mathbf{M}_L(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) = \operatorname{vec}(\boldsymbol{\Sigma})(\boldsymbol{\Pi}_L\operatorname{vec}(\boldsymbol{\Sigma}))^\top$, and since $\boldsymbol{\Pi}_L^\top(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{\Pi}_L = (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{\Pi}_L$, we also find that $\mathbf{M}_L(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{M}_L^\top = \pi_L\operatorname{vec}(\boldsymbol{\Sigma})\operatorname{vec}(\boldsymbol{\Sigma})^\top$, where $\mathbf{M}_L$ and π_L are defined in (16) and (A.10). Since $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta})$ is symmetric, also $\mathbf{L}_j = \partial\mathbf{V}/\partial\theta_j$ is symmetric, for $j \in \{1,\ldots,\ell\}$. This means that $\mathbf{K}_{k,k}\mathbf{L} = \mathbf{L}$ and $\mathbf{K}_{k,k}\boldsymbol{\Pi}_L = \boldsymbol{\Pi}_L$. Furthermore, $\mathbf{K}_{k,k}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) = (\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\mathbf{K}_{k,k}$. Hence, it follows that the first term on the right hand side of (A.14) is equal to

$$\begin{aligned}
&a^2\delta_1\boldsymbol{\Pi}_L\left(\mathbf{I}_{k^2} + \mathbf{K}_{k,k}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\boldsymbol{\Pi}_L^\top + 2ab\delta_1\boldsymbol{\Pi}_L\left\{\boldsymbol{\Pi}_L\operatorname{vec}(\boldsymbol{\Sigma})\operatorname{vec}(\boldsymbol{\Sigma})^\top + \operatorname{vec}(\boldsymbol{\Sigma})\operatorname{vec}(\boldsymbol{\Sigma})^\top\boldsymbol{\Pi}_L^\top\right\}\boldsymbol{\Pi}_L^\top \\
&\qquad + 2\pi_L b^2\delta_1\boldsymbol{\Pi}_L\operatorname{vec}(\boldsymbol{\Sigma})\operatorname{vec}(\boldsymbol{\Sigma})^\top\boldsymbol{\Pi}_L^\top,
\end{aligned}$$

where a, b, δ_1, and δ_2 are defined in (A.9) and (A.13). Therefore, together with (A.2), the limiting variance of $\sqrt{n}(\operatorname{vec}(\mathbf{V}_n) - \operatorname{vec}(\boldsymbol{\Sigma}))$ becomes

$$2\sigma_1\mathbf{L}\left(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1}\mathbf{L}^\top + \sigma_2\boldsymbol{\Pi}_L\operatorname{vec}(\boldsymbol{\Sigma})\operatorname{vec}(\boldsymbol{\Sigma})^\top\boldsymbol{\Pi}_L^\top,$$

where

$$\sigma_1 = a^2\delta_1 = \frac{\delta_1}{\gamma_1^2} = \frac{k(k+2)\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{\left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3 + k(k+2)w_3(\|\mathbf{z}\|)\right]\right)^2}$$

and

$$\sigma_2 = 4ab\delta_1 + 2\pi_L b^2\delta_1 + \delta_2(a+b\pi_L)^2.$$

For convenience, we first write $\delta_2 = -2\delta_1/k + \delta_3$, where δ_1, δ_2 are defined in (A.13) and

$$\delta_3 = \frac{1}{k^2}\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(w_2(\|\mathbf{z}\|)\|\mathbf{z}\|^2 - kw_3(\|\mathbf{z}\|)\right)^2\right],$$

and $\sigma_2 = -2\sigma_1/k + \sigma_3$, where

$$\sigma_3 = 2b\delta_1(2a+b\pi_L)\left(1-\frac{\pi_L}{k}\right) + \delta_3(a+b\pi_L)^2.$$

Furthermore, $a + b\pi_L = 1/(\gamma_1 - \gamma_2\pi_L)$ and

$$2a + b\pi_L = \frac{2}{\gamma_1} + \frac{\gamma_2\pi_L}{\gamma_1(\gamma_1-\gamma_2\pi_L)} = \frac{2\gamma_1 - \gamma_2\pi_L}{\gamma_1(\gamma_1-\gamma_2\pi_L)},$$

so that

$$\sigma_3 = \frac{2\delta_1\gamma_2(2\gamma_1-\gamma_2\pi_L)}{\gamma_1^2(\gamma_1-\gamma_2\pi_L)^2}\left(1-\frac{\pi_L}{k}\right) + \frac{\delta_3}{(\gamma_1-\gamma_2\pi_L)^2}.$$

Writing γ_3 and γ_4 instead of δ_1 and δ_3, respectively, proves part (i) of the theorem.

From expansion (A.4) it follows immediately that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ is asymptotically normal with mean zero and variance

$$(\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\left\{2\sigma_1\mathbf{L}\left(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1}\mathbf{L}^\top + \sigma_2\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top\boldsymbol{\Pi}_L^\top\right\}\mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1}.$$

Let $\boldsymbol{\theta}_L$ be such that $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_L$, i.e., $\boldsymbol{\theta}_L = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma})$. Then the limiting variance of $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ can be written as

$$2\sigma_1\left(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1} + \sigma_2\boldsymbol{\theta}_L\boldsymbol{\theta}_L^\top.$$

This finishes the proof. □

Proof of Corollary 2. When $\mathbf{V}$ is linear, then $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}) = \mathrm{vec}(\boldsymbol{\Sigma})$ and $\pi_L = k$, where π_L is defined in (16). It follows from Theorem 2 that $\sqrt{n}(\mathrm{vec}(\mathbf{V}_n) - \mathrm{vec}(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and variance (3). Furthermore, because $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}) = \mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_0$, it follows that $\boldsymbol{\theta}_L$ in Theorem 2 is equal to $\boldsymbol{\theta}_0$. It follows from Theorem 2 that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ is asymptotically normal with mean zero and variance (2). Finally, since $\pi_L = k$, from the expression in (17), it follows that $\sigma_2 = -2\sigma_1/k + \sigma_3$, where $\sigma_3 = \gamma_4/(\gamma_1 - k\gamma_2)^2$, where γ_1, γ_2, and γ_4 are defined in (16) and (18). We find

$$\gamma_1 - k\gamma_2 = \frac{1}{2k}\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3 + 2kw_3(\|\mathbf{z}\|) - kw_3'(\|\mathbf{z}\|)\|\mathbf{z}\|\right].$$

so that

$$\sigma_3 = \frac{4\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(w_2(\|\mathbf{z}\|)\|\mathbf{z}\|^2 - kw_3(\|\mathbf{z}\|)\right)^2\right]}{\left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3 + 2kw_3(\|\mathbf{z}\|) - kw_3'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\right)^2}. \quad \square$$

Proof of Remark 1. Consider the expansion (A.8). When $\mathbf{V}$ is linear, we find

$$\begin{aligned}\boldsymbol{\Pi}_L(a\mathbf{I}_{k^2} + b\mathbf{M}_L)\mathrm{vec}\left\{\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\right\} &= a\boldsymbol{\Pi}_L\mathrm{vec}\left\{\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\right\} + b\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^\top\mathrm{vec}\left\{\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\right\}\\ &= a\boldsymbol{\Pi}_L\mathrm{vec}\left\{\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\right\} + b\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\right\} = a\boldsymbol{\Pi}_L\mathrm{vec}\left\{\boldsymbol{\Psi}_{\mathbf{C}}(\mathbf{s},\boldsymbol{\xi}_0)\right\} + b\mathrm{vec}(\boldsymbol{\Sigma})\left(w_2(d_0)d_0^2 - kw_3(d_0)\right).\end{aligned}$$

Using once more that $\boldsymbol{\Pi}_L\mathrm{vec}(\boldsymbol{\Sigma}) = \mathrm{vec}(\boldsymbol{\Sigma})$, we conclude that the right hand side can be written as $\boldsymbol{\Pi}_L\mathrm{vec}\left\{\boldsymbol{\Psi}_{\mathbf{N}}(\mathbf{s},\boldsymbol{\xi}_0)\right\}$, where

$$\boldsymbol{\Psi}_{\mathbf{N}}(\mathbf{s},\boldsymbol{\xi}) = v_1(d)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top - v_2(d)\boldsymbol{\Sigma}, \tag{A.15}$$

with $d^2 = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$ and

$$v_1(s) = aw_2(s), \quad v_2(s) = -bw_2(s)s^2 + (a+bk)w_3(s), \tag{A.16}$$

where a and b are defined in (A.9). Hence, if we define

$$\mathbf{N}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Psi}_{\mathbf{N}}(\mathbf{s}_i,\boldsymbol{\xi}_0),$$

with $\Psi_{\mathbf{N}}$ defined in (A.15), then it follows that

$$\sqrt{n}(\text{vec}(\mathbf{V}_n) - \text{vec}(\boldsymbol{\Sigma})) = -\boldsymbol{\Pi}_L \text{vec}\left\{\sqrt{n}(\mathbf{N}_n - \mathbb{E}[\mathbf{N}_n])\right\} + o_P(1).$$

This proves the first claim in Remark 1.

To prove the second claim, note that from $\Lambda(\xi_0) = \mathbf{0}$, together with (A.5) and (A.7), it follows that

$$0 = \boldsymbol{\theta}_0^\top \mathbb{E}\left[\Psi_{\boldsymbol{\theta}}(\mathbf{s}, \xi_0)\right] = \mathbb{E}\left[\text{vec}(\boldsymbol{\Sigma}^{-1})^\top \text{vec}\left\{\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0)\right\}\right] = \mathbb{E}\left[\text{tr}\left(\boldsymbol{\Sigma}^{-1}\Psi_{\mathbf{C}}(\mathbf{s}, \xi_0)\right)\right] = \mathbb{E}\left[w_2(d_0)d_0^2 - kw_3(d_0)\right],$$

where $\Psi_{\mathbf{C}}$ is defined in (A.6). Then, from the properties of elliptically contoured densities, together with (A.16), one finds $\mathbb{E}[\Psi_{\mathbf{N}}(\mathbf{s}, \xi_0)] = \mathbf{0}$. This means that $\sqrt{n}(\mathbf{N}_n - \mathbb{E}[\mathbf{N}_n])$ is asymptotically normal with mean zero and variance $\mathbb{E}\left[\text{vec}(\Psi_{\mathbf{N}}(\mathbf{s}, \xi_0))\text{vec}(\Psi_{\mathbf{N}}(\mathbf{s}, \xi_0))^\top\right]$. Similar to (A.12) one finds that this variance is of the form (1). □

Proof of Theorem 3. Let $H : \mathbb{R}^{k\times k} \to \mathbb{R}^m$ and let

$$H'(\mathbf{V}) = \frac{\partial H(\mathbf{V})}{\partial \text{vec}(\mathbf{V})^\top} = \left(\frac{\partial H_i(\mathbf{V})}{\partial v_{st}}\right)_{i\in\{1,\ldots,m\};\, s,t\in\{1,\ldots,k\}} \tag{A.17}$$

be the $m \times k^2$ matrix of partial derivatives. According to the delta method $\sqrt{n}(H(\mathbf{V}_n) - H(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and variance $H'(\boldsymbol{\Sigma})\text{var}\{\text{vec}(\mathbf{M})\}H'(\boldsymbol{\Sigma})^\top$. Because H is continuously differentiable and satisfies (21), it follows that

$$\sum_{j=1}^{l} v_j \frac{\partial H(\mathbf{v})}{\partial v_j} = \mathbf{0}. \tag{A.18}$$

This means that $H'(\boldsymbol{\Sigma})\text{vec}(\boldsymbol{\Sigma}) = \mathbf{0}$. Then, after inserting (3) for $\text{var}\{\text{vec}(\mathbf{M})\}$ (see Corollary 2), this finishes the proof of part (i).

For part (ii), let $H : \mathbb{R}^\ell \to \mathbb{R}^m$, and let

$$H'(\boldsymbol{\theta}) = \frac{\partial H(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} = \left(\frac{\partial H_i(\boldsymbol{\theta})}{\partial \theta_j}\right)_{i\in\{1,\ldots,m\};\, j\in\{1,\ldots,\ell\}} \tag{A.19}$$

be the $m\times\ell$ matrix of partial derivatives. According to Corollary 2 and the delta method $\sqrt{n}(H(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}_0))$ is asymptotically normal with mean zero and variance

$$H'(\boldsymbol{\theta}_0)\left\{2\sigma_1\left(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\right)^{-1} + \sigma_2\boldsymbol{\theta}_0\boldsymbol{\theta}_0^\top\right\}H'(\boldsymbol{\theta}_0)^\top.$$

Because H satisfies (21) and (A.18), it follows immediately that $H'(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 = \mathbf{0}$. This finishes the proof of part (ii). □

Proof of Lemma 1. We apply Lemma 1 in [5]. Although the lemma is established for the $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, the proof holds for any distribution with an elliptically contoured density. According to [5], there exist two functions $\alpha_C, \beta_C : [0,\infty) \to \mathbb{R}$, such that

$$\text{IF}(\mathbf{y}; \mathbf{C}, P_{\boldsymbol{\mu},\boldsymbol{\Sigma}}) = \alpha_C(d(\mathbf{y}))(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^\top - \beta_C(d(\mathbf{y}))\boldsymbol{\Sigma}. \tag{A.20}$$

We have that

$$\begin{aligned}\text{IF}(\mathbf{y}; \text{vec}(\mathbf{M}), P_{\boldsymbol{\mu},\boldsymbol{\Sigma}}) &= \lim_{h\downarrow 0}\frac{\text{vec}(\mathbf{M}((1-h)P_{\boldsymbol{\mu},\boldsymbol{\Sigma}} + h\delta_{\mathbf{y}})) - \text{vec}(\mathbf{M})(P_{\boldsymbol{\mu},\boldsymbol{\Sigma}})}{h}\\ &= \boldsymbol{\Pi}_L \lim_{h\downarrow 0}\frac{\text{vec}(\mathbf{C}((1-h)P_{\boldsymbol{\mu},\boldsymbol{\Sigma}} + h\delta_{\mathbf{y}})) - \text{vec}(\mathbf{C})(P_{\boldsymbol{\mu},\boldsymbol{\Sigma}})}{h} = \boldsymbol{\Pi}_L \text{vec}\left(\text{IF}(\mathbf{y}; \mathbf{C}, P_{\boldsymbol{\mu},\boldsymbol{\Sigma}})\right).\end{aligned}$$

When we insert the expression (A.1) for $\boldsymbol{\Pi}_L$, together with (A.20) and the fact that $(\mathbf{B}^\top\otimes\mathbf{A})\text{vec}(\mathbf{v}\mathbf{v}^\top) = \text{vec}(\mathbf{A}\mathbf{v}\mathbf{v}^\top\mathbf{B})$ according to (A.7), this finishes the proof of part (i). Since $\mathbf{L}$ has full column rank, $(\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\text{vec}(\mathbf{M}(P)) = \boldsymbol{\theta}(P)$, which yields

$$\text{IF}(\mathbf{y}; \boldsymbol{\theta}, P_{\boldsymbol{\mu},\boldsymbol{\Sigma}}) = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\text{IF}(\mathbf{y}; \text{vec}(\mathbf{M}), P_{\boldsymbol{\mu},\boldsymbol{\Sigma}}).$$

Part (i), together with (A.1) finishes the proof of part (ii). □

Proof of Lemma 2. Let $H : \mathbb{R}^{k\times k} \to \mathbb{R}^m$ with derivative H' defined in (A.17). From the definition of the influence function, it follows that

$$\text{IF}(\mathbf{y}; H(\mathbf{M}), P) = \left.\frac{\partial H(\mathbf{M}(P_{h,\mathbf{y}}))}{\partial h}\right|_{h=0} = \left.\frac{\partial H(\mathbf{C})}{\partial \text{vec}(\mathbf{C})^\top}\right|_{\mathbf{C}=\mathbf{M}(P)}\left.\frac{\partial \text{vec}(\mathbf{M}(P_{h,\mathbf{y}}))}{\partial h}\right|_{h=0} = H'(\mathbf{M}(P))\cdot\text{IF}(\mathbf{y}; \text{vec}(\mathbf{M}), P). \tag{A.21}$$

By Fisher consistency we have $\mathbf{M}(P) = \boldsymbol{\Sigma}$, and since $\mathbf{V}$ is linear, the expression in Lemma 1(i) holds with $\boldsymbol{\Pi}_L\text{vec}(\boldsymbol{\Sigma}) = \text{vec}(\boldsymbol{\Sigma})$. After inserting this in the right hand side of (A.21), together with $\text{vec}(\mathbf{v}\mathbf{v}^\top) = \mathbf{v}\otimes\mathbf{v}$ and (A.18), this proves part (i). Next, let $H : \mathbb{R}^\ell \to \mathbb{R}^m$ with derivative H' defined by (A.19). It follows that

$$\text{IF}(\mathbf{y}; H(\boldsymbol{\theta}), P) = H'(\boldsymbol{\theta}(P))\cdot\text{IF}(\mathbf{y}; \boldsymbol{\theta}, P). \tag{A.22}$$

By Fisher consistency we have $\boldsymbol{\theta}(P) = \boldsymbol{\theta}_0$, and since $\mathbf{V}$ is linear, the expression in Lemma 1(ii) holds with $\boldsymbol{\theta}_L = \boldsymbol{\theta}_0$. After inserting this on the right hand side of (A.22), together with (A.18), this proves part (ii). □

Appendix B. Derivation of σ_1 and σ_2

We compare the expressions for σ_1 and σ_2 derived in Theorem 2 with the ones obtained for specific cases in Tyler [27] and Lopuhaä et al. [17].

Proof of Example 1. Inserting $w_1 = w_2 = w_3 = 1$ in the expressions for σ_1 and σ_2 in Theorem 2 gives

$$\sigma_1 = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\|\mathbf{z}\|^4\right]}{k(k+2)},$$

which equals 1 for the multivariate normal. Furthermore, since $\gamma_1 = 1$ and $\gamma_2 = 0$ in (16), we find

$$\sigma_2 = -\frac{2}{k} + \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[(\|\mathbf{z}\|^2 - k)^2\right]}{k^2} = -\frac{2}{k} + \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\|\mathbf{z}\|^4\right] - 2k\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\|\mathbf{z}\|^2\right] + k^2}{k^2} = -\frac{2}{k} + \frac{k(k+2) - 2k^2 + k^2}{k^2} = 0. \qquad \square$$

Proof of Example 2. First consider the special case of maximum likelihood, with $w_1(s) = w_2(s) = -2g'(s^2)/g(s^2)$ and $w_3(s) = 1$. Note that

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[z(\|\mathbf{z}\|)\right] = \frac{2\pi^{k/2}}{\Gamma(k/2)}\int_0^\infty z(r)g(r^2)r^{k-1}\,\mathrm{d}r,$$

see, e.g., Lemma 1 in Lopuhaä [15]. When $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[\|\mathbf{z}^2\|] < \infty$, then by means of integration by parts we get

$$\begin{aligned}\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3\right] &= \frac{2\pi^{k/2}}{\Gamma(k/2)}\int_0^\infty \frac{4g'(r^2)^2}{g(r^2)^2}g(r^2)r^{k+3}\mathrm{d}r - \frac{2\pi^{k/2}}{\Gamma(k/2)}\int_0^\infty \frac{4g''(r^2)}{g(r^2)}g(r^2)r^{k+3}\mathrm{d}r \\ &= 4\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right] - k(k+2).\end{aligned}$$

It follows that

$$\sigma_1 = \frac{k(k+2)\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{\left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3] + k(k+2)\right)^2} = \frac{k(k+2)}{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}, \tag{B.1}$$

which coincides with the expression found in Example 2 in Tyler [27], who expresses expectations in terms of the random variable $T = \|\mathbf{z}\|^2$. To compute σ_2 in Corollary 2, first note that by means of integration by parts it follows that $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[w_2(\|\mathbf{z}\|)\|\mathbf{z}\|^2] = k$. When we insert this in the expression for σ_2, this gives

$$\sigma_2 = -\frac{2}{k}\sigma_1 + \frac{4\left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right] - k^2\right)}{\left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4] - k(k+2) + 2k\right)^2} = -\frac{2}{k}\sigma_1 + \frac{4}{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4] - k^2}.$$

After inserting $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4] = k(k+2)/\sigma_1$, as follows from (B.1), we find

$$\sigma_2 = -\frac{2}{k}\sigma_1 + \frac{4}{k(k+2)/\sigma_1 - k^2} = \frac{2\sigma_1(1-\sigma_1)}{k+2-k\sigma_1},$$

which coincides with the expression found in Example 2 in Tyler [27].

Next, consider the general case of M-estimators, with $w_3 = 1$. First note that Tyler [27] uses a function u_2, which relates to our function w_2 as $w_2(s) = u_2(s^2)$. Then, since $\boldsymbol{\xi}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ satisfies (15), we find that

$$\begin{aligned}0 = \boldsymbol{\theta}_0^\top\mathbb{E}\left[\boldsymbol{\Psi}_\theta(\mathbf{s},\boldsymbol{\xi}_0)\right] &= \mathbb{E}\left[\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^\top\mathrm{vec}\left\{w_2(d_0)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)^\top - \boldsymbol{\Sigma}\right\}\right] \\ &= \mathbb{E}\left[\mathrm{tr}\left\{w_2(d_0)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)^\top\boldsymbol{\Sigma}^{-1} - \mathbf{I}_k\right\}\right] = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)\|\mathbf{z}\|^2 - k\right],\end{aligned}$$

where $d_0^2 = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)$, so that $k = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}[w_2(\|\mathbf{z}\|)\|\mathbf{z}\|^2] = \mathbb{E}[u_2(T)T]$. It then follows that

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right] = \mathbb{E}\left[u_2(T)^2T^2\right] = k(k+2)\psi_1, \quad \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[w_2'(\|\mathbf{z}\|)\|\mathbf{z}\|^3\right] = 2\mathbb{E}\left[u_2'(T)T^2\right] = 2k(\psi_2 - 1),$$

where ψ_1 and ψ_2 are defined in Example 3 in Tyler [27]. Then from the expression for σ_1 provided in Theorem 2 we find

$$\sigma_1 = \frac{k^2(k+2)^2\psi_1}{\left(2k(\psi_2-1) + k(k+2)\right)^2} = \frac{(k+2)^2\psi_1}{(2\psi_2 + k)^2},$$

which coincides with the one in Example 3 in Tyler [27]. For σ_2 in Corollary 2 we obtain

$$\sigma_2 = -\frac{2\sigma_1}{k} + \frac{4\left\{k(k+2)\psi_1 - k^2\right\}}{(2k\psi_2)^2}.$$

After inserting the expression for σ_1 in σ_2, one can verify that also the expression for σ_2 coincides with one in Example 3 in Tyler [27]. $\square$

Proof of Example 3. With $w_1(s) = \rho'(s)/s$, $w_2(s) = k\rho'(s)/s$ and $w_3(s) = \rho'(s)s - \rho(s) + b_0$, one can easily verify that the expressions for σ_1 and σ_2 in Corollary 2 coincide with the ones in Corollary 9.2 in Lopuhaä et al. [17]. $\square$

Appendix C. Details for Examples 4 and 5

Proof of Example 4. From (22) we find

$$\begin{aligned} H'(\boldsymbol{\Sigma})\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top H'(\boldsymbol{\Sigma})^\top &= \frac{1}{k^2}|\boldsymbol{\Sigma}|^{-2/k}\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^\top\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{\Sigma})^\top \\ &- \frac{1}{k}|\boldsymbol{\Sigma}|^{-2/k}\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma}^{-1})^\top\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top - \frac{1}{k}|\boldsymbol{\Sigma}|^{-2/k}\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top\mathrm{vec}(\boldsymbol{\Sigma}^{-1})\mathrm{vec}(\boldsymbol{\Sigma})^\top \\ &+ |\boldsymbol{\Sigma}|^{-2/k}\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top. \end{aligned} \tag{C.1}$$

Using that $\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_0$ and

$$\mathrm{vec}(\boldsymbol{\Sigma}^{-1}) = \mathrm{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}) = \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\boldsymbol{\Sigma}) = \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\boldsymbol{\theta}_0,$$

the first term on the right hand side of (C.1) reduces to $(1/k)|\boldsymbol{\Sigma}|^{-2/k}\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top$. Similarly, the second and third term on the right hand side of (C.1) are equal to $-(1/k)|\boldsymbol{\Sigma}|^{-2/k}\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top$. Putting everything together, we find that the limiting variance of $\sqrt{n}(H(\mathbf{V}_n) - H(\boldsymbol{\Sigma}))$ is given by (23). □

Proof of Example 5. From Example 4 and the delta method, it follows that the limiting variance of $\sqrt{n}(H(\boldsymbol{\theta}_n) - H(\boldsymbol{\theta}))$ is given by

$$\begin{aligned} &(\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\left[2\sigma_1|\boldsymbol{\Sigma}|^{-2/k}\left\{\mathbf{L}\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^\top - \frac{1}{k}\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top\right\}\right]\mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1} \\ &\quad= 2\sigma_1|\boldsymbol{\Sigma}|^{-2/k}\left\{\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1} - \frac{1}{k}(\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top\mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1}\right\} \\ &\quad= \frac{2\sigma_1}{|\boldsymbol{\Sigma}|^{2/k}}\left\{\Big(\mathbf{L}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1} - \frac{1}{k}\boldsymbol{\theta}_0\boldsymbol{\theta}_0^\top\right\}, \end{aligned}$$

using that $\mathrm{vec}(\boldsymbol{\Sigma}) = \mathbf{L}\boldsymbol{\theta}_0$. □

References

[1] I. Chervoneva, M. Vishnyakov, Constrained *S*-estimators for linear mixed effects models with covariance components, Stat. Med. 30 (14) (2011) 1735–1750.
[2] I. Chervoneva, M. Vishnyakov, Generalized S-estimators for linear mixed effects models, Statist. Sinica 24 (3) (2014) 1257–1276.
[3] S. Copt, S. Heritier, Robust alternatives to the F-test in mixed linear models based on MM-estimates, Biometrics 63 (4) (2007) 1045–1052.
[4] S. Copt, M.P. Victoria-Feser, High-breakdown inference for mixed linear models, J. Amer. Statist. Assoc. 101 (473) (2006) 292–300.
[5] C. Croux, G. Haesbroeck, Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies, Biometrika 87 (3) (2000) 603–618.
[6] G.M. Fitzmaurice, N.M. Laird, J.H. Ware, Applied Longitudinal Analysis, second ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, 2011, p. xxviii+701.
[7] F.R. Hampel, The influence curve and its role in robust estimation, J. Amer. Statist. Assoc. 69 (1974) 383–393.
[8] H.O. Hartley, J.N.K. Rao, Maximum-likelihood estimation for the mixed analysis of variance model, Biometrika 54 (1967) 93–108.
[9] S. Heritier, E. Ronchetti, Robust bounded-influence tests in general parametric models, J. Amer. Statist. Assoc. 89 (427) (1994) 897–904.
[10] P.J. Huber, Robust Statistics, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1981, p. ix+308.
[11] R.I. Jennrich, M.D. Schluchter, Unbalanced repeated-measures models with structured covariance matrices, Biometrics 42 (4) (1986) 805–820.
[12] J.T. Kent, D.E. Tyler, Constrained *M*-estimation for multivariate location and scatter, Ann. Statist. 24 (3) (1996) 1346–1370.
[13] N.L. Kudraszow, R.A. Maronna, Estimates of MM type for the multivariate linear model, J. Multivariate Anal. 102 (9) (2011) 1280–1292.
[14] H.P. Lopuhaä, On the relation between *S*-estimators and *M*-estimators of multivariate location and covariance, Ann. Statist. 17 (4) (1989) 1662–1683.
[15] H.P. Lopuhaä, Asymptotic expansion of *S*-estimators of location and covariance, Stat. Neerl. 51 (2) (1997) 220–237.
[16] H.P. Lopuhaä, Highly efficient estimators with high breakdown point for linear models with structured covariance matrices, Econ. Stat. (2023).
[17] H.P. Lopuhaä, V. Gares, A. Ruiz-Gazen, S-estimation in linear models with structured covariance matrices, Ann. Statist. 51 (6) (2023) 2415–2439.
[18] H.P. Lopuhaä, V. Gares, A. Ruiz-Gazen, Supplement to "S-estimation in linear models with structured covariance matrices", Ann. Statist. 51 (6) (2023).
[19] J.R. Magnus, 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, p. xviii+393.
[20] C.L. Mallows, Latent vectors of random symmetric matrices, Biometrika 48 (1961) 133–149.
[21] K.V. Mardia, R.J. Marshall, Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika 71 (1) (1984) 135–146.
[22] R.A. Maronna, Robust *M*-estimators of multivariate location and scatter, Ann. Statist. 4 (1) (1976) 51–67.
[23] J.J. Miller, Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance, Ann. Statist. 5 (4) (1977) 746–762.
[24] P. Rousseeuw, V. Yohai, Robust regression by means of S-estimators, in: Robust and Nonlinear Time Series Analysis (Heidelberg, 1983), in: Lect. Notes Stat., vol. 26, Springer, 1984, pp. 256–272.
[25] M. Salibián-Barrera, S. Van Aelst, G. Willems, Principal components analysis based on multivariate MM estimators with fast and robust bootstrap, J. Amer. Statist. Assoc. 101 (475) (2006) 1198–1211.
[26] K.S. Tatsuoka, D.E. Tyler, On the uniqueness of *S*-functionals and *M*-functionals under nonelliptical distributions, Ann. Statist. 28 (4) (2000) 1219–1243.
[27] D.E. Tyler, Radial estimates and the test for sphericity, Biometrika 69 (2) (1982) 429–436.
[28] D.E. Tyler, Robustness and efficiency properties of scatter matrices, Biometrika 70 (2) (1983) 411–420.
[29] S. Van Aelst, G. Willems, Multivariate regression *S*-estimators for robust estimation and inference, Statist. Sinica 15 (4) (2005) 981–1001.
[30] S. Van Aelst, G. Willems, Robust and efficient one-way MANOVA tests, J. Amer. Statist. Assoc. 106 (494) (2011) 706–718.