Robust Nonparametric Inference

Klaus Nordhausen[1] and Hannu Oja[2]

[1]Department of Mathematics and Statistics, University of Turku, Turku, Finland, 20014
[2]Department of Mathematics and Statistics, University of Turku, Turku, Finland, 20014; email: hannu.oja@utu.fi

Xxxx. Xxx. Xxx. Xxx. YYYY. AA:1–32

This article's doi:
10.1146/((please add article doi))

Keywords

affine equivariance and invariance, location, scatter, signs and ranks, skewness and kurtosis, subspace estimation

Abstract

In the paper we provide a personal review of the literature on nonparametric and robust tools in the standard univariate and multivariate location-scatter as well as linear regression problems with a special focus on sign and rank methods, their equivariance and invariance properties and their robustness and efficiency. Beyond parametric models the population quantities of interest are then often formulated as location, scatter, skewness, kurtosis and other functionals. Some old and recent tools for model checking, dimension reduction and subspace estimation in wide semiparametric models are discussed. We also discuss recent extensions of procedures in certain non-standard semiparametric cases including clustered data, time series, matrix valued observations and functional data. Our personal list of important unsolved and future issues is provided.

1. INTRODUCTION

Classical multivariate statistical inference methods are based on the explicit or implicit assumption that the observed data come from a multivariate normal (Gaussian) distribution. See, for example, the monographs by Anderson (2003) and Mardia et al. (1979). The use of the mean vector and covariance matrix for location and scatter problems are often optimal under the assumption of multivariate normality but poor in efficiency for heavy-tailed distributions and highly sensitive to outlying observations. The robust approach to statistical data analysis aims at finding reliable inference tools not only for the true parametric model but also for other distributions close enough or similar enough to a target model, see e.g. the books by Huber (1980), Hampel et al. (1986), and Maronna et al. (2006). The nonparametric methods, often based on signs and ranks, tend to be robust in the global sense and valid in wide semiparametric and nonparametric models. Univariate nonparametric methods are treated in Hajek & Sidak (1967), Lehmann (1975), and Hettmansperger & McKean (2011), for example. The classical book by Puri & Sen (1971) gives a complete presentation of multivariate methods based on marginal signs and ranks. These methods are not affine invariant or equivariant, however, and other equivariant notions of sign and rank have been proposed in the literature. See Oja (1999) for a full description of affine equivariant signs and ranks and their use in multivariate location problems. Oja (2010) described analysis tools for multivariate data that are based on rotation equivariant spatial signs and ranks. For robust and nonparametric statistics, see also Jurečková et al. (2013).

Robust and nonparametric analysis faces new challenges raised by highly complex modern data sets. The number of measured variables may be huge as compared to the number of observations, the measurements on each individual are curves or figures rather than numbers, the variables are non-Gaussian, the dependence between the variables is not linear, the observations may be cluster-dependent, time-dependent, spatially dependent, or dependent in other ways, and so on. Nonparametric and robust tests and estimates in location, scatter and linear regression problems are quite well developed for independent and identically distributed observations and are valid in wide semiparametric and nonparametric models, far beyond the standard Gaussian model. The extensions of these tools to non-standard data analyses are not straightforward and only sparse in the literature. Different goodness-of-fit tests and tests for Gaussianity often help in the choice of the model and appropriate inference tools. Dimension reduction tools such as principal component analysis, independent component analysis, invariant coordinate selection and sliced inverse regression can also be formulated using scatter functionals in relevant semiparametric models and robustifying these tools is necessary as well.

The paper is structured as follows. We first provide a highly personal and far from complete review of the literature on nonparametric and robust tools in the standard univariate (Chapter 2) and multivariate (Chapter 3) location and scatter as well as in linear regression cases, with a special focus on the sign and rank methods, their equivariance and invariance properties, and their robustness and efficiency properties. Beyond parametric models, the procedures are then formulated as statistical, e.g. location and scatter, functionals. In Chapter 4, some tools for dimension reduction and subspace estimations in wide semiparametric models are discussed. Finally in Chapter 5, we provide some notions on recent extensions of procedures in certain non-standard semiparametric cases including time series, clustered data, matrix and tensor valued observations, and functional data. Our personal lists of important unsolved and future issues are provided. Throughout the paper, the variables of interest are assumed to be continuous.

2. ROBUST NONPARAMETRIC METHODS - UNIVARIATE CASE

2.1. Location-Scale Models

We start with a simple univariate model and assume that $y_1, ..., y_n$ is a random sample from a distribution which can be expressed as

$$y = az + b,$$

where z is a continuous standardized random variable such that $E(z) = 0$ and $Var(z) = 1$ and $a, b \in \Re$ are unknown constants. The aim is to draw inference on location b and scale $|a|$. How to do that depends strongly on what one can assume about the distribution of z. One can for example assume that z has (i) a normal (Gaussian) distribution, (ii) a symmetric distribution ($-z \sim z$) or (iii) has any distribution with a finite second moment. Model (i) is a *parametric model* while models (ii) and (iii) are *nonparametric models* of all symmetric distributions or of all distributions, respectively, with finite second moments. Models (ii) and (iii) may be seen also as *semiparametric models* with two parameters, standard deviation $|a|$ and mean b, due to the specific standardization of z. To avoid the assumption that the second moment exists, the moment assumptions may be replaced for example by the assumptions that the median and inter-quartile range of z are 0 and 1, respectively. Inference on the distribution of z – often in terms of skewness and kurtosis – is then helpful in the choice of estimation and testing methods.

If one does not trust in Gaussianity, it is of course possible to adopt a parametric model with the normal distribution as a special or limiting case, such as the t-distribution family or the skew-normal family of distributions. Another possibility is to consider models allowing small deviations from the normal distributions, ϵ-neighborhood models of the $N(0, 1)$ density φ such as a mixture model $\{f : f \geq (1 - \epsilon)\varphi\}$.

2.2. Descriptive Functionals and Statistics

In wide semiparametric and nonparametric models $\mathcal{F}$, closed under linear transformations ($F_y \in \mathcal{F} \Rightarrow F_{ay+b} \in \mathcal{F}$) and large enough to include empirical distributions, the properties of distributions may be be operationalized by descriptive functionals and corresponding sample statistics. The characteristics of univariate distributions most commonly considered are location, scale, skewness and kurtosis. Location is often most interesting and scale is mainly needed for the accuracy of its estimate. Skewness and kurtosis have often, but not always, a secondary role as tools for model checking. Kurtosis may sometimes be seen as an accuracy measure for a scale estimate.

The comparison of two distributions could be made as follows. Let $x \sim F$ and $y \sim G$ and let F and G be strictly increasing. The transport function $R(\cdot) := G^{-1}(F(\cdot))$ then has the property that $R(x) \sim y$ and it minimizes $E[x - R(x)]^2$ or maximizes $Cor(x, R(x))$ over all functions R for which $R(x) \sim G$. See Chapter 2 in Santambrogio (2015). It provides partial orderings for location ($R(x) \geq x$ is positive), scale ($R(x)$ is increasing), skewness ($R(x)$ is convex) and kurtosis ($R(x)$ is concave-convex for symmetrical distributions). See Oja (1981) and references therein for original contributors to these ideas. The measures for these properties are then hoped to preserve the corresponding partial orderings and they should be location-scale or *affine equivariant or invariant* in a natural way.

Using affine equivariant location and scale functionals $\mu, \mu_1, \mu_2, ...$ and $\sigma, \sigma_1, \sigma_2, ...$, skewness and kurtosis functionals can be found as ratios $(\mu_1 - \mu_2)^2/\sigma^2$ and σ_2^2/σ_1^2, respectively.

Univariate Location and Scale Functionals

Univariate location and scale functionals are affine equivariant in the following way.

(i) For a location functional $\mu : \mathcal{F} \to \Re$, $\mu(F_{ay+b}) = a\mu(F_y) + b$ for all $a, b \in \Re$.
(ii) For a scale functional $\sigma : \mathcal{F} \to \Re_+$, $\sigma(F_{ay+b}) = |a|t(F_y)$ for all $a, b \in \Re$.

These functionals are affine invariant and can be used to check the assumptions on the distribution of z.

The mean μ and standard deviation σ serve as first examples of location and scatter functionals. Other moment based location and scale functionals are for example $E(z^2 y)$ (location) and $E[z^2(y-\mu)^2]$ (squared scale) where $z = (y-\mu)/\sigma$, and these four measures together provide the classical (moment based) measures of skewness and kurtosis, $E[z^3]$ and $E[z^4]$. In his celebrated system of frequency curves, Pearson (1895) identified different parametric families of distributions and the classification was simply based on $E[z^3]$ and $E[z^4]$. For $0 < \alpha < \beta < 0.5$, location, scale, skewness and scatter functionals based on quantiles are given e.g. by $F^{-1}(0.5)$, $F^{-1}(1-\alpha) - F^{-1}(\alpha)$, $(F^{-1}(1-\alpha) + F^{-1}(\alpha) - 2F^{-1}(0.5))/(F^{-1}(1-\alpha) - F^{-1}(\alpha))$ and $(F^{-1}(1-\beta) - F^{-1}(\beta))/(F^{-1}(1-\alpha) - F^{-1}(\alpha))$.

Consider next the tools for considering the robustness of a functional $\mu : \mathcal{F} \to \Re$. First, it is hoped that the functional is *weakly continuous* in the sense that $F_n \to_d F$ implies that $\mu(F_n) \to \mu(F)$ guaranteeing the consistency of the sample statistic. *Local robustness* properties are usually considered using the *influence function (IF)*. The influence function of μ at F is defined as

$$IF(x; \mu, F) = \lim_{\epsilon \to 0} \frac{\mu((1-\epsilon)F + \epsilon\delta_x) - \mu(F)}{\epsilon}$$

where δ_x is a distribution with the probability mass one at x. It is thus the Gâteaux derivative of μ at F in the direction of δ_x. The IF is bounded for robust functionals. For *global robustness* one considers the ϵ-neighborhoods of F, say $\mathcal{F}_\epsilon$, given by $\{g : g \geq (1-\epsilon)f\}$ ('ϵ-mixture model') or $\left\{g : \int (f-g)_+ \geq \epsilon\right\}$ ('ϵ-replacement model'). We then say for example that location functional μ breaks down in $\mathcal{F}_\epsilon$ if $\mu(\mathcal{F}_\epsilon) = \Re$ as in this (close) neighborhood of F it can be forced to have any value. The smallest ϵ for this to happen is called the *breakdown point* of μ at F. The breakdown points of the mean and variance are zero and those of $F^{-1}\left(\frac{1}{2}\right)$ and $F^{-1}(1-\alpha) - F^{-1}(\alpha)$ are $\frac{1}{2}$ and α, respectively. For robustness of the sample statistics, sensitivity curves and finite sample breakdown points are defined and used in a similar way.

2.3. Inference on Location and Scale

2.3.1. M-functionals for location and scale. For known distribution F (or f) of z, symmetric around zero, the *maximum likelihood (ML) functionals* for location and scale, say μ an σ, maximize $L(\mu, \sigma) = E\left[\log\left(\frac{1}{\sigma} f\left(\frac{y-\mu}{\sigma}\right)\right)\right]$ and are then determined by implicit equations

$$E[w(|z|)z] = 0 \quad \text{and} \quad E[w(|z|)z^2] = 1$$

where $z = \frac{y-\mu}{\sigma}$ and $w(z) = \frac{l(z)}{z}$ and $l = -f'/f$ is the optimal location score function. In the $N(0,1)$-case, μ and σ are the mean and variance (functionals) and in the double exponential (Laplace) case μ is the median and $\sigma = E[|y-\mu|]$ is the mean deviation around the median. The so called *M-functionals* μ and σ then satisfy, for any choices of positive weight functions w_1, w_2 and w_3 (not necessarily related to any distribution),

$$E[w_1(|z|)z] = 0 \quad \text{and} \quad E[w_2(|z|)z^2] = E[w_3(|z|)]. \tag{1}$$

Note that the functionals μ and σ are location-scale equivariant and the above estimating equations further suggest a fixed-point algorithm of the form

$$\mu \leftarrow \frac{E[w_1(|z|)y]}{E[w_1(|z|)]} \quad \text{and} \quad \sigma^2 \leftarrow \frac{E[w_2(|z|)(y-\mu)^2]}{E[w_3(|z|)]}, \tag{2}$$

as well as one-step functionals starting with z that is standardized by preliminary affine equivariant functionals μ an σ. The influence functions of M-functionals are easily derived from their weight functions. The highly robust Huber's estimate for example uses weight functions $w_1(r) = \min(k/r, 1)$, $w_2(r) = \min(k^2/r^2, 1)$ and $w_3(r) \equiv 1$, and w_1 provides in the pure location case, with some choice $k = k(\epsilon)$, a minimax solution in an ϵ-mixture neighborhood of a Gaussian model. For M-estimation, see e.g. Huber (1980) and Hampel et al. (1986).

2.3.2. R-functionals for location. Rank based location functionals, R-functionals, use sign, rank and signed-rank scores in the estimation. For a distribution F, the *sign, centered rank and signed-rank* score functions, s, r and q are given by

$$s(z) := sign(z), \quad r(z) := 2F(z) - 1 \quad \text{and} \quad q(z) := s(z)\left[F(|z|) - F(-|z|)\right].$$

These are the scores corresponding to the three criterion functions

$$E(|z|) = E[s(z)z], \quad E(|z_1 - z_2|) = E[r(z)z] \quad \text{and} \quad E(|z_1 - z_2|) + E(|z_1 + z_2|) = 4E[q(z)z],$$

for z and its independent copies z_1 and z_2, see Hettmansperger & Aubuchon (1988). Let y, y_1, y_2 be independent observations (copies) from F. *Median* $\mu : \mathcal{F} \to \Re$ is then a functional that minimizes $E[|y-\mu|]$ or solves $E[s(y-\mu)] = 0$ while the *Hodges-Lehmann (HL) functional* minimizes $E[|y_1 + y_2 - 2\mu|]$ or solves $E[q(y-\mu)] = 0$. To attain higher efficiency, one sometimes uses *rank scores* $h(r(z))$ or $s(z)h(|q(z)|)$ instead of $r(z)$ and $q(z)$ with some additional score function h. See e.g. Hettmansperger & McKean (2011).

For the empirical distribution F_n from a random sample $y_1, ..., y_n$, we get the empirical centered rank and signed-rank functions $r_n(z) = 2F_n(z) - 1$ and $q_n(z) := s(z)\left[F_n(|z|) - F_n(-|z|)\right]$. The sample median and HL-estimate $\hat{\mu}$ then center the observations so that, for $z_i = y_i - \hat{\mu}$, $i = 1, ..., n$, $\sum_{i=1}^n s(z_i) = 0$ and $\sum_{i=1}^n q_n(z_i) = 0$, respectively. Note that the HL estimate is the median of all pairwise averages $\frac{1}{2}(y_i + y_j)$, $1 \le i \le j \le n$.

2.3.3. Other functionals. Let $y_{1:n}, y_{2:n}, ..., y_{n:n}$ be an ordered sample of size n. The so called L-estimate, linear combination of order statistics, is then $\sum_{i=1}^n c_{i,n} y_{i:n}$ where often $c_{i:n} = h\left(\frac{i}{n+1}\right)$. For different choices for function h, a whole class of location and scale estimates are obtained for the corresponding population quantities, *L-functionals* $E[h(F(y))y]$. Descriptive functionals may be created also by starting from the so called *elemental estimates*: Assume that k is the smallest sample size to find a (mean or median) unbiased

estimate $h(y_1, ..., y_k)$ of a population quantity. In the mean unbiased case, the sample statistic is the *U-statistic*, the average of all elemental estimates $h(y_{i_1}, ..., y_{i_k})$, $i_1 < ... < i_k$, with a well developed asymptotic theory. See e.g. Serfling (1980). In the spirit of partial orderings for location, scale, skewness and kurtosis discussed in Section 2.1, one can define the corresponding measures as the expected values of $y_{1:1}$, $y_{2:2} - y_{1:2}$, $(y_{3:3} - y_{2:3})/(y_{3:3} - y_{1:3})$ and $(y_{4:4} - y_{1:4})/(y_{3:4} - y_{2:4})$ or use *L-moments* $\lambda_1 = E(y_{1:1})$, $\lambda_2 = \frac{1}{2}E(y_{2:2} - y_{1:2})$,

$$\lambda_3 = \frac{1}{3}E(y_{3:3} - 2y_{2:3} + y_{1:3}) \quad \text{and} \quad \lambda_4 = \frac{1}{4}E(y_{4:4} - 3y_{3:4} + 3y_{2:4} - y_{1:4}),$$

or their robustified trimmed versions. See Hosking (1990). The sample L-moments are then L-estimates as well as U-statistics. Note that λ_3/λ_2 and λ_4/λ_2 are skewness and kurtosis functionals. For the HL functional, $k = 2$ and the functional is the median of $\frac{1}{2}(y_1 + y_2)$. To obtain highly robust estimates, the subsamples can be used also in other ways: Find the shortest interval covering at least half of the observations. The so called *Shorth estimates of location and scale* are then the midpoint and length of that interval.

ASYMPTOTIC RELATIVE EFFICIENCY (ARE): In the normal case, ARE of the median with respect to the mean is 0.64. Roughly speaking, this means that the median with $n = 100$ has the same variance as the mean with $n = 64$.

2.3.4. Location estimates and tests based on a score function. We next speak about the asymptotic efficiency of the location estimates and companion tests that are based on a monotone and odd score function $k(y)$. The score function may come from M-estimation (with known or estimated σ) or R-estimation, possibly with a rank score function. Let $y_1, ..., y_n$ be a random sample from a distribution f symmetric around μ. The estimate $\hat{\mu}$ centers the observed scores, that is, it solves the estimation equation $\sum_{i=1}^{n} k(y_i - \hat{\mu}) = 0$. Under general assumptions and with the true location center μ_0, the limiting distribution of $\sqrt{n}(\hat{\mu} - \mu_0)$ is $N(0, \tau)$ where $\tau = E[k^2(z)]/E^2[k(z)l(z))]$ with $z = y - \mu_0$. The asymptotic efficiency of the estimate with respect to the optimal ML-estimate is then simply $Cor(k(z), l(z))$. With unknown l, the estimation of τ may be tricky but the bootstrap estimation of the variance of the estimate is easy to implement. For the estimation of the limiting variance of the median and HL estimate, see e.g. Oja (1999) for a short review. For the median and HL estimate, the efficiencies with respect to the mean are in the normal case 0.64 and 0.95. In the t_3 and t_6 cases the efficiencies are 1.62 and 0.88 for the median and 1.90 and 1.16 for the HL-estimate. Using the so called van der Waerden score gives an estimate that is never outperformed by the mean.

The companion M-test statistic for $H_0 : \mu = \mu_0$ is $T_n = n^{-1/2} \sum_{i=1}^{n} k(y_i - \mu_0)$ with the limiting null distribution $N(0, E[k^2(z)])$. Under a contiguous sequence of alternatives $H_{1,n} : \mu = \mu_0 + \frac{1}{\sqrt{n}}\delta$, the limiting distribution of $\sqrt{n}T_n$ is $N(\delta E[k(z)l(z)], E[k^2(z)])$ (LeCam's third lemma) and then $T_n^2/(\frac{1}{n}\sum_i k^2(y_i - \mu_0))$ converges to a non-central chi-squared distribution with non-centrality parameter $\delta^2 E^2[k(z)l(z)]/E[k^2(z)]$. The Pitman efficiencies w.r.t. the t-test then equal to the efficiencies of the corresponding estimate w.r.t. the mean. Note also that the sign test and signed-rank test are in fact distribution-free, that is, their (finite-sample as well as asymptotic) distribution does not depend on the symmetric f. Thus, they also provide distribution-free confidence intervals for μ. Conditionally distribution-free sign-change test versions are in fact available for any choice of an odd score function k.

2.3.5. Tests for normality. Skewness and kurtosis statistics are often used to test for normality. Let $z_i = (y_i - \bar{y})/s_y$, be the observations standardized with the sample mean $\bar{y}$ and sample standard deviation s_y. The third and fourth cumulants (for skewness and kurtosis)

are then $\frac{1}{n}\sum_{i=1}^n z_i^3$ and $\frac{1}{n}\sum_{i=1}^n z_i^4 - 3$ and, for the observations from $N(\mu, \sigma^2)$,

$$\frac{n}{6}\left[\frac{1}{n}\sum_{i=1}^n z_i^3\right]^2 + \frac{n}{24}\left[\frac{1}{n}\sum_{i=1}^n z_i^4 - 3\right]^2 \to_d \chi_2^2 \quad \text{as } n \to \infty.$$

Other skewness and kurtosis statistics can be used in model checking as well. Interesting choices for the test statistics are also $\sum_{i=1}^n z_{i:n}\Phi^{-1}\left(\frac{i}{n+1}\right)$ and $\sum_{i=1}^n s(y_i - \bar{y})$.

2.4. Linear Regression Analysis

2.4.1. Linear regression model. The conditional distribution of y_i given $\mathbf{x}_i \in \Re^q$. $i = 1, ..., n$, is now that of

$$y_i = az_i + b_0 + \mathbf{b}'\mathbf{x}_i,$$

where $a, b_0 \in \Re, \mathbf{b} \in \Re^q$, $z_1, ..., z_n$ is a random sample from the distribution of z with $E(z) = 0$ and $Var(z) = 1$. Again, one may assume that z has a normal distribution, a symmetric distribution or a distribution in other families listed in the location-scale case. The data set is given by the matrix $(\mathbf{X}, \mathbf{y}) \in \Re^{n\times(q+1)}$. Note that the models for two and several samples as well as for randomized blocks can also be written in this way. The vectors $\mathbf{x}_i$ may be design variables (preassigned in the experiment) or observational variables. In the latter case, $\mathbf{x}_i$ and z_i should be independent and the distribution of $\mathbf{x}_i$ should not depend on the parameters of interest, $|a|$, b_0 and $\mathbf{b}$. The aim is to draw inference on these parameters and the distribution of z has an crucial role in the choice of the method.

2.4.2. M-estimates of the regression coefficients. For known f symmetric around zero and fixed $\mathbf{x}_1, ..., \mathbf{x}_n$, the *maximum likelihood (ML) estimates* for regression coefficients and scale, say $\hat{\beta}_0 \in \Re$, $\hat{\boldsymbol{\beta}} \in \Re^q$ and $\hat{\sigma} \in \Re_+$ are determined by implicit equations

$$\frac{1}{n}\sum_{i=1}^n w(|z_i|)z_i\tilde{\mathbf{x}}_i = \mathbf{0} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n w(|z_i|)z_i^2 = 1$$

where $\tilde{\mathbf{x}}_i = (1, \mathbf{x}_i')'$, $z_i = \frac{y_i - \beta_0 - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}$ and $w(z) = \frac{l(z)}{z}$. In the normal error case $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$ are the so called least squares (LS) estimates and minimize $\sum_{i=1}^n (y_i - \beta_0 - \mathbf{x}_i'\boldsymbol{\beta})^2$. The extended family of estimates, *M-estimates for regression* solve

$$\sum_{i=1}^n w_1(|z_i|)z_i\tilde{\mathbf{x}}_i = \mathbf{0} \quad \text{and} \quad \sum_{i=1}^n w_2(|z_i|)z_i^2 = \sum_{i=1}^n w_3(|z_i|)$$

with some choices of positive weight functions w_1, w_2, w_3.

2.4.3. R-estimates of the regression coefficients. The estimates based on signs and ranks minimize the mean deviation and the mean difference of the residuals, that is,

$$\sum_{i=1}^n |y_i - \beta_0 - \mathbf{x}_i'\boldsymbol{\beta}| \quad \text{and} \quad \sum_{i<j} |(y_j - y_i) - (\mathbf{x}_j - \mathbf{x}_i)'\boldsymbol{\beta}|,$$

respectively. They then also solve the corresponding estimation equations

$$\sum_{i=1}^n s(z_i)\tilde{\mathbf{x}}_i = \mathbf{0} \quad \text{and} \quad \sum_{i=1}^n r_n(z_i)\mathbf{x}_i = \mathbf{0}.$$

(The rank function r_n is for the estimated residuals $z_1, ..., z_n$.) Note that the rank estimate is only for $\mathbf{b}$ and, if needed, b_0 must be estimated separately by using the residuals $y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$, $i = 1, ..., n$. For the sign method, the signs of the residuals and the components of $\mathbf{x}$ are uncorrelated while, for the rank method, the centered ranks of the residuals and the components are uncorrelated. Note also that simultaneous estimation of $|a|$ is not needed. In the rank case, assumption on the symmetry of the residuals is not needed. Again, using rank scores instead of pure ranks in the second estimation equation, may yield gains in efficiency of the corresponding estimate (and test).

2.4.4. Other estimates of the regression coefficients. We discuss the use of elemental estimates in the regression and first consider the simple regression model $y = az + b_0 + bx$, The *elemental estimates* based on the ith and jth observations, $1 \leq i < j \leq n$, are then

$$\hat{b}_{0;ij} = \frac{y_i x_j - y_j x_i}{x_j - x_i} \quad \text{and} \quad \hat{b}_{ij} = \frac{y_j - y_i}{x_j - x_i}.$$

The LS estimate is then weighted average of all $\binom{n}{2}$ elemental estimates with weights $(x_j - x_i)^2$. In the simple linear regression, the classical slope and intercept estimates of Theil and Maritz are just medians of the $\hat{b}_{ij}$'s and the $\hat{b}_{0;ij}$'s. The rank regression estimate of the slope is the weighted median of the $\hat{b}_{ij}$'s, the weights being proportional to $|x_j - x_i|$. See Hettmansperger & McKean (2011). For the elemental estimates in the general multivariate multiple linear regression case, see Chapter 3.

2.4.5. Properties of the regression estimates based on a location score function. Consider next the limiting distribution of $\hat{\boldsymbol{\beta}}$ for a general score function. As in the location case, the score function $k(z)$ is from the M-estimation (with known or estimated $|a|$) or in R-estimation with sign or rank scores. Write $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)' \in \Re^{n \times q}$ and assume that $\frac{1}{n}\mathbf{X}'\mathbf{X} \to \boldsymbol{\Sigma}_x \in \Re^{q \times q}$ as $n \to \infty$. Then, under general assumptions, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \mathbf{b}) \to N_q(\mathbf{0}, \tau\boldsymbol{\Sigma}_x^{-1})$ with the same τ as in the location case. This means that the relative efficiencies are inherited from the simple location case. Note that the asymptotic tests for regression coefficients can be based on this result. In the several sample location problem, also conditionally distribution-free permutation tests are available. For a thorough discussion on robustness in linear regression, see Chapters 4 and 5 in Maronna et al. (2006).

2.5. R Packages for Univariate Methods

We list some `R` packages for nonparametric tools in the univariate case. Our short list is subjective and not comprehensive. The basic installation of `R` (R Core Team 2016) already contains many nonparametric estimates and tests. Further `R` packages are `exactRankTests` (Hothorn & Hornik 2017) and `gMWT` (Fischer & Oja 2015). Linear regression based on signs and ranks can be done for example using the packages `quantreg` (Koenker 2016) and `Rfit` (Kloke & McKean 2012). Many robust methods are implemented in the packages `MASS` (Venables & Ripley 2002) and `robustbase` (Mächler et al. 2016).

3. ROBUST NONPARAMETRIC METHODS - MULTIVARIATE CASE

3.1. Location-Scatter Model

In the multivariate location-scatter model we assume that $\mathbf{y}_1, ..., \mathbf{y}_n$ is a random sample from a p-variate distribution of

$$\mathbf{y} = \mathbf{A}\mathbf{z} + \mathbf{b}$$

where $\mathbf{A} \in \Re^{p\times p}, \mathbf{b} \in \Re^p$ and $\mathbf{z}$ is a p-variate random vectors standardized so that $E(\mathbf{z}) = \mathbf{0}$ and $Cov(\mathbf{z}) = \mathbf{I}_p$. We may further assume that $\mathbf{z} \in \Re^p$ has (i) a multivariate normal distribution, (ii) a spherical distribution or (iii) has any distribution with finite second moments. One can also avoid the silent assumption on the existence of second moments with another standardization. Note that if $\mathbf{z}$ is *spherical or spherically symmetric* then $\mathbf{U}\mathbf{z} \sim \mathbf{z}$ for all orthogonal matrices $\mathbf{U} \in \Re^{p\times p}$. Weaker extensions of symmetry are *marginal symmetry* and *(regular) symmetry*, that is, $(\pm z_1, ..., \pm z_p)' \sim (z_1, ..., z_p)'$ and $-\mathbf{z} \sim \mathbf{z}$, respectively. Model (i) is parametric, models (ii) and (iii) are examples of semiparametric *elliptical* and fully nonparametric models. In the semiparametric *independent component (IC) model* one assumes that the components $z_1, ..., z_p$ of $\mathbf{z}$ are mutually independent. In all cases the aim is to make inference on location $\mathbf{b}$ and scatter $\mathbf{A}\mathbf{A}'$. In the IC model, one also wishes to estimate the mixing matrix $\mathbf{A}$ or its inverse.

As in the univariate case, mixture neighborhood models of $N_p(\mathbf{0}, \mathbf{I}_p)$ with the density functions φ can be given as $\{f : f \geq (1-\epsilon)\varphi\}$. Parametric multivariate models with multivariate normality as a special or limiting case are, for example, multivariate skew-normal or multivariate t-distribution families. Copulas provide semiparametric models with unspecified marginal distributions and parameters to control the dependence.

3.2. Descriptive Functionals and Statistics

In the univariate case, one could define partial orderings for location, scale, skewness and kurtosis by using the transport function $R = G^{-1}F$. In the multivariate case, the *Monge-Brenier transport function* $R : \Re^p \to \Re^p$ similarly minimizes $E\|\mathbf{x} - R(\mathbf{x})\|^2$ under the conditions that $\mathbf{x} \sim F \Rightarrow R(\mathbf{x}) \sim G$. It is still an open question whether R could be used in the comparison of the distributions F and G. Recently, Chernozhukov et al. (2017) used this transport function for data depth construction. See also Santambrogio (2015).

Let $\mathcal{F}$ be a family of p-variate distributions that is closed under affine transformations and large enough to include all finite sample distributions. Location and scatter functionals $\boldsymbol{\mu} : \mathcal{F} \to \Re^p$ and $\boldsymbol{\Sigma} : \mathcal{F} \to \Re_+^{p\times p}$ are functionals that satisfy natural affine equivariant (AE) conditions.

Multivariate Location and Scatter Functionals

Multivariate location and scatter functionals satisfy the AE conditions

$$\boldsymbol{\mu}(F_{\mathbf{A}\mathbf{y}+\mathbf{b}}) = \mathbf{A}\boldsymbol{\mu}(F_{\mathbf{y}}) + \mathbf{b} \quad \text{and} \quad \boldsymbol{\Sigma}(F_{\mathbf{A}\mathbf{y}+\mathbf{b}}) = \mathbf{A}\boldsymbol{\Sigma}(F_{\mathbf{y}})\mathbf{A}'.$$

The mean vector $E(\mathbf{y})$ and the covariance matrix $Cov(\mathbf{y}) = E[(\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))']$ serve as first examples. A scatter functional $\boldsymbol{\Sigma}$ possesses the so called *independence property*

if it is diagonal at random vectors with independent components. This does not follow from the AE property but it is true for the covariance matrix. Starting from any scatter matrix $\mathbf{\Sigma}$, a scatter matrix with independence property is obtained via symmetrization: $\mathbf{\Sigma}_{sym}(F_{\mathbf{y}}) := \mathbf{\Sigma}(F_{\mathbf{y}_1-\mathbf{y}_2})$ where $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent copies of $\mathbf{y}$ (Nordhausen & Tyler 2015).

Let $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$ be the eigendecomposition of $\mathbf{\Sigma}$. The eigenvalues in $\mathbf{\Lambda}$ and eigenvectors in $\mathbf{U}$ provide important tools to consider the multivariate dependence and variation of a random vector. The first two moments and the variance of the eigenvalues are given by

$$m_1(\mathbf{\Sigma}) = \frac{1}{p}tr(\mathbf{\Sigma}), \quad m_2(\mathbf{\Sigma}) = \frac{1}{p}tr(\mathbf{\Sigma}^2) \quad \text{and} \quad s^2(\mathbf{\Sigma}) = m_2(\mathbf{\Sigma}) - m_1^2(\mathbf{\Sigma}).$$

The arithmetic mean of the eigenvalues $m_1(\mathbf{\Sigma})$ and the geometric mean of the eigenvalues, the generalized variance $\det(\mathbf{\Sigma})^{1/p}$, are global measures of variation while $s^2(\mathbf{\Sigma})$ measures the deviation from sphericity in the elliptic case. Matrix $\mathbf{U}$ is sometimes called orientation of the distribution (and the data cloud) and the rotated random vector $\mathbf{U}'\mathbf{z}$ has uncorrelated components, the so called *principal components*.

As in the univariate case, location functionals $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and scatter functionals $\mathbf{\Sigma}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ can be used to build measures of multivariate skewness and kurtosis such as $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ and $tr\left(\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_2\right)$. All third and fourth moments of the standardized $\mathbf{z}$ can be collected into p vectors and p^2 matrices $E(z_i\mathbf{z})$, $i = 1, ..., p$ and $E(z_i z_j \mathbf{z}\mathbf{z}')$, $i, j = 1, ..., p$, and summing over i and $i = j$, respectively gives vector- and matrix-valued skewness and kurtosis measures $E(\mathbf{z}\mathbf{z}'\mathbf{z})$ and $E(\mathbf{z}\mathbf{z}'\mathbf{z}\mathbf{z}')$.

The sample mean vector and sample covariance matrix are $\bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^n \mathbf{y}_i$ and $\mathbf{S}_y = \frac{1}{n}\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$, respectively. For invertible $\mathbf{S}_y$, the *squared Mahalanobis distances* between $\mathbf{y}_i$ and $\bar{\mathbf{y}}$ and between $\mathbf{y}_i$ and $\mathbf{y}_j$ are given by $(\mathbf{y}_i - \bar{\mathbf{y}})'\mathbf{S}_y^{-1}(\mathbf{y}_i - \bar{\mathbf{y}})$ and $(\mathbf{y}_i - \mathbf{y}_j)'\mathbf{S}_y^{-1}(\mathbf{y}_i - \mathbf{y}_j)$, respectively.

Maximal Invariant Statistics

For the data matrix $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_n)' \in \Re^{n\times p}$, the matrix $\mathbf{D} = (\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{y}}')\mathbf{S}_y^{-1}(\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{y}}')' \in \Re^{n\times n}$ is *maximal invariant* under affine transformations $\mathbf{Y} \to \mathbf{Y}\mathbf{A}' + \mathbf{1}_n\mathbf{b}'$.

$\mathbf{D} = (\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{y}}')\mathbf{S}_y^{-1}(\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{y}}')'$ provides all squared Mahalanobis distances (d_{ii}) as well as squared Mahalanobis distances between the observations ($d_{ii} + d_{jj} - 2d_{ij}$). In fact, $(\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}')\hat{\mathbf{\Sigma}}^{-1}(\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}')'$ is maximal invariant for any choices of location and scatter functionals $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. Moreover, starting with a location functional $\boldsymbol{\mu}$ and two scatter functionals $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, the data matrix can be moved to an *invariant coordinate system (ICS)* by $\hat{\mathbf{Z}} \leftarrow (\mathbf{Y} - \mathbf{1}_n\hat{\boldsymbol{\mu}}')\hat{\mathbf{W}}$ where the transformation matrix $\hat{\mathbf{W}} \in \Re^{p\times p}$ satisfies $\hat{\mathbf{W}}'\hat{\mathbf{\Sigma}}_1\hat{\mathbf{W}} = \mathbf{I}_p$ and $\hat{\mathbf{W}}'\hat{\mathbf{\Sigma}}_2\hat{\mathbf{W}} = \hat{\mathbf{\Lambda}}$, where $\hat{\mathbf{\Lambda}}$ is a diagonal matrix with the diagonal elements in decreasing order. The matrix $\hat{\mathbf{Z}} \in \Re^{n\times p}$ is again maximal invariant and the $\mathbf{z}$-observations are centered with $\boldsymbol{\mu}$, standardized w.r.t $\mathbf{\Sigma}_1$ and also 'uncorrelated' w.r.t. $\mathbf{\Sigma}_2$ (Tyler et al. 2009; Ilmonen et al. 2012).

The Large-p-small-n Case

Tyler (2010) has shown that, if $n \leq p+1$, all AE location statistics are equal to the sample mean vector, and all AE scatter statistics are proportional to the sample covariance matrix. For $n = p+1$, the sample covariance matrix $\mathbf{S}_y$ is still invertible (w.p.1) and the maximal invariant statistic $(\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{y}}')\mathbf{S}_y^{-1}(\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{y}}')' = n\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n'$ is a constant matrix and all pairwise Mahalanobis distances are the same. Therefore when $n \leq p+1$, one cannot construct any affine invariant (coordinate-free) multivariate location tests, for example, and the requirement of affine equivariance/invariance must be relaxed.

3.3. Inference on Location and Scatter

3.3.1. M-estimation. Corresponding to any choices of positive weight functions w_1, w_2 and w_3 (not necessarily related to any distribution) the values of *location and scatter M-functionals* $\boldsymbol{\mu} \in \Re^p$ and $\boldsymbol{\Sigma} \in \Re_+^{p\times p}$ at the distribution of $\mathbf{y}$ satisfy the estimating equations

$$E[w_1(|\mathbf{z}|)\mathbf{z}] = \mathbf{0} \quad \text{and} \quad E[w_2(|\mathbf{z}|)\mathbf{z}\mathbf{z}'] = E[w_3(|\mathbf{z}|)]\mathbf{I}_p$$

with $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$. Note that the functionals $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are affine equivariant (AE) and, as in the univariate case, the above equations further suggest a fixed-point algorithm as well as corresponding one-step functionals. One has often $w_3(r) \equiv 1$ and the following choices. (i) $w_1(r) = w_2(r) = 1$ gives the mean vector and the covariance matrix. (ii) $w_1(r) = 1/r$ and $w_2(r) = p/r^2$ gives the AE spatial median and the so called Tyler's scatter matrix. (iii) For $f(\mathbf{z}) = \exp(-\rho(||\mathbf{z}||)$ and $\psi(r) = \rho'(r)$ then the ML functional is obtained as $w_1(r) = w_2(r) = \left|\frac{\psi(r)}{r}\right|$. (iv) Huber's estimate $w_1(r) = \min(k/r, 1)$ and $w_2(r) = \min(k^2/r^2, 1)$. The efficiency and robustness properties of M-functionals are determined by the weight functions. For a review and computation details see Dümbgen et al. (2015, 2016).

For a general discussion of the M-functionals, closely related S-functionals, the minimum covariance determinant (MCD) and the minimum volume ellipsoid (MVE), a multivariate extension of the Shorth, see Chapter 6 in Maronna et al. (2006) and the references therein. In the following we discuss the estimates and tests that are based on three different multivariate extensions of the concepts of sign, rank and signed-rank.

Multivariate Signs and Ranks

There is no natural ordering of multivariate observations. The multivariate notions of sign, rank and signed-rank are based on the multivariate extensions of the mean deviation and mean difference. The affine equivariance and invariance of the sign and rank methods then depend on the norm used in these extensions.

3.3.2. R-estimation and testing: Marginal sign and ranks. The univariate concepts of sign and rank are based on the complete ordering of univariate data points manifested with the univariate sign function. In the multivariate case there is no natural ordering but the concepts of sign and rank can be extended using certain objective or criterion functions. Let $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_n)'$ be a random sample from a p-variate distribution of $\mathbf{y}$. The first extensions

are obtained using the so called *Manhattan norm* or L_1 *norm* $||\mathbf{y}||_1 = |y_1| + ... + |y_p|$. The multivariate empirical signs, centered ranks, and signed-ranks may be implicitly defined through multivariate L_1 type objective functions

$$\sum_{i=1}^{n} ||\mathbf{y}_i||_1, \quad \sum_{i=1}^{n}\sum_{j=1}^{n} ||\mathbf{y}_i - \mathbf{y}_j||_1 \quad \text{and} \quad \sum_{i=1}^{n}\sum_{j=1}^{n} \left(||\mathbf{y}_i - \mathbf{y}_j||_1 + ||\mathbf{y}_i + \mathbf{y}_j||_1\right)$$

It is easy to see that the resulting scores $\mathbf{s}_i$, $\mathbf{r}_i$ and $\mathbf{q}_i$, $i = 1, ..., n$, are simply the vectors of *marginal (univariate) signs, centered ranks, and signed-ranks*. The population sign, centered rank and signed-rank functions $\mathbf{s}(\mathbf{z})$, $\mathbf{r}(\mathbf{z})$ and $\mathbf{q}(\mathbf{z})$ are vectors of corresponding marginal functions. For a complete treatment of this approach consult Puri & Sen (1971).

The vectors of marginal signs, ranks and centered ranks are not affine invariant or equivariant as their behavior under rotations $\mathbf{y}_i \to \mathbf{U}\mathbf{y}_i$ with orthogonal $\mathbf{U} \in \Re^{p\times p}$ is unpredictable. The vector of marginal ranks is however invariant under componentwise monotone transformations. The multivariate median $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}(\mathbf{Y})$, that is, the *vector of the marginal medians*, minimizes $\sum_{i=1}^{n} ||\mathbf{y}_i - \boldsymbol{\mu}||_1$ and centers the observations so that $\sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}) = \mathbf{0}$. Similarly, the *vector of marginal HL estimates* is the choice $\hat{\boldsymbol{\mu}}$ for which the signed-ranks $\hat{\mathbf{q}}_i$ of centered observations $\mathbf{y}_i - \hat{\boldsymbol{\mu}}$, $i = 1, ..., n$, satisfy $\sum_{i=1}^{n} \hat{\mathbf{q}}_i = \mathbf{0}$. The median is equivariant under monotone componentwise transformations but not affine equivariant. An affine equivariant median is obtained by using the so called transformation retransformation technique proposed by Chakraborty & Chaudhuri (1998). If $\hat{\mathbf{W}} \in \Re^{p\times p}$ is the ICS transformation matrix, the AE median (similarly HL estimate) is obtained as $(\hat{\mathbf{W}}')^{-1}\boldsymbol{\mu}(\mathbf{Y}\hat{\mathbf{W}})$. Invariant sign and rank based location tests are obtained using the marginal sign and ranks in the invariant coordinate system (Nordhausen et al. 2006).

Consistent estimation of the covariance matrix using the marginal signs and ranks is not possible but, in the elliptic case, $\frac{1}{n}\hat{\mathbf{s}}_i\hat{\mathbf{s}}_i'$ and $\binom{n}{2}^{-1}\sum_{i<j}\mathbf{s}(\mathbf{y}_j - \mathbf{y}_i)\mathbf{s}(\mathbf{y}_j - \mathbf{y}_i)'$ (Kendall's tau) are consistent to $\frac{2}{\pi}\sin^{-1}(Cor(\mathbf{y}_i, \mathbf{y}_j))$ and $\frac{1}{n}\sum_{i=1}^{n}\mathbf{r}_i\mathbf{r}_i'$ (Spearman's rho) and $\frac{1}{n}\sum_{i=1}^{n}\hat{\mathbf{q}}_i\hat{\mathbf{q}}_i'$ are consistent to $\frac{6}{\pi}\sin^{-1}\left(Cor(\mathbf{y}_i, \mathbf{y}_j)/2\right)$ which then surprisingly allows consistent estimation of the correlation matrix. See Visuri et al. (2000) and references therein.

3.3.3. R- estimation and testing: Spatial signs and ranks. Let $\mathbf{y}_1, ..., \mathbf{y}_n$ be again a random sample for a p-variate distribution $F_{\mathbf{y}}$. This approach is based on the use of *Euclidean or* L_2 *norm*

$$||\mathbf{y}||_2 = \left(y_1^2 + ... + y_p^2\right)^{1/2}.$$

The multivariate concepts of *spatial sign*, *spatial rank* and *spatial signed-rank* functions at F_y are then $\mathbf{s}(\mathbf{y}) = ||\mathbf{y}||_2^{-1}\mathbf{y}$ for $\mathbf{y} \neq \mathbf{0}$ and $\mathbf{s}(\mathbf{0}) = \mathbf{0}$, and

$$\mathbf{r}(\mathbf{y}) = E\left[\mathbf{s}(\mathbf{y} - \mathbf{y}_1)\right] \quad \text{and} \quad \mathbf{q}(\mathbf{y}) = \frac{1}{2}E\left[\mathbf{s}(\mathbf{y} - \mathbf{y}_1) + \mathbf{s}(\mathbf{y} + \mathbf{y}_1)\right],$$

respectively. The observed spatial signs, spatial ranks and spatial signed-ranks are $\mathbf{s}_i = \mathbf{s}(\mathbf{y}_i)$, $\mathbf{r}_i = \frac{1}{n}\sum_{j=1}^{n}\mathbf{s}(\mathbf{y}_i - \mathbf{y}_j)$ and $\mathbf{q}_i = \frac{1}{2n}\sum_{j=1}^{n}\left[\mathbf{s}(\mathbf{y}_i - \mathbf{y}_j) + \mathbf{s}(\mathbf{y}_i + \mathbf{y}_j)\right]$, $i = 1, ..., n$. Again, the signs, ranks, and signed-ranks are scores corresponding to the objective functions

$$\sum_{i=1}^{n} ||\mathbf{y}_i||_2, \quad \sum_{i=1}^{n}\sum_{j=1}^{n} ||\mathbf{y}_i - \mathbf{y}_j||_2 \quad \text{and} \quad \sum_{i=1}^{n}\sum_{j=1}^{n} \left(||\mathbf{y}_i - \mathbf{y}_j||_2 + ||\mathbf{y}_i + \mathbf{y}_j||_2\right).$$

See Oja (2010) for a complete discussion of this approach and for further references.

The spatial sign $\mathbf{s}_i$ is just a direction vector of length one whenever $\mathbf{y}_i \neq \mathbf{0}$. The centered ranks $\mathbf{r}_i$ and signed-ranks $\mathbf{q}_i$ have the lengths at most 1 and therefore lie in the unit p-ball. Roughly speaking, the direction of $\mathbf{r}_i$ tells in which direction $\mathbf{y}_i$ is from the center of the data cloud, and its length tells how far away this point is from the center. $\mathbf{q}_i$ is the rank of $\mathbf{y}_i$ among $\pm\mathbf{y}_1, ..., \pm\mathbf{y}_n$. The spatial signs, ranks and signed-ranks are orthogonal equivariant, that is, for any orthogonal $\mathbf{U} \in \Re^{p\times p}$, transformations $\mathbf{y}_i \to \mathbf{U}\mathbf{y}_i$, $i = 1, ..., n$, induce transformations $\mathbf{s}_i \to \mathbf{U}\mathbf{s}_i$, $\mathbf{r}_i \to \mathbf{U}\mathbf{r}_i$ and $\mathbf{q}_i \to \mathbf{U}\mathbf{q}_i$, $i = 1, ..., n$. Note also that the spatial ranks are invariant under location shifts $\mathbf{y}_i \to \mathbf{y}_i + \mathbf{b}$, $i = 1, ..., n$.

To estimate $\boldsymbol{\mu}$, let $\hat{\mathbf{s}}_i$ and $\hat{\mathbf{q}}_i$ be the signs and signed ranks of the shifted observations $\mathbf{y}_i - \hat{\boldsymbol{\mu}}$, $i = 1, ..., n$. The *spatial median* and *spatial HL estimate* are then choices of $\hat{\boldsymbol{\mu}}$ for which $\sum_{i=1}^n \hat{\mathbf{s}}_i = \mathbf{0}$ and $\sum_{i=1}^n \hat{\mathbf{q}}_i = \mathbf{0}$, respectively. They are shift and orthogonal equivariant. For fully affine equivariant estimation of $\boldsymbol{\mu}$, a simultaneous estimate of $\boldsymbol{\Sigma}$ is needed. Let then $\hat{\mathbf{s}}_i$ and $\hat{\mathbf{q}}_i$ be the signs and signed ranks of the standardized observations $\hat{\mathbf{z}}_i = \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})$, $i = 1, ..., n$. The *spatial median* $\hat{\boldsymbol{\mu}}$ and *Tyler's scatter estimate* $\hat{\boldsymbol{\Sigma}}$ then satisfy $\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{s}}_i = \mathbf{0}$ and $\frac{p}{n}\sum_{i=1}^n \hat{\mathbf{s}}_i\hat{\mathbf{s}}_i' = \mathbf{I}_p$, and together give the so called *Hettmansperger-Randles estimate* (Hettmansperger & Randles 2002). Similarly, for the affine equivariant spatial HL estimate and the related scatter estimate, the signed-ranks are standardized so that $\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{q}}_i = \mathbf{0}$ and $\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{q}}_i\hat{\mathbf{q}}_i' \propto \mathbf{I}_p$. See Tyler (1987) and Hettmansperger & Randles (2002). *Dümbgen's scatter estimate* (Dümbgen 1998), avoids location estimation and finds $\hat{\boldsymbol{\Sigma}}$ and $\hat{\mathbf{z}}_i = \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}_i$, $i = 1, ..., n$ such that $\binom{n}{2}^{-1}\sum_{i<j}\mathbf{s}(\hat{\mathbf{z}}_j - \hat{\mathbf{z}}_i))\mathbf{s}(\hat{\mathbf{z}}_j - \hat{\mathbf{z}}_i))' \propto \mathbf{I}_p$. Note that, in these approaches, $\boldsymbol{\Sigma}$ is estimated only up to its size, i.e., if $\hat{\boldsymbol{\Sigma}}$ is a solution, so is $c\hat{\boldsymbol{\Sigma}}$, for any $c > 0$.

If spatial sign and ranks are used to analyze the data, the *spatial sign and signed-rank covariance matrices* $E\left[\mathbf{s}(\mathbf{y} - \boldsymbol{\mu})\mathbf{s}(\mathbf{y} - \boldsymbol{\mu})'\right]$ and $E\left[\mathbf{q}(\mathbf{y} - \boldsymbol{\mu})\mathbf{q}(\mathbf{y} - \boldsymbol{\mu})'\right]$, as well as the *spatial Spearman's rho* and *spatial Kendall's tau* matrices $E\left[\mathbf{s}(\mathbf{y}_1 - \mathbf{y}_2)\mathbf{s}(\mathbf{y}_1 - \mathbf{y}_3)'\right]$ and $E\left[\mathbf{s}(\mathbf{y}_1 - \mathbf{y}_2)\mathbf{s}(\mathbf{y}_1 - \mathbf{y}_2)'\right]$, where $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ are independent copies of $\mathbf{y}$, provide simple and robust tools to estimate the eigenvectors of the covariance matrix. For an elliptical distribution, they have the same eigenvectors with the covariance matrix and, for the spatial sign and Kendall's tau matrices for example, there is a one-to-one correspondence between their eigenvalues and the eigenvalues of the covariance matrix. See again Visuri et al. (2000).

Finally, we mention that Chaudhuri (1996) considered the inverse of the spatial rank function and called it the *spatial quantile function*. Serfling (2004) gives a review of the inference methods based on the concept of the spatial quantile, and studies nonparametric measures of multivariate location, spread, skewness and kurtosis in terms of these quantiles.

3.3.4. R-estimation and testing: Affine equivariant signs and ranks. The previous extensions of the sign and rank concepts were based on the mean deviation and mean difference with the L_1 and L_2 norms, and the behavior of the signs and ranks were not predictable under affine transformations. In this section we discuss alternative criterion functions, also multivariate extensions of the mean deviation and mean difference yielding affine equivariant signs and ranks for the location and scatter problems. See Oja (1999) for a review.

Assume that $\mathbf{y}_1, ..., \mathbf{y}_n$ is a random sample from a distribution of $\mathbf{y}$. The volume of the simplex with $p + 1$ vertices $\mathbf{y}_1, ..., \mathbf{y}_{p+1} \in \Re^p$ is

$$V(\mathbf{y}_1, ..., \mathbf{y}_{p+1}) = \frac{1}{p!}\left|\det(\tilde{\mathbf{y}}_1, ..., \tilde{\mathbf{y}}_{p+1})\right|,$$

where $\tilde{\mathbf{y}} = (1, \mathbf{y}')' \in \Re^{p+1}$. The mean vector $E(\mathbf{y})$ then minimizes $E[V^2(\mathbf{y}_1, ..., \mathbf{y}_p, \boldsymbol{\mu})]$

and the generalized variance $[\det(Cov(\mathbf{y}))]^{1/p}$ is a constant times the minimum value $E[V^2(\mathbf{y}_1, ..., \mathbf{y}_p, E(\mathbf{y}))]$. This suggests the use of the averages of the volumes

$$V(\mathbf{0}, \mathbf{z}_{i_1}, ..., \mathbf{z}_{i_p}) \text{ or } V(\mathbf{z}_{i_1}, \mathbf{z}_{i_1}, ..., \mathbf{z}_{i_{p+1}})$$

over all subsets of the residuals $\mathbf{z}_1, ..., \mathbf{z}_n$ as multivariate extensions of the *mean deviation* and the *mean difference*, respectively.

Write $I = (i_1, ..., i_{p-1})$, $1 \leq i_1 < i_2 < ... < i_{p-1} \leq n$, for the $p-1$ vector of separate indices. Index I then refers to a $(p-1)$-set of observations or to a *hyperplane* going through the origin and $p-1$ data points, that is, $\{\mathbf{y} \in \Re^p : \det\left(\mathbf{y}_{i_1} \ ... \ \mathbf{y}_{i_{p-1}} \ \mathbf{y}\right) = 0\}$. Let next the vector $\mathbf{e}(I) \in \Re^p$ be implicitly defined as cofactors of $\mathbf{y}$ in $\det\left(\mathbf{y}_{i_1} \ ... \ \mathbf{y}_{i_{p-1}} \ \mathbf{y}\right) = \mathbf{e}'(I)\mathbf{y}$. The mean deviation is then a constant times $\binom{n}{p-1}^{-1} \sum_I |\mathbf{e}'(I)\mathbf{y}|$ and its gradient at $\mathbf{y}_i$ gives the p-variate affine equivariant sign or *Oja sign*

$$\mathbf{s}_i = \binom{n}{p-1}^{-1} \sum_I sign(\mathbf{e}'(I)\mathbf{y}_i)\mathbf{e}(I), \quad i = 1, ..., n.$$

The sign vectors are affine equivariant in the sense that the affine transformations $\mathbf{y}_i \to \mathbf{A}\mathbf{y}_i$, $i = 1, ..., n$, induce the transformations $\mathbf{s}_i \to |\det(\mathbf{A})|(\mathbf{A}^{-1})'\mathbf{s}_i$, $i = 1, ..., n$. The affine equivariant multivariate median $\hat{\boldsymbol{\mu}}$, called *Oja median*, is the choice to minimize the mean deviation of the residuals $\mathbf{y}_i - \boldsymbol{\mu}$, $i = 1, ..., n$ and then again $\sum_{i=1}^n \hat{\mathbf{s}}_i = 0$.

Find next in the similar way the affine equivariant multivariate ranks or *Oja ranks* of the observation vectors $\mathbf{y}_1, ..., \mathbf{y}_n$. In this case we need the p-subsets of observation with index sets $J = (i_1, ..., i_p)$, $1 \leq i_1 < i_2 < ... < i_p \leq n$. The *hyperplane* going through the observations listed in J is $\{\mathbf{y} \in \Re^p : \det\left(\tilde{\mathbf{y}}_{i_1} \ ... \ \tilde{\mathbf{y}}_{i_p} \ \tilde{\mathbf{y}}\right) = 0\}$. Define next $\mathbf{d}(J) = (d_0(J), \mathbf{d}_1(J)') \in \Re^{p+1}$ by the implicit equation $\det\left(\tilde{\mathbf{y}}_{i_1} \ ... \ \tilde{\mathbf{y}}_{i_p} \ \tilde{\mathbf{y}}\right) = \mathbf{d}'(J)\tilde{\mathbf{y}}$. The multivariate mean difference is then a constant times $\binom{n}{p}^{-1} \sum_J |\mathbf{d}'(J)\tilde{\mathbf{y}}|$ and its gradient at $\mathbf{y}_i$ gives the p-variate rank vector

$$\mathbf{r}_i = \binom{n}{p}^{-1} \sum_J sign(\mathbf{d}'(J)\tilde{\mathbf{y}}_i)\mathbf{d}_1(J), \quad i = 1, ..., n.$$

The rank vectors are affine equivariant and location invariant so that the affine transformations $\mathbf{y}_i \to \mathbf{A}\mathbf{y}_i + \mathbf{b}$, $i = 1, ..., n$, induce transformations $\mathbf{r}_i \to |\det(\mathbf{A})|(\mathbf{A}^{-1})'\mathbf{r}_i$, $i = 1, ..., n$. As in other extensions of signs and ranks, the signed-rank vector $\mathbf{q}_i$ for the ith observation $\mathbf{y}_i$ can be found as its rank among $2n$-set of observations $\pm\mathbf{y}_1, ..., \pm\mathbf{y}_n$, that is, among the original observations and their reflections. The affine equivariant HL estimate $\hat{\boldsymbol{\mu}}$ is the choice for which the sum of signed-ranks of the estimated residuals is zero.

Due to the peculiar type of affine equivariance of signs and ranks and signed-ranks, the sign and rank covariance matrices $\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{s}}_i\hat{\mathbf{s}}_i'$, $\frac{1}{n}\sum_{i=1}^n \mathbf{r}_i\mathbf{r}_i'$ and $\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{q}}_i\hat{\mathbf{q}}_i'$ are, in the elliptic case, consistent estimates of population quantities that are proportional to the inverse of the covariance matrix. This means for example that the eigenvector and proportional eigenvalues of the covariance matrix can be consistently estimated by them. For the influence functions and limiting distributions of these estimates, see Ollila et al. (2003a, 2004). Estimation of the sign covariance matrix, for example, is equivalent to finding $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ such that if $\hat{\mathbf{s}}_i$ is the sign of $\hat{\mathbf{z}}_i = \hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})$ among the $\hat{\mathbf{z}}_j$'s, $i = 1, ..., n$, then $\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{s}}_i = \mathbf{0}$ and $\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{s}}_i\hat{\mathbf{s}}_i' \propto \mathbf{I}_p$

How to Choose Between Multivariate Signs and Ranks?

Marginal signs and ranks are easiest to compute and their transformation re-transformation versions are most natural in an IC model. Affine equivariant spatial sign and rank methods are preferable in the elliptic case and easily applied using the R package `MNM`. The Oja sign and ranks have nice geometrical and theoretical properties but with highest computational burden. Our favourite location-scatter estimate is the Hettmansperger-Randles estimate.

We end this section with a discussion on related approaches in the literature. It is interesting that the mean deviation and the mean difference are proportional to the volumes of the *zonotope* and *lift zonotope* of $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_n)'$, that is, the Minkowski averages $Z(\mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n}[-\mathbf{y}_i, \mathbf{y}_i]$ and $LZ(\mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n}[-\tilde{\mathbf{y}}_i, \tilde{\mathbf{y}}_i]$, respectively. See Koshevoy et al. (2004) for the duality between Oja signs and ranks and zonotopes and lift zonotopes. Randles (1989) developed an affine invariant sign tests for the one and two sample cases based on *interdirection counts*, angular distances between $\mathbf{y}_i$ and $\mathbf{y}_j$ relative to the rest of data given, using our notation, by $\sum_I sign(\mathbf{e}'(I)\mathbf{y}_i)sign(\mathbf{e}'(I)\mathbf{y}_j)$. In a similar way, the distance between $\mathbf{y}_i$ and $\mathbf{y}_j$ can be defined by counting the number of data hyperplanes separating them, $\sum_J sign(\mathbf{d}'(J)\tilde{\mathbf{y}}_i)sign(\mathbf{d}'(J)\tilde{\mathbf{y}}_j)$, see Oja & Paindaveine (2005). A related and important approach is to combine Randles' interdirections (or spatial signs) and the univariate rank scores for the Mahalanobis distances from the origin. In this way, Hallin & Paindaveine (2002) developed optimal signed-rank location tests in the elliptic model. See Nordhausen et al. (2009) for a similar development in the independent component model. Related affine equivariant multivariate generalizations of the median working with hyperplanes and simplices are the half-space median by Tukey (1975) and Liu's median (Liu 1990).

3.3.5. Location estimates and tests based on a score function. Let $\mathbf{y}_1, ..., \mathbf{y}_n$ be a random sample from a p-variate distribution $f(\mathbf{y} - \boldsymbol{\mu})$ where f is symmetric around $\mathbf{0}$. Let $\mathbf{l}(\mathbf{z}) = -\nabla \log f(\mathbf{z})$ be the optimal multivariate location score. We use an odd score function $\mathbf{k} : \Re^p \to \Re^p$, coming for example from M- or R-estimation to draw inference on the unknown $\boldsymbol{\mu}$. For efficiency comparisons we need the matrices $\boldsymbol{\Omega} = E[\mathbf{k}(\mathbf{z})\mathbf{k}(\mathbf{z})'] \in \Re^{p\times p}$ and $\boldsymbol{\Gamma} = E[\mathbf{k}(\mathbf{z})\mathbf{l}(\mathbf{z})'] \in \Re^{p\times p}$.

Let $\hat{\boldsymbol{\mu}}$ center the observed scores by solving the estimation equation $\sum_{i=1}^{n}\mathbf{k}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}) = \mathbf{0}$. Under general assumptions and with the true location center $\boldsymbol{\mu}_0$, the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)$ is $N_p(\mathbf{0}, \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}(\boldsymbol{\Gamma}^{-1})')$. The geometrical mean of the eigenvalues of $\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}(\boldsymbol{\Gamma}^{-1})'$ is a global measure of variation that can be used to compare the estimates. The asymptotic relative efficiencies of the AE spatial median (resp. the AE spatial HL estimate) with respect to the mean vector are in 2-, 4- and 10-variate normal case are 0.78, 0.88 and 0.95 (resp. 0.97, 0.98 and 0.99). In the multivariate t_3 case, for example, the efficiencies are 2.00, 2.25 and 2.42 (resp. 1.95, 2.02 and 2.09). For other sign and rank based estimates and other distributions, see e.g. Oja (1999, 2010).

The companion M-test statistic for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is $\mathbf{T}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{k}(\mathbf{y}_i - \boldsymbol{\mu}_0)$ with the null distribution $N_p(\mathbf{0}, \boldsymbol{\Omega})$. Under a contiguous sequence of alternatives $H_{1,n} : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \frac{1}{\sqrt{n}}\delta$, the limiting distribution of $\sqrt{n}\mathbf{T}_n$ is $N_p(\boldsymbol{\Gamma}\delta, \boldsymbol{\Omega})$ and then $n\mathbf{T}_n'\boldsymbol{\Omega}^{-1}\mathbf{T}_n$ converges to a

non-central chi-squared distribution with p degrees of freedom non-centrality parameter $\delta'\mathbf{\Gamma}'\mathbf{\Omega}^{-1}\mathbf{\Gamma}\delta$. The Pitman efficiency with respect to Hotelling's T^2 test is then the ratio $\delta'\mathbf{\Gamma}'\mathbf{\Omega}^{-1}\mathbf{\Gamma}\delta/\delta'\mathbf{\Sigma}^{-1}\delta$ providing the same efficiencies as in the estimation case. The multivariate sign and signed-rank tests are not distribution-free but conditionally distribution-free sign-change test versions are naturally available for any choice of the odd score function k.

3.3.6. Testing for multivariate normality. Classical Mardia's test statistics for testing multivariate normality (Mardia 1970) are based on the maximal invariant statistic $\mathbf{D} = (d_{ij}) = (\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{y}}')\hat{\mathbf{S}}_y^{-1}(\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{y}}')'$ and computed as follows. First, multivariate measures of skewness and kurtosis are given by $b_1 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n d_{ij}^3$ and $b_2 = \frac{1}{n}\sum_{i=1}^n d_{ii}^4$ Then, under the null hypothesis of multinormality, the limiting distribution of $nb_1/6$ is a chi-square distribution with $p(p+1)(p+2)/6$ degrees of freedom and the limiting distribution of $\sqrt{n}(b_2 - p(kp+2))/\sqrt{8k(k+2)}$ is $N(0,1)$.

Starting from location functionals $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_1$ and scatter functionals $\mathbf{\Sigma}$, $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$, (rescaled to give the covariance matrix in the multivariate normal case), skewness and kurtosis statistics $(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'\hat{\mathbf{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)$ and $m_2(\hat{\mathbf{\Sigma}}_1^{-1}\hat{\mathbf{\Sigma}}_2 - \mathbf{I}_p)$ can also be used to test for multivariate normality. For testing ellipticity, the kurtosis should be replaced by $s^2(\hat{\mathbf{\Sigma}}_1^{-1}\hat{\mathbf{\Sigma}}_2)$. See Kankainen et al. (2007).

Also univariate tests of normality can be utilized in testing for multivariate normality. Then for $\mathbf{z}_i = \mathbf{S}_y^{-1/2}(\mathbf{y}_i - \bar{\mathbf{y}})$, $i = 1, ..., n$, the test statistic can be for example

$$\sup_{\mathbf{u}:\mathbf{u}'\mathbf{u}=1}\left(\frac{n}{6}\left[\sum_{i=1}^n(\mathbf{u}'\mathbf{z}_i)^3\right]^2 + \frac{n}{24}\left[\sum_{i=1}^n(\mathbf{u}'\mathbf{z}_i)^4 - 3\right]^2\right).$$

3.4. Multivariate Linear Regression

3.4.1. Linear regression model. The conditional distribution of $\mathbf{y}_i \in \Re^p$ given $\mathbf{x}_i \in \Re^q$, $i = 1, ..., n$, is that of

$$\mathbf{y}_i = \mathbf{A}\mathbf{z}_i + \mathbf{b}_0 + \mathbf{B}'\mathbf{x}_i$$

where $\mathbf{A} \in \Re^{p\times p}, \mathbf{b}_0 \in \Re^p, \mathbf{B} \in \Re^{q\times p}$, $\mathbf{z}_1, ..., \mathbf{z}_n$ is a random sample from the p-variate distribution of $\mathbf{z}$, $E(\mathbf{z}) = \mathbf{0}$ and $Cov(\mathbf{z}) = \mathbf{I}_p$. Again, one may assume that $\mathbf{z}$ has a $N_p(\mathbf{0}, \mathbf{I}_p)$-distribution or a distribution in other models listed in the location-scatter case. The data matrix is now $(\mathbf{X}, \mathbf{Y}) \in \Re^{n\times(q+p)}$ where $\mathbf{X}$ is a design matrix or the matrix of observational explaining variables with the same assumptions as in the univariate response variable case. The aim is to make inference on regression parameters $\mathbf{b}_0$ and $\mathbf{B}$ and scatter matrix $\mathbf{A}\mathbf{A}'$ based on the knowledge or assumptions on the distribution of $\mathbf{z}$.

3.4.2. M-estimates of the regression coefficients. Write $\mathbf{z}_i = \mathbf{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\beta}_0 - \boldsymbol{\beta}'\mathbf{x}_i)$, $i = 1, ..., n.$. The *M-estimates for multivariate regression* then solve

$$\sum_{i=1}^n w_1(|\mathbf{z}_i|)\tilde{\mathbf{x}}_i\mathbf{z}_i' = \mathbf{0} \quad \text{and} \quad \sum_{i=1}^n w_2(|\mathbf{z}_i|)\mathbf{z}_i\mathbf{z}_i' = \sum_{i=1}^n w_3(|\mathbf{z}_i|)\mathbf{I}_p$$

with some choices of positive weight functions w_1, w_2, w_3. One-step estimates are

$$\begin{pmatrix}\beta_0' \\ \boldsymbol{\beta}\end{pmatrix} \leftarrow \left[\sum_{i=1}^n w_1(|\mathbf{z}_i|)\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i'\right]^{-1}\sum_{i=1}^n w_1(|\mathbf{z}_i|)\tilde{\mathbf{x}}_i y_i'$$

and

$$\boldsymbol{\Sigma} \leftarrow \frac{\sum_{i=1}^n w_2(|\mathbf{z}_i|)(\mathbf{y}_i - \boldsymbol{\beta}_0 - \boldsymbol{\beta}'\mathbf{x}_i)(\mathbf{y}_i - \boldsymbol{\beta}_0 - \boldsymbol{\beta}'\mathbf{x}_i)'}{\sum_{i=1}^n w_3(|\mathbf{z}_i|)}.$$

3.4.3. R-estimates of the regression coefficients. Regression R-estimates in the first two sign-rank approaches are based on the objective functions

$$\sum_{i=1}^n \|\mathbf{y}_i - \boldsymbol{\beta}_0 - \boldsymbol{\beta}'\mathbf{x}_i\| \quad \text{and} \quad \sum_{i<j} \|(\mathbf{y}_j - \mathbf{y}_i) - \boldsymbol{\beta}'(\mathbf{x}_j - \mathbf{x}_i)\|$$

with L_1 or L_2 norms as in the location case. The estimates then solve the estimation equations

$$\sum_{i=1}^n \tilde{\mathbf{x}}_i \mathbf{s}(\mathbf{z}_i)' = \mathbf{0} \quad \text{and} \quad \sum_{i=1}^n \mathbf{x}_i \mathbf{r}_n(\mathbf{z}_i)' = \mathbf{0}$$

with a corresponding notion of sign and rank. Note that the rank estimate is only for $\mathbf{B}$ and, if needed, $\mathbf{b}_0$ must be estimated separately from the residuals $\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i'$, $i = 1, ..., n$ (Puri & Sen 1971; Oja 2010). Similar results for the Oja signs and ranks are still missing.

3.4.4. Elemental estimates of the regression coefficients. We briefly discuss the use of elemental estimates in the multivariate multiple regression. The hyperplane (fit) going through $p + q$ observations listed in $K = (i_1, ..., i_{p+q})$, $i_1 < ... < i_{p+q}$, is now

$$det \begin{pmatrix} 1 & 1 & ... & 1 & 1 \\ \mathbf{x}_{i_1} & \mathbf{x}_{i_2} & ... & \mathbf{x}_{i_{p+q}} & \mathbf{x} \\ \mathbf{y}_{i_1} & \mathbf{y}_{i_2} & ... & \mathbf{y}_{i_{p+q}} & \mathbf{y} \end{pmatrix} = (d_0(K), \mathbf{d}_1'(K), \mathbf{d}_2'(K)) \begin{pmatrix} 1 \\ \mathbf{x} \\ \mathbf{y} \end{pmatrix} = 0.$$

The elemental estimate, unbiased for $\mathbf{b}_0$ and $\mathbf{B}(K)$ are then

$$\mathbf{b}_0(K) = -\frac{d_0(K)\mathbf{d}_2(K)}{||\mathbf{d}_2(K)||} \quad \text{and} \quad \mathbf{B}(K) = -\frac{\mathbf{d}_1(K)\mathbf{d}_2'(K)}{||\mathbf{d}_2(K)||}.$$

The estimate may then be a weighted average of the elemental estimate such as

$$\left[\sum_K \mathbf{B}(K)\mathbf{W}(K)\right]\left[\sum_K \mathbf{W}(K)\right]^{-1}.$$

The LS estimate for example is obtained using $\mathbf{W}(K) = \mathbf{d}_2(K)\mathbf{d}_2(K)'$ (Ollila et al. 2003b).

3.5. R Packages for Multivariate Methods with an Example

Most multivariate methods based on the mean vector and covariance matrix and classical multivariate regression are a part of the default `R` installation. Various M-estimates of location and scatter as well as MVE and MCD are found in the packages `ICS` (Nordhausen et al. 2008b), `ICSNP` (Nordhausen et al. 2015), `fastM` (Dümbgen et al. 2014), `MASS` and `robustbase` for example. Methods based on marginal signs and ranks are for example in the package `ICSNP`. Spatial sign and rank methods are provided in `MNM` (Nordhausen & Oja 2011), which is almost a full implementation of all methods described in Oja (2010). The package `OjaNP` (Fischer et al. 2016) has several algorithms for the Oja median, signs and ranks.

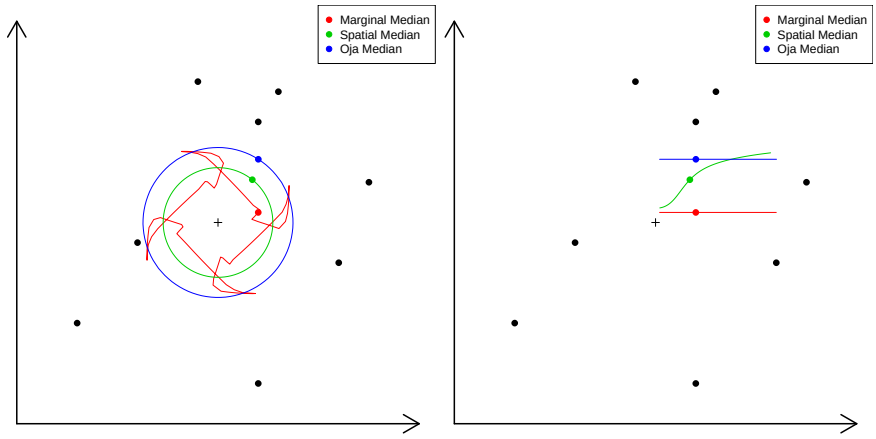

Figure 1

On the left, the colored curves describe how the medians move when the data points are rotated around the cross. On the right, the lines describe how the medians move when the x-axis is multiplied by a positive scalar.

In Figure 1 we visualize the equivariance properties of the three multivariate medians. The left panel shows how the medians move when the original data points shown in the figure are rotated around the cross by an orthogonal matrix. Orthogonal invariant estimators move along circles whereas the marginal median, which does not have this property, takes an unpredictable path. Similarly, in the right panel, the observed x-values are rescaled by a positive constant and scale equivariant estimators then move along parallel lines. The spatial median fails in an unpredictable way. We strongly encourage the use of affine equivariant methods and affine equivariant versions of spatial signs and rank methods for example are accessible in R package `MNM`.

In Figure 2 we compare the robustness of three affine equivariant location-scatter estimates: (i) Hettmansperger-Randles estimate based on spatial signs, (ii) AE spatial Hodges-Lehmann estimate with as a companion scatter estimate based on spatial signed-ranks, and (iii) the regular mean vector and covariance matrix. On the left panel, for a sample of size 200 from bivariate normal distribution, the three pairs of estimators are plotted and there is hardly any difference visible. Then, in the right panel, three points have been moved to the top left corner. While this clearly affects the regular mean and covariance matrix, the nonparametric robust estimates are only slightly affected by this change. Note the scatter matrix estimates are here visualized using the corresponding tolerance ellipses.

4. SUBSPACE ESTIMATION AND SCATTER MATRICES

4.1. Principal Component Analysis

4.1.1. Estimation of principal components. Assume that $\mathbf{y}_1, ..., \mathbf{y}_n$ is a random sample from a p-variate distribution of $\mathbf{y} = \mathbf{A}\mathbf{z} + \mathbf{b}$ where $\mathbf{A} \in \Re^{p \times p}, \mathbf{b} \in \Re^p$ $E(\mathbf{z}) = \mathbf{0}$ and $Cov(\mathbf{z}) = \mathbf{I}_p$. Write then $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}'$ for the covariance matrix of $\mathbf{y}$. In the principal component analysis the idea is to replace (without losing too much information) the original p variables in $\mathbf{y}$ by a much smaller number of *uncorrelated linear combinations* $\mathbf{v}_i'\mathbf{y}$, $i = 1, ..., q$, such that

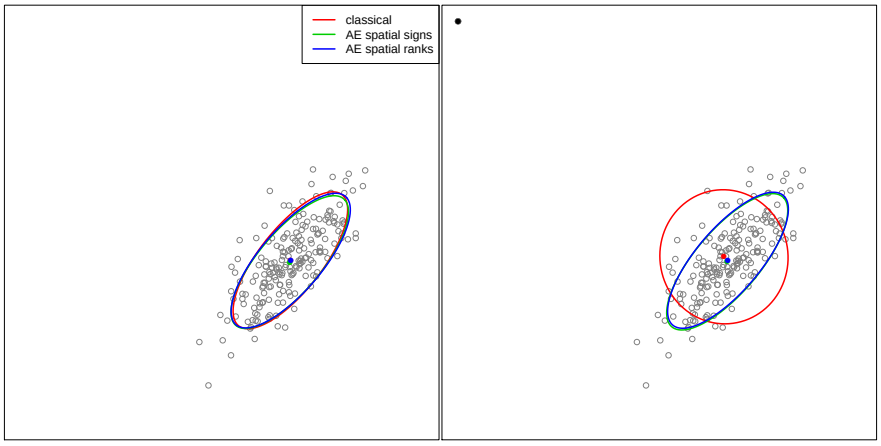

Figure 2

Behaviour of mean vector and covariance matrix and affine equivariant location-scatter estimates based on spatial signs and spatial signed-ranks, respectively, for the original (left panel) and contaminated (right panel) data.

$\mathbf{v}_i'\mathbf{v}_i = 1$, $i = 1, ..., q$. If $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}' = \sum_{i=1}^{p} \lambda_i \mathbf{u}_i \mathbf{u}_i'$ is the eigendecomposition of $\mathbf{\Sigma}$ with $\lambda_1 \geq ... \geq \lambda_q > \lambda_{q+1} \geq ... \geq \lambda_p$. then the sum of the variances $\sum_{i=1}^{q} Var(\mathbf{v}_i'\mathbf{y})$ (and the amount of information in this sense) is maximized if we choose $\mathbf{v}_i = \mathbf{u}_i$, $i = 1, ..., q$. For general review of PCA see for example Jolliffe (2002).

In the following assume that $\mathbf{z} \in \Re^p$ has a spherical distribution, that is $\mathbf{U}\mathbf{z} \sim \mathbf{z}$ for all orthogonal $\mathbf{U} \in \Re^{p \times p}$. It then holds under general assumptions that, if $\hat{\mathbf{\Sigma}}_z$ is the scatter matrix estimate obtained from $\mathbf{z}_1, ..., \mathbf{z}_n$, the limiting distribution of $\sqrt{n}\ vec(\hat{\mathbf{\Sigma}}_z - \mathbf{I}_p)$ is

$$N_{p^2}\left(\mathbf{0}, \sigma_1(\mathbf{I}_{p^2} + \mathbf{K}_{p,p}) + \sigma_2 vec(\mathbf{I}_p)vec(\mathbf{I}_p)'\right)$$

where $\mathbf{K}_{p,p} = \sum_{i=1}^{p}\sum_{j=1}^{p}(\mathbf{e}_i\mathbf{e}_j') \otimes (\mathbf{e}_j\mathbf{e}_i')$ is the so called commutation matrix. The limiting distribution depends only on two constants $\sigma_1 = AsVar(\mathbf{\Sigma}_{12})$ and $\sigma_2 = AsCov(\mathbf{\Sigma}_{11}, \mathbf{\Sigma}_{22})$ and then $AsVar(\mathbf{\Sigma}_{11}) = 2\sigma_1 + \sigma_2$. As $vec(\mathbf{\Sigma}_y) = (\mathbf{A} \otimes \mathbf{A})vec(\mathbf{\Sigma})_z$, the limiting distribution of $\hat{\mathbf{\Sigma}}_y$ is easily derived from that of $\hat{\mathbf{\Sigma}}_z$ (see e.g. Oja 2010).

Let next $\hat{\mathbf{\Sigma}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}' = \sum_{i=1}^{p} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i'$. be the eigendecomposition of $\hat{\mathbf{\Sigma}}$. Assume that the ith eigenvalue λ_i of $\mathbf{\Sigma}$ is distinct from other eigenvalues. Then $\sqrt{n}(\hat{\mathbf{u}}_i - \mathbf{u}_i)$ has a limiting p-variate normal distribution with zero mean vector and covariance matrix

$$AsCov(\hat{\mathbf{u}}_i) = \sigma_1 \sum_{j \neq i} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} \mathbf{u}_i \mathbf{u}_i', \quad i = 1, ..., q,$$

and the limiting distribution of $\sqrt{n}(\hat{\lambda}_i - \lambda_i)$ is a normal distribution with zero mean and variance $2\sigma_1 + \sigma_2$. For the regular covariance matrix and $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$, $\sigma_1 = 1$ and $\sigma_2 = 0$. The influence function of the scatter functional $\mathbf{\Sigma}$ at a spherical F_z is

$$IF(\mathbf{z}; \mathbf{\Sigma}, F) = \alpha(||\mathbf{z}||)\frac{\mathbf{z}\mathbf{z}'}{||\mathbf{z}||^2} - \beta(||\mathbf{z}||)\mathbf{I}_p$$

and, if function α is known, $\sigma_1 = E(\alpha^2(||\mathbf{z}||))/(p(p+2))$ can be consistently estimated from the data $\mathbf{y}_1, ..., \mathbf{y}_n$. See e.g. Tyler (1983); Croux & Haesbrock (2000). For details about

PCA using multivariate sign and rank covariance matrices see for example Locantore et al. (1999),Croux et al. (2002), Oja (2010) and Taskinen et al. (2012).

4.1.2. Testing for dimension. Write $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}$ for the eigendecomposition of $\mathbf{\Sigma}$ where the diagonal elements of $\mathbf{\Lambda}$ are $\lambda_1 \geq ... \geq \lambda_p > 0$. We wish to test the null hypothesis of *subsphericity*,

$$H_{0q}: \quad \lambda_1 \geq ... \geq \lambda_q > \lambda_{q+1} = ... = \lambda_p = \lambda \quad \text{for some unknown } \lambda.$$

It is then thought that the q first principal components carry all the information and the last $p-q$ components represent spherical noise. To test the null hypothesis, a natural test statistic is the variance of the $p-q$ smallest eigenvalues of $\hat{\mathbf{\Sigma}}$, that is,

$$T_q := s^2(\hat{\mathbf{U}}_q'\hat{\mathbf{\Sigma}}\hat{\mathbf{U}}_q), \quad \text{where} \quad \hat{\mathbf{U}}_q := \underset{\mathbf{U}\in\mathcal{O}^{p\times(p-q)}}{\arg\min} \; s^2(\mathbf{U}'\hat{\mathbf{\Sigma}}\mathbf{U})$$

and $\mathcal{O}^{p\times(p-q)}$ is the set of $p \times (p-q)$ matrices with orthonormal columns. For the test construction we also need to estimate one population constant σ_1, namely, the limiting variance of the off-diagonal element of $\hat{\mathbf{\Sigma}}$ in a spherical case; see the discussion above. The unknown λ must be estimated as well; a consistent estimate is given by $\hat{\lambda} = m_1(\hat{\mathbf{U}}_q'\hat{\mathbf{\Sigma}}\hat{\mathbf{U}}_q)$. Then, under H_{0q},

$$\frac{n(p-q)T_q}{2\hat{\lambda}^2\hat{\sigma}_1} \to_d \chi^2_{(p-q-1)(p-q+2)/2}.$$

For these results, estimation of q and bootstrap testing for the same problem, see for example Nordhausen et al. (2016) and references therein.

4.2. Independent Component Analysis and Related Methods

4.2.1. Estimation of the unmixing matrix. In the semiparametric *independent component (IC) model* it is assumed that $\mathbf{y}_1, ..., \mathbf{y}_n$ is a random sample from a distribution of

$$\mathbf{y} = \mathbf{A}\mathbf{z} + \mathbf{b},$$

where $\mathbf{A} \in \Re^{p\times p}$ is non-singular, $\mathbf{b} \in \Re^p$, and $\mathbf{z}$ is a random p-vector with independent and standardized components, that is, $E(\mathbf{z}) = \mathbf{0}$ and $Cov(\mathbf{z}) = \mathbf{I}_p$. We assume that q components of $\mathbf{z}$ are non-Gaussian (signal), and $p-q$ components Gaussian (noise). In the classical IC model, at most one component is Gaussian, that is, $q \geq p-1$. The idea is then to estimate the transformations back to the non-Gaussian components and to test for the dimension.

In classical *fourth order blind identification (FOBI)* (Cardoso 1989) one uses two moment based scatter matrices,

$$\mathbf{\Sigma}_1 = E\left[(\mathbf{y}-E(\mathbf{y}))(\mathbf{y}-E(\mathbf{y}))'\right] \quad \text{and} \quad \mathbf{\Sigma}_2 = E\left[r^2(\mathbf{y}-E(\mathbf{y}))(\mathbf{y}-E(\mathbf{y}))'\right],$$

where $r^2 = (\mathbf{y}-E(\mathbf{y}))'\mathbf{\Sigma}_1^{-1}(\mathbf{y}-E(\mathbf{y}))$. Then a matrix $\mathbf{W} = (\mathbf{w}_1, ..., \mathbf{w}_p) \in \Re^{p\times p}$ and a diagonal matrix $\mathbf{\Lambda} \in \Re^{p\times p}$ are found such that

$$\mathbf{W}'\mathbf{\Sigma}_1\mathbf{W} = \mathbf{I}_p \quad \text{and} \quad \mathbf{W}'\mathbf{\Sigma}_2\mathbf{W} = \mathbf{\Lambda}.$$

The matrix $\mathbf{W}$ is called an *unmixing matrix* as, under certain assumptions (including that there is at most one gaussian component), $\mathbf{W}'\mathbf{y} = \mathbf{z}$ up to the order, shifts and signs of the

components. The eigenvalues in $\mathbf{\Lambda}$ measure the kurtoses of the marginal variables in $\mathbf{Wy}$ and equal $p+2$ for gaussian components. The eigenvalues can then be used to separate the Gaussian and non-Gaussian components. Notice that if $\mathbf{U} \in \Re^{p\times p}$ is the matrix of eigenvectors of $\mathbf{R} := \mathbf{\Sigma}_1^{-1/2}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1/2}$ then $\mathbf{W}' = \mathbf{U}'\mathbf{\Sigma}_1^{-1/2}$. FOBI can be robustified by replacing the moment based scatter matrices by any robust scatter matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ possessing the independence property (Oja et al. 2006; Nordhausen et al. 2008a).

In the *Joint Approximate Diagonalization of Eigenmatrices (JADE)* (Cardoso & Souloumiac 1993) all possible fourth order cumulants are used. The fourth order cumulants of $\mathbf{z} = Cov(\mathbf{y})^{-1/2}(\mathbf{y} - E(\mathbf{y}))$ are given in the matrices $\mathbf{C}^{ij} = E\left(\mathbf{zz}'\mathbf{E}^{ij}\mathbf{zz}'\right) - \mathbf{E}^{ij} - \mathbf{E}^{ji} - tr(\mathbf{E}^{ij})\mathbf{I}_p$, where $\mathbf{E}^{ij} = \mathbf{e}_i\mathbf{e}_j'$, $i,j = 1,\ldots,p$. If an orthogonal $\mathbf{U}$ is chosen to maximize $\sum_{i=1}^p\sum_{j=1}^p \left\|diag\left(\mathbf{U}'\mathbf{C}^{ij}\mathbf{U}\right)\right\|^2$, the independent components are surprisingly found in $\mathbf{U}'Cov(\mathbf{y})^{-1/2}\mathbf{y}$. For a throughout discussion of FOBI and JADE, see Miettinen et al. (2015b). The third main approach is the so called *projection pursuit* or *fastICA* (Huber 1985; Hyvärinen & Oja 1997): Choose a affine invariant measure $G(y)$ of non-gaussianity of a univariate random variable y such as

$$G(y) = \alpha E^2\left[\left(\frac{y-E(y)}{Var(y)^{1/2}}\right)^3\right] + (1-\alpha)E^2\left[\left(\frac{y-E(y)}{Var(y)^{1/2}}\right)^4 - 3\right]$$

and then find $\mathbf{W} = (\mathbf{w}_1,...,\mathbf{w}_q)' \in \Re^{q\times p}$ to maximize $\sum_{i=1}^q G(\mathbf{w}_i'\mathbf{y})$ under the constraint $\mathbf{W\Sigma W}' = \mathbf{I}_q$. For more details, see Virta et al. (2016); Miettinen et al. (2017a). For the use of signs and ranks in ICA, see Ilmonen & Paindaveine (2011); Hallin & Mehta (2015). Finally, for a general overview of ICA, see Comon & Jutten (2010).

4.2.2. Inference on the dimension of non-Gaussian subspace. In this section we discuss the use of the eigenvalues of the estimated FOBI matrix $\hat{\mathbf{R}} = \hat{\mathbf{\Sigma}}_1^{-1/2}\hat{\mathbf{\Sigma}}_2\hat{\mathbf{\Sigma}}_1^{-1/2}$ for the inference on unknown q. We then need to assume that, for the q-subvector of non-Gaussian components of $\mathbf{z}$, say $\mathbf{z}_{NG}$, the fourth moments exist and $E[(\mathbf{u}'\mathbf{z}_{NG})^4] \neq 3$ for all $\mathbf{u}'\mathbf{u} = 1$, $\mathbf{u} \in \Re^q$. We wish to test the null hypothesis

$$H_{0,q}: \quad \text{exactly } p-q \text{ eigenvalues of } \mathbf{R} \text{ equal } p+2\ ,$$

stating that the dimension of the signal space is q. To test the null hypothesis $H_{0,q}$, we use the test statistic

$$T_q = \min_{\mathbf{U}\in\mathcal{O}^{p\times(p-q)}} m_2\left(\mathbf{U}'(\hat{\mathbf{R}} - (p+2)\mathbf{I}_p)\mathbf{U}\right),$$

that is, the smallest sum of squared distances of $p-k$ eigenvalues from $p+2$. If $H_{0,q}$ is true, then under general assumptions

$$n(p-q)T_q \to_d \sim 2\sigma_1\chi^2_{(p-q-1)(p-q+2)/2} + (2\sigma_1 + \sigma_2(p-q))\,\chi^2_1$$

with independent chi-squared variables $\chi^2_{(p-q-1)(p-q+2)/2}$ and χ^2_1 and $\sigma_1 = Var\left(\|z\|^2\right) + 8$ and $\sigma_2 = 4$. Further, if $(c_{k,n})$ is a sequence such that $c_{k,n} \to \infty$ and $\frac{c_{k,n}}{n} \to 0$ as $n \to \infty$, then $\hat{q} = \min\{k \ : \ n(p-k)T_k < c_{k,n}\} \to_P q$. For bootstrap testing strategies and more details, see Nordhausen et al. (2016, 2017).

Recall that FOBI is just a simple special case of the so called two-scatter method; $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are then replaced by any two scatter matrices. Deriving asymptotic tests or using bootstrap testing strategy with $T_q = \min_{\mathbf{U}\in\mathcal{O}^{p\times(p-q)}} s^2\left(\mathbf{U}'\hat{\mathbf{R}}\mathbf{U}\right)$ is then also an option for

testing $H_{0,q}$. The test above is valid also in the wider *Non-Gaussian Component Analysis (NGCA)* model where the non-Gaussian and Gaussian parts are still independent but the Gaussian components are allowed to be mutually dependent. See Nordhausen et al. (2017).

4.3. Sliced Inverse Regression and Related Methods

4.3.1. Subspace estimation with known dimension. In a regression context, let

$$(\mathbf{y}, \mathbf{X}) = \left(\begin{pmatrix} y_1 \\ \mathbf{x}_1 \end{pmatrix}, ..., \begin{pmatrix} y_n \\ \mathbf{x}_n \end{pmatrix} \right)' \in \Re^{n \times (p+1)}$$

be a random sample from a distribution of $(y, \mathbf{x}')'$ where now $\mathbf{x} = \mathbf{A}\mathbf{z} + \mathbf{b}$, with non-singular $\mathbf{A} \in \Re^{p \times p}$, $\mathbf{b} \in \Re^p$ and random $\mathbf{z} = (\mathbf{z}_1', \mathbf{z}_2')'$ such that $E(\mathbf{z}) = \mathbf{0}$, $Cov(\mathbf{z}) = \mathbf{I}_p$ and $(y, \mathbf{z}_1')'$ and $\mathbf{z}_2$ are independent. We assume that the partition $\mathbf{z} = (\mathbf{z}_1', \mathbf{z}_2')'$ is the unique one with the smallest q (up to inner rotations). The aim is to estimate the projections to the signal subspace of $\mathbf{x}$ determined by $\mathbf{z}_1$. Our assumption on the independence between $(y, \mathbf{z}_1')'$ and $\mathbf{z}_2$ is stronger than the regular assumptions in sliced inverse regression and therefore allows easier derivations of the results sketched below. For more details, see Nordhausen et al. (2017).

The *sliced inverse regression (SIR)* (Li 1991) again uses two scatter matrices, namely,

$$\mathbf{\Sigma}_1 := E\left[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))'\right] \quad \text{and} \quad \mathbf{\Sigma}_2 := E\left[E(\mathbf{x} - E(\mathbf{x})|y)E(\mathbf{x} - E(\mathbf{x})|y)'\right].$$

Note that the second scatter matrix is *supervised* in the sense that it depends on the joint distribution of $\mathbf{x}$ and y. One then finds a transformation matrix $\mathbf{W} \in \Re^{p \times p}$ and a diagonal matrix $\mathbf{\Lambda} \in \Re^{p \times p}$ such that $\mathbf{W}'\mathbf{\Sigma}_1\mathbf{W} = \mathbf{I}_q$ and $\mathbf{W}'\mathbf{\Sigma}_1\mathbf{W} = \mathbf{\Lambda}$ where the diagonal elements of $\mathbf{\Lambda}$ are in a decreasing order. Again, $\mathbf{W}' = \mathbf{U}'\mathbf{\Sigma}_1^{-1/2}$ with an orthogonal $\mathbf{U} \in \Re^{p \times p}$ that is the matrix of eigenvectors of $\mathbf{R} := \mathbf{\Sigma}_1^{-1/2}\mathbf{\Sigma}_2\mathbf{\Sigma}_1^{-1/2}$. In practice, the random variable y is replaced by its discrete approximation $\sum_{h=1}^{H} y_h 1_{y \in S_h}$ for some disjoint intervals (slices) $S_1, \ldots, S_H$ such that $\Re = S_1 + \ldots + S_H$ and for some choices $y_h \in S_h$, $h = 1, \ldots, H$. The first q columns of $\mathbf{W}$ give the transformation to $\mathbf{z}_1$ (up to rotation).

In the *sliced average variance estimate (SAVE)* (Cook & Weisberg 1991) one uses the variation in conditional covariance matrices to find the subspace and the second scatter matrix is then $\mathbf{\Sigma}_2 = E\left[(Cov(\mathbf{x}|y) - Cov(\mathbf{x}))\, Cov(\mathbf{x})^{-1}\, (Cov(\mathbf{x}|y) - Cov(\mathbf{x}))\right]$. In the *canonical correlation analysis (CCA)* for example the second scatter matrix is $\mathbf{\Sigma}_2 = Cov(\mathbf{x}, \mathbf{y})Cov(\mathbf{y})^{-1}Cov(\mathbf{y}, \mathbf{x})$. The *directional regression (DR)* (Li & Wang 2007) approach and *principal Hessian directions (PHD)* (Li 1992) can be written as well using the regular covariance matrix and a supervised scatter matrix (Liski et al. 2014). For the influence functions of these approaches see for example Prendergast & Smith (2010) and references therein and a robust version of SIR was suggested in Gather et al. (2001).

4.3.2. Inference on unknown dimension. We wish to test the null hypothesis

$$H_{0q}: \quad \text{exactly } p - q \text{ eigenvalues of } \mathbf{R} \text{ are zero,}$$

stating that the dimension of the signal space is exactly q. Let again $\hat{\mathbf{R}} = \hat{\mathbf{\Sigma}}_1^{-1/2}\hat{\mathbf{\Sigma}}_2\hat{\mathbf{\Sigma}}_1^{-1/2}$. The number of slices in the estimation is assumed to be large enough so that $H > q + 1$ To test the null hypothesis, we use the test statistic that is the average of the $p - q$ smallest eigenvalues, that is,

$$T_q := \min_{\mathbf{U} \in \mathcal{O}^{p \times (p-q)}} m_1(\widehat{\mathbf{U}}'\widehat{\mathbf{R}}\widehat{\mathbf{U}}).$$

Then, under H_q and under general assumptions (including that $H > q + 1$), we have $n(p-q)T_q \to_d \chi^2_{(p-q)(H-q-1)}$. For further details and a bootstrapping testing strategy, see for example Li (1991), Bura & Cook (2001) and Nordhausen et al. (2016).

4.4. R Packages for Subspace Estimation with an Example

PCA has many implementations in `R`, already the standard `R` has two functions for it. Our preferred packages in `R` for ICA are `ICS`, `JADE` (Miettinen et al. 2017b) and `fICA` (Miettinen et al. 2015a). For many ICA estimators in these packages asymptotic quantities can be computed using the package `BSSasymp` (Miettinen et al. 2017b). Supervised dimension reductions methods like SIR, SAVE and PHD are available in the package `dr` (Weisberg 2002). The subspace dimension estimation methods as discussed here are implemented in `ICtest` (Nordhausen et al. 2016).

Using these packages we first visualize the different subspaces. Figure 3 shows in the left panel a random sample from an IC model with one uniform and one Gaussian component. The next three panels then show the data after standardization, the principal

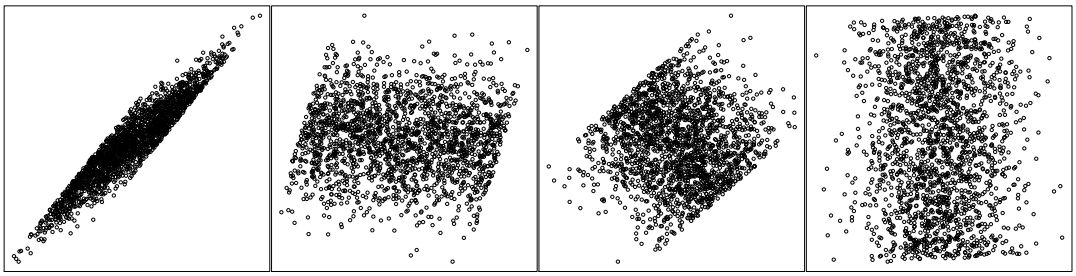

Figure 3

The left panel shows observations from a bivariate IC model with one uniform and one Gaussian component. The next three panels are then (from left to right) the standardized components, the principal components and the independent components.

components based on the regular covariance matrix and the independent components using FOBI. Recall that the independent components are found just by rotating and/or reflecting the standardized data until the observations are correlated also w.r.t. the second scatter matrix.

To conclude this section we consider the data in the left panel of Figure 4 which looks like well behaving Gaussian or elliptic data. Regular PCA gives the principal components in the middle panel showing some tendency towards non-Gaussiniaty. The FOBI solution in the right panel fully recovers the hidden structure of non-Gaussian and Gaussian parts. The asymptotic test described in Section 4.2.2 yields for the null hypotheses of four Gaussian components and that of three Gaussian components p-values < 0.001 and 0.281, respectively, indicating that there are indeed three normal components independent from the bivariate signal.

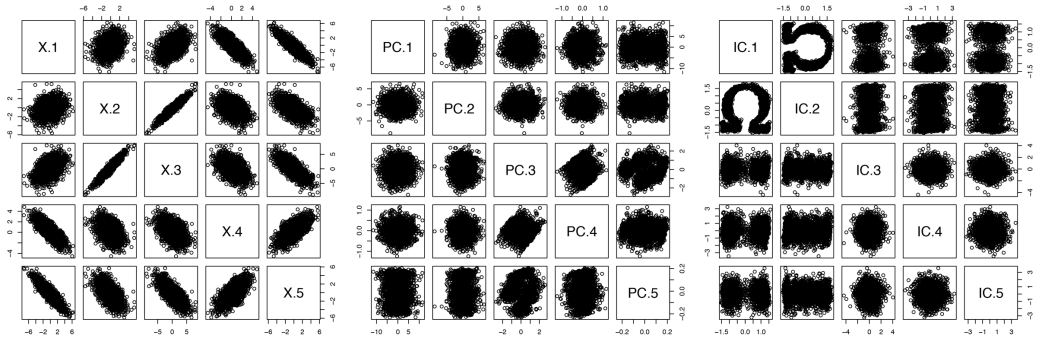

Figure 4

The left scatter plot matrix shows the observed data, the middle one the principle components and the right one independent components from the FOBI approach.

5. ANALYSIS OF DEPENDENT AND FUNCTIONAL DATA

5.1. Clustered Data

Clustered data problems are encountered almost everywhere in applied research. A typical situation is that instead of sampling n independent and identically distributed random variables, the observations come in m clusters with known cluster memberships. To state the assumptions to model clustered data, we write $\mathbf{y}_{i1}, ..., \mathbf{y}_{in_i}$ for the n_i p-variate observations in the ith cluster, $i = 1, ..., m$. Let $\mathbf{x}_{ij}$ be a possible vector of (design or observational) explanatory variables for the response $\mathbf{y}_{ij}$, $i = 1, ..., m$; $j = 1, ..., n_i$. Assume first that the cluster sizes $n_1, \ldots, n_m$ are fixed design constants. In the linear regression model it is then often assumed that,

(i) the p-variate distributions of $\mathbf{z}_{ij} = \mathbf{y}_{ij} - \mathbf{b}_0 - \mathbf{B}'\mathbf{x}_{ij}$ are all the same,
(ii) all the joint $2p$-variate distributions of $\mathbf{z}_{ij}$ and $\mathbf{z}_{ij'}$, $j \neq j'$ are the same, and
(iii) $\mathbf{z}_{ij}$ and $\mathbf{z}_{i'j'}$, $i \neq i'$ are independent.

If joint multivariate normality of the random errors $\mathbf{z}_{ij}$ can be assumed (parametric model), regular multivariate mixed models with cluster random effects can be employed. Alternatively, for sign and rank based methods for example, variance adjusted test and estimating procedures based on different weighting schemes all provide valid statistical inference. If $n_1, ..., n_m$ are random and the joint distribution of the random errors $\mathbf{z}_{ij}$ does not depend on $n_1, ..., n_m$, it is still reasonable to assume that (i)–(iii) hold conditionally on $n_1, ..., n_m$. For sign and rank tests, see for example Kloke et al. (2009), Konietschke & Brunner (2009) and Nevalainen et al. (2010) and references therein.

A much more complex setting arises when the random cluster size n_i may have an influence on the measured values $\mathbf{x}_{ij}$ and $\mathbf{y}_{ij}$, or vice versa, possibly due to an unobservable latent variable. The setting is termed *informative cluster size*, because the cluster size could then carry information about the quantities or parameters of interest. The standard approaches are not sufficient any more if the cluster size is informative. Nevalainen et al. (2014) introduced extensions of common population quantities of interest—such as location and scale functionals, and regression coefficients including M-functionals and R-functionals

and discussed their sample counterparts as well. Nevalainen et al. (2017) considered the problem of testing for informative cluster size.

`R` packages useful in clustered data analysis are for example `lme4` (Bates et al. 2015), `robustlmm` (Koller 2016), `lqmm` (Geraci 2014), `ClusterRankTest` (Dutta & Datta 2016) and `mvctm` (Larocque 2017).

5.2. Matrix Valued Data

In some applications, the independent and identically distributed random variables $\mathbf{Y}_1, ..., \mathbf{Y}_n$ are structured as $p \times q$ matrices rather than as pq vectors. For multivariate clustered data with p measurements on q individuals in each cluster, for example, the measurements in a single cluster may be seen as a random matrix $\mathbf{Y} \in \Re^{p\times q}$. For repeated measures design with p variables and q repetitions, the observations for each individual are $p \times q$ matrices. In applications of matrix and, more generally, tensor valued data, the problem itself often suggests the Kronecker structure in modelling so that

$$\mathbf{Y} = \mathbf{A}_L \mathbf{Z} \mathbf{A}_R' + \mathbf{B} \ \in \Re^{p\times q},$$

where $\mathbf{A}_L \in \Re^{p\times p}$, $\mathbf{A}_R \in \Re^{q\times q}$, $\mathbf{B} \in \Re^{p\times q}$ and $\mathbf{Z} \in \Re^{p\times q}$ is a standardized matrix valued random variable with $E(\mathbf{Z}) = \mathbf{0}$ and $Cov(vec(\mathbf{Z})) = \mathbf{I}_{pq}$. Thus, $E(\mathbf{Y}) = \mathbf{B}$ and the covariance matrix of the vectorized observation has the Kronecker covariance structure, $Cov(vec(\mathbf{Y})) = (\mathbf{A}_R\mathbf{A}_R') \otimes (\mathbf{A}_L\mathbf{A}_L')$. If the columns of $\mathbf{Y}$ are exchangeable random vectors, as is the case with clustered data, then $\mathbf{A}_R\mathbf{A}_R' \propto (1-\rho)\mathbf{I}_q + \rho\mathbf{1}_q\mathbf{1}_q^T$ (the intraclass correlation structure).

The parametric multivariate normal model or the semiparametric elliptical model are obtained if one assumes that $vec(\mathbf{Z})$ has the $N_{pq}(\mathbf{0}, \mathbf{I}_{pq})$ distribution or a spherically symmetric distribution, respectively. In the independent component model, one again assumes that the elements of $\mathbf{Z}$ are independent. See for example Gupta & Nagar (2010) for an overview of matrix-valued distributions. In the multivariate normal case Srivastava et al. (2008) introduced likelihood ratio test for the null hypotheses of Kronecker covariance structure. For other approaches to this estimation problem, see Wiesel (2012) and Ros et al. (2016). Sun et al. (2016) for example considered robust estimation of a structured covariance matrix, including Kronecker covariance structure, under heavy-tailed elliptical distributions. Analysis tools for matrix and tensor valued observations such as principal component analysis, independent component analysis and sliced inverse regression have been increasingly discussed in the literature. See e.g. Virta et al. (2016,a,b); Virta & Nordhausen (2017) and the discussions and references therein.

5.3. Multivariate Time Series

In this section we consider p-variate time series $\mathbf{y} = (\mathbf{y}_t)_{t=0,\pm1,\pm2}$. For a full-rank matrix $\mathbf{A} \in \Re^{p\times p}$ and a vector $\mathbf{b} \in \Re^p$, $\mathbf{A}\mathbf{y}+\mathbf{b}$ is then a time series with values $(\mathbf{A}\mathbf{y}+\mathbf{b})_t = \mathbf{A}\mathbf{y}_t+\mathbf{b}$, $t = 0, \pm1, \pm2, ...$ We assume that the observed p-variate time series $\mathbf{y}_1, ..., \mathbf{y}_T$ follows the model $\mathbf{y} = \mathbf{A}\mathbf{z} + \mathbf{b}$, where $\mathbf{b} \in \Re^p$ is a location vector, $\mathbf{A} \in \Re^{p\times p}$ is a mixing matrix, and the components of the p-variate marginal time series in $\mathbf{z}$ are uncorrelated and stationary with $E(\mathbf{z}_t) = \mathbf{0}$ and $Cov(\mathbf{z}_t) = \mathbf{I}_p$. The goal in the so called *blind source separation (BSS)* is to find an unmixing matrix functional $\mathbf{W} = \mathbf{W}(\mathbf{y}) \in \Re^{p\times p}$ such that the marginal time series in $\mathbf{W}(\mathbf{y})\mathbf{y}$ are standardized and uncorrelated.

The most popular BSS methods designed for time series is *SOBI (Second Order Blind Identification)* (Belouchrani et al. 1997). SOBI uses the cross-autocovariance matrices $\mathbf{\Sigma}_\tau(\mathbf{y}) = E\left[(\mathbf{y}_t - E(\mathbf{y}_t))(\mathbf{y}_{t+\tau} - E(\mathbf{y}_t))'\right]$, $\tau = 0, 1, \ldots, K$, and finds $\mathbf{W}$ that maximizes $\sum_{\tau=1}^{K}\sum_{i=1}^{p}\left(\mathbf{\Sigma}_\tau(\mathbf{W}\mathbf{y})\right)_{ii}^2$ under the constraint $\mathbf{\Sigma}_0(\mathbf{W}\mathbf{y}) = \mathbf{I}_p$. Different algorithms for this so called joint diagonalization and the statistical properties of the resulting estimates are discussed for example Miettinen et al. (2014, 2016). SOBI is a valid procedure if the latent time series are linear processes but does not work for times series exhibiting stochastic volatility. For such cases, time series extensions of FOBI and JADE using fourth order cross-moment matrices were suggested in Matilainen et al. (2015). For BSS methods that work with spatial signs of $\mathbf{y}_t$, see Nordhausen (2014); Ilmonen et al. (2015).

5.4. Functional Data

Functional observations are often assumed to belong to $\mathcal{H} = L^2(0,1)$, a separable Hilbert space with the scalar product scalar product $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$, and norm $||f|| = \langle f, f\rangle^{1/2}$. In the separable Hilbert space $\mathcal{H} = L^2(0,1)$, there exists an *orthonormal basis* (f_j), that is, $\langle f_j, f_k\rangle = \delta_{jk}$ such that $\mathcal{H} = \{y \ : \ y = \sum_{j=1}^{\infty} y_j f_j\}$. In practice, one chooses orthonormal basis functions (g_j) such as (i) Fourier basis functions, (ii) spline functions, (iii) wavelets, (iv) basis functions based on kernel smoothing of data functions, etc. The final data to be analyzed is then often located in a truncated space $\{y \ : \ y = \sum_{j=1}^{q} y_j g_j\}$.

In functional data analysis, one speaks about *linear operators* rather than matrices. The class of *Hilbert-Schmidt operators* is particularly interesting: A linear operator $A : \mathcal{H} \to \mathcal{H}$ is Hilbert-Schmidt if it allows a singular value decomposition (SVD) $Ay = \sum_{j=1}^{\infty} \lambda_j \langle g_j, y\rangle f_j$ with two orthonormal bases (g_j) and (f_j) and a sequence (λ_j) of non-negative real numbers such that $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$. Operator A is then an infinite weighted sum of elemental (tensor product) operators $f \otimes e : y \to \langle e, y\rangle f$.

By the Riesz representation theorem, there is a function $E(y) \in \mathcal{H}$, called the *mean function*, such that $\langle f, E(y)\rangle = E\langle f, y\rangle, f \in \mathcal{H}$. The *covariance operator* $Cov(y) : \mathcal{H} \to \mathcal{H}$ is defined by $Cov(y)(f) = E\left(\langle f, y - E(y)\rangle (y - E(y))\right), \quad f \in \mathcal{H}$. We can also write $Cov(y) = E\left((y - E(y)) \otimes (y - E(y))\right)$. The covariance operator $Cov(y)$ is also an integral operator as $(Cov(y)f)(t) = \int K(t,s)f(s)ds$ with $K(t,s) = Cov(y(t), y(s))$. It is easy to see that $E(y)$ and $Cov(y)$ are affine equivariant in the sense that $E(Ay + b) = AE(y) + b$ and $Cov(Ay + b) = ACov(y)A^*$ for all functions $b \in \mathcal{H}$ and for all linear operators $A : \mathcal{H} \to \mathcal{H}$. Our conjecture is that there is no other location function or scatter operator possessing the affine equivariant property. For recent extensions of (unitary equivariant/invariant) spatial sign and ranks methods for the data in infinite-dimensional Banach spaces, see e.g. Chakraborty & Chaudhuri (2014) and the references therein.

The covariance operator $Cov(y)$ is symmetric and positive definite and, if λ_j, e_j, $j = 1, 2, ...$, are the eigenvalues and eigenfunctions of $Cov(y)$, $\lambda_1 \geq \lambda_2 \geq ...$ and $\sum_j \lambda_j < \infty$, the eigendecomposition of $Cov(y)$ is $\sum_{j=1}^{\infty} \lambda_j e_j \otimes e_j$. Further, by Karhunen-Loeve expansion, $y = \sum_{i=1}^{\infty} z_i e_i$ where $z_i = \langle e_i, y\rangle$ are uncorrelated random variables, that is, *functional principal components* with variances λ_i, $i = 1, 2, ...$ Many statistical methods, such as linear regression, principal component analysis, canonical correlation and sliced inverse regression have been extended to the functional setting as well. See, for example, Ramsay & Silverman (2005) , Yao et al. (2005a,b), Ferraty & Vieu (2006), Horvath & Kokoszka (2012), Ferre & Yao (2003, 2005), Hsing & Ren (2009) and Li & Song (2017).

SUMMARY POINTS

1. In the univariate case, the concepts of sign and rank may be seen as location scores corresponding to the mean deviation and mean difference. The tests are often distribution-free in wide nonparametric models. Most methods are implemented in R and are easy to apply in practice.
2. Multivariate signs, ranks and signed-ranks are based on multivariate extensions of the mean deviation and mean difference. They offer efficient and fairly robust alternatives to analyze multivariate data. The choice between different extensions depend on the model assumptions and required equivariance/invariance properties. The tests are conditionally distribution-free with known asymptotics. The R package `MNM` is available for the computation of the spatial procedures for example.
3. The model assumptions can be tested using skewness and kurtosis statistics which are often based on simultaneous uses of several location and scatter statistics. It is discussed how scatter statistics can be used in PCA, ICA, SIR and in estimating spherical or Gaussian subspaces and their dimensions.
4. Recent extensions of semiparametric methods to clustered, matrix valued, time series and functional data analysis are briefly discussed.

FUTURE ISSUES

1. Robust methods have been developed mainly for neighboring models in the iid case. In the multivariate case, ellipticity is often assumed. The literature on robust methods for ICA, BSS, time series, clustered, matrix valued or functional data is still sparse and the proper theory is often missing. Robust and efficient tools are needed for dimension reduction and subspace estimation as well.
2. There are huge challenges in the analysis of high dimensional and functional data: What is for example the role of affine equivariance/invariance in the case of functional data? What can we say about affine equivariance of the covariance operator in the 'true' functional space if the observed functions are in practice in a truncated finite dimensional space?
3. Asymptotics in the small-n-large-p case: It is easy to understand the idea of consistency and limiting normality of an estimate for fixed p and $n \to \infty$. The normal approximation is then good if $n \gg p$. However, we often have data sets with $p \gg n$ but there is no logical way of choosing the convergence rates for n and p or the families of distributions with increasing dimension p.

Disclosure statement

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Acknowledgements

This work was supported by the Academy of Finland (grant 268703).

LITERATURE CITED

Anderson TW. 2003. An introduction to multivariate statistical analysis. Hobocken, US: John Wiley & Sons, 3rd ed.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1–48

Belouchrani A, Abed-Meraim K, Cardoso JF, Moulines E. 1997. A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* 45:434–44

Bura E, Cook RD. 2001. Extending sliced inverse regression: The weighted chi-squared test. *J. Am. Stat. Assoc.* 96:996–03

Cardoso JF. 1989. Source separation using higher order moments, In *Proceedings of IEEE International Conference on Accoustics, Speech and Signal Processing*

Cardoso JF, Souloumiac A. 1993. Blind beamforming for non-gaussian signals. *IEE Proc. F* 140:362–70

Chakraborty A, Chaudhuri P. 2014. The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Stat.* 42:1203–31

Chakraborty B, Chaudhuri P. 1998. On an adaptive transformation and retransformation estimate of multivariate location. *J. R. Stat. Soc. Series B Stat. Methodol.* 60:145–57

Chaudhuri P. 1996. On a geometric notion of quantiles for multivariate data. *J. Am. Stat. Assoc.* 91:862–68

Chernozhukov V, Galichon A, Hallin M, Henry M. 2017. Monge-Kantorovich depth, quantiles, ranks and signs. *Ann. Stat.* 45:223–56

Comon P, Jutten C, eds. 2010. Handbook of blind source sepparation. Independent component analysis and applications. Oxford: Academic Press

Cook RD, Weisberg S. 1991. Sliced inverse regression for dimension reduction: Comment. *J. Am. Stat. Assoc.* 86:328–32

Croux C, Haesbrock G. 2000. Principal component analysis based on robust estimators of the covariance and correlation matrix: Influence functions and efficiencies. *Biometrika* 87:603–18

Croux C, Ollila E, Oja H. 2002. Sign and rank covariance matrices: Statistical properties and application to principal component analysis. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, ed. Y Dodge. Basel, Switzerland: Birkäuser Verlag, 257–69

Dümbgen L. 1998. On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Stat. Math.* 50:471–91

Dümbgen L, Nordhausen K, Schuhmacher H. 2014. fastM: Fast computation of multivariate M-estimators. Https://CRAN.R-project.org/package=fastM

Dümbgen L, Nordhausen K, Schuhmacher H. 2016. New algorithms for M-estimation of multivariate scatter and location. *J. Multivar. Anal.* 144:200–17

Dümbgen L, Pauly M, Schweizer T. 2015. M-functionals of multivariate scatter. *Stat. Surv.* 9:32–105

Dutta S, Datta S. 2016. ClusterRankTest: Rank tests for clustered data. Https://CRAN.R-project.org/package=ClusterRankTest

Ferraty F, Vieu P. 2006. Nonparametric functional data analysis: Theory and practice. New York: Springer

Ferre L, Yao AF. 2003. Functional sliced inverse regression analysis. *Statistics* 37:475–88

Ferre L, Yao AF. 2005. Smoothed functional inverse regression. *Stat. Sin.* 15:665–83

Fischer D, Mosler K, Möttönen J, Nordhausen K, Pokotylo O, Vogel D. 2016. Computing the Oja Median in R: The Package OjaNP. *Preprint available as arXiv:1606.07620*

Fischer D, Oja H. 2015. Mann-Whitney type tests for microarray experiments: The R package gMWT. *J. Stat. Softw.* 65:1–19

Gather U, Hilker T, Becker C. 2001. A robustified version of sliced inverse regression. In *Statistics in Genetics and Environmental Sciences*, eds. L Fernholz, S Morgenthaler, W Stahel. Basel, Switzerland: Birkhäuser Verlag, 147–57

Geraci M. 2014. Linear quantile mixed models: The lqmm package for laplace quantile regression. *J. Stat. Softw.* 57:1–29

Gupta AK, Nagar DK. 2010. Matrix variate distributions. Boca Raton: Chapman & Hall/CRC

Hajek Z, Sidak J. 1967. Theory of rank tests. New York: Academic Press

Hallin M, Mehta C. 2015. R-estimation for asymmetric independent component analysis. *J. Am. Stat. Assoc.* 110:218–32

Hallin M, Paindaveine D. 2002. Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Stat.* 30:1103–33

Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. 1986. Robust statistics: The approach based on influence functions. New York, USA: John Wiley & Sons

Hettmansperger TP, Aubuchon JC. 1988. Comment on "Rank-based robust analysis of linear models. I. Exposition and review" by David Draper. *Stat. Sci.* 3:262–63

Hettmansperger TP, McKean JW. 2011. Robust nonparametric statistical methods. Boca Ranton: CRC Press, 2nd ed.

Hettmansperger TP, Randles RH. 2002. A practical affine equivariant multivariate median. *Biometrika* 89:851–60

Horvath L, Kokoszka P. 2012. Inference for functional data with applications. New York: Springer

Hosking JRM. 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Series B Stat. Methodol.* 52:105–24

Hothorn T, Hornik K. 2017. exactRankTests: Exact distributions for rank and permutation tests. Https://CRAN.R-project.org/package=exactRankTests

Hsing T, Ren H. 2009. An RKHS formulation of the inverse regression dimension-reduction problem. *Ann. Stat.* 37:726–55

Huber PJ. 1980. Robust statistics. New York, USA: John Wiley & Sons

Huber PJ. 1985. Projection pursuit. *Ann. Stat.* 13:435–75

Hyvärinen A, Oja E. 1997. A fast fixed-point algorithm for independent component analyis. *Neural Comput.* 9:1483–92

Ilmonen P, Nordhausen K, Oja H, Theis F. 2015. An affine equivariant robust second order blind source separation method. In *Latent Variable Analysis and Signal Separation*, eds. E Vincent, A Yeredor, Z Koldovsky, P Tichavsky. Springer, 328–35

Ilmonen P, Oja H, Serfling R. 2012. On invariant coordinate system (ICS) functionals. *Int. Stat. Rev.* 80:93–110

Ilmonen P, Paindaveine D. 2011. Semiparametrically efficient inference based on signed ranks in symmetric independent components models. *Ann. Stat.* 39:2448–76

Jolliffe IT. 2002. Prinicpal component analysis. New York, US: Springer, 2nd ed.

Jurečková J, Sen PK, Picek J. 2013. Methodology in robust and nonparametric statistics. Boca Ranton: CRC Press

Kankainen A, Taskinen S, Oja H. 2007. Tests of multinormality based on location vectors and scatter matrices. *Stat. Method Appl.* 16:357–79

Kloke JD, McKean JW. 2012. Rfit: Rank-based estimation for linear models,. *R J.* 4:57–64

Kloke JD, McKean JW, Rashid MM. 2009. Rank-based estimation and associated inferences for linear models with cluster correlated errors. *J. Am. Stat. Assoc.* 104:384–90

Koenker R. 2016. quantreg: Quantile regression. Https://CRAN.R-project.org/package=quantreg

Koller M. 2016. robustlmm: An R package for robust estimation of linear mixed-effects models. *J. Stat. Softw.* 75:1–24

Konietschke F, Brunner E. 2009. Nonparametric analysis of clustered data in diagnostic trials: Estimation problems in small sample sizes. *Comput. Stat. Data. Anal.* 53:730–41

Koshevoy G, Möttönen J, Oja H. 2004. On the geometry of multivariate L1 objective functions.

All. Stat. Arch. 88:137–54

Larocque D. 2017. mvctm: Multivariate variance components tests for multilevel data. Https://CRAN.R-project.org/package=mvctm

Lehmann EL. 1975. Nonparametrics: Statistical methods based on ranks. San Francisco, USA: Holden-Day

Li B, Song J. 2017. Nonlinear sufficient dimension reduction for functional data. *Ann. Stat.* :to appear

Li B, Wang S. 2007. On directional regression for dimension reduction. *J. Am. Stat. Assoc.* 102:997–08

Li KC. 1991. Sliced inverse regression for dimension reduction. *J. Am. Stat. Assoc.* 86:316–27

Li KC. 1992. On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Am. Stat. Assoc.* 87:1025–39

Liski E, Nordhausen K, Oja H. 2014. Supervised invariant coordinate selection. *Statistics* 48:711–31

Liu R. 1990. On a notion of data depth based upon random simplices. *Ann. Stat.* 18:405–14

Locantore N, Marron JS, Simpson DG, Tripoli N, Zhang JT, Cohen KL. 1999. Robust principal component analysis for functional data. *Test* 8:1–73

Mächler M, Rousseeuw P, Croux C, Valentin Todorov AR, Salibian-Barrera M, et al. 2016. robustbase: Basic robust statistics. Https://CRAN.R-project.org/package=robustbase

Mardia KV. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57:519–30

Mardia KV, Kent JT, Bibby JM. 1979. Multivariate analysis. Orlando: Academic Press

Maronna RA, Martin RD, Yohai VJ. 2006. Robust statistics - theory and methods. Chichester, UK: John Wiley & Sons

Matilainen M, Nordhausen K, Oja H. 2015. New independent component analysis tools for time series. *Stat. Probab. Lett.* 105:80–87

Miettinen J, Illner K, Nordhausen K, Oja H, Taskinen S, Theis F. 2016. Separation of uncorrelated stationary time series using autocovariance matrices. *J. Time Ser. Anal.* 37:337–54

Miettinen J, Nordhausen K, Oja H, Taskinen S. 2014. Deflation-based separation of uncorrelated stationary time series. *J. Multivar. Anal.* 123:214–27

Miettinen J, Nordhausen K, Oja H, Taskinen S. 2015a. fICA: Classical, reloaded and adaptive FastICA algorithms. Https://CRAN.R-project.org/package=fICA

Miettinen J, Nordhausen K, Oja H, Taskinen S, Virta J. 2017a. The squared symmetric FastICA estimator. *Signal Process.* 131:402–11

Miettinen J, Nordhausen K, Taskinen S. 2017b. Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp. *J. Stat. Softw.* 76:1–31

Miettinen J, Taskinen S, Nordhausen K, Oja H. 2015b. Fourth moments and independent component analysis. *Stat. Sci.* 30:372–90

Nevalainen J, Datta S, Oja H. 2014. Inference on the marginal distribution of clustered data with informative cluster size. *Stat. Pap.* 55:71–92

Nevalainen J, Larocque D, Oja H, Pörsti I. 2010. Nonparametric analysis of clustered multivariate data. *J. Am. Stat. Assoc.* 105:864–72

Nevalainen J, Oja H, Datta S. 2017. Tests for informative cluster size using a novel balanced bootstrap scheme. *Stat. Med.* :n/a–n/a

Nordhausen K. 2014. On robustifying some second order blind source separation methods for nonstationary time series. *Stat. Pap.* 55:141–56

Nordhausen K, Oja H. 2011. Multivariate L1 methods: The package MNM. *J. Stat. Softw.* 43:1–28

Nordhausen K, Oja H, Ollila E. 2008a. Robust independent component analysis based on two scatter matrices. *Aust. J. Stat.* 37:91–100

Nordhausen K, Oja H, Paindaveine D. 2009. Signed-rank tests for location in the symmetric independent component model. *J. Multivar. Anal.* 100:821–34

Nordhausen K, Oja H, Tyler DE. 2006. On the efficiency of invariant multivariate sign and rank

test. In *Festschrift for Tarmo Pukkila on his 60th Birthday*, eds. EP Liski, J Isotalo, J Niemelä, S Puntanen, GPH Styan. Tampere, Finland: University of Tampere, 217–31

Nordhausen K, Oja H, Tyler DE. 2008b. Tools for exploring multivariate data: The package ICS. *J. Stat. Softw.* 28:1–31

Nordhausen K, Oja H, Tyler DE. 2016. Asymptotic and bootstrap tests for subspace dimension. *Preprint available as arXiv:1611.04908*

Nordhausen K, Oja H, Tyler DE, Virta J. 2016. ICtest: Estimating and testing the number of interesting components in linear dimension reduction. Https://CRAN.R-project.org/package=ICtest

Nordhausen K, Oja H, Tyler DE, Virta J. 2017. Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace. *Preprint available as arXiv:1701.06836*

Nordhausen K, Sirkiä S, Oja H, Tyler DE. 2015. ICSNP: Tools for multivariate nonparametrics. Https://CRAN.R-project.org/package=ICSNP

Nordhausen K, Tyler DE. 2015. A cautionary note on robust covariance plug-in methods. *Biometrika* 102:573–88

Oja H. 1981. On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Stat.* 8:154–68

Oja H. 1999. Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scand. J. Stat.* 26:319–43

Oja H. 2010. Multivariate nonparametric methods with R. an approach based on spatial signs and ranks. New York: Springer

Oja H, Paindaveine D. 2005. Optimal signed-rank tests based on hyperplanes. *J. Stat. Plann. Inf.* 135:300–23

Oja H, Sirkiä S, Eriksson J. 2006. Scatter matrices and independent component analysis. *Aust. J. Stat.* 35:175–89

Ollila E, Croux C, Oja H. 2004. Influence function and asymptotic efficiency of the affine equivariant rank covariance matrix. *Stat. Sin.* 14:297–16

Ollila E, Oja H, Croux C. 2003a. The affine equivariant sign covariance matrix: asymptotic behavior and efficiencies. *J. Multivar. Anal.* 87:328–55

Ollila E, Oja H, Koivunen V. 2003b. Estimates of regression coefficients based on lift rank covariance matrix. *J. Am. Stat. Assoc.* 98:90–98

Pearson K. 1895. Contributions to the mathematical theory of evolution, ii: Skew variation in homogeneous material. *Phil. Trans. R. Soc.* 186:343–14

Prendergast LA, Smith JA. 2010. Influence functions for dimension reduction methods: An example influence study of principal hessian direction analysis. *Scand. J. Stat.* 37:588–611

Puri ML, Sen PK. 1971. Nonparametric methods in multivariate analysis. New York, USA: John Wiley & Sons

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Https://www.R-project.org/

Ramsay JO, Silverman BW. 2005. Functional data analysis. New York: Springer

Randles RH. 1989. A distribution-free multivariate sign test based on interdirections. *J. Am. Stat. Assoc.* 84:1045–50

Ros B, Bijma F, de Munck JC, de Gunst MC. 2016. Existence and uniqueness of the maximum likelihood estimator for models with a Kronecker product covariance structure. *J. Multivar. Anal.* 143:345–61

Santambrogio F. 2015. Optimal transport for applied mathematicians. Basel, Switzerland: Birkhäuser Verlag

Serfling RJ. 1980. Approximation theorems of mathematical statistics. New York, USA: John Wiley & Sons

Serfling RJ. 2004. Nonparametric multivariate descriptive measures based on spatial quantiles. *J. Stat. Plann. Inf.* 123:259–78

Srivastava MS, von Rosen T, von Rosen D. 2008. Models with a Kronecker product covariance

structure: estimation and testing. *Math. Methods Stat.* 17:357–70

Sun Y, Babu P, Palomar DP. 2016. Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions. *IEEE Trans. Signal Process.* 64:3576–90

Taskinen S, Koch I, Oja H. 2012. Robustifying principal component analysis with spatial sign vectors. *Stat. Probab. Lett.* 82:765–74

Tukey J. 1975. Mathematics and the picturing of data. *Proceedings of International Congress of Mathematics* 2:523–31

Tyler DE. 1983. The asymptotic distribution of principal component roots under local alternatives to multiple roots. *Ann. Stat.* 11:1232–42

Tyler DE. 1987. A distribution-free M-estimator of multivariate scatter. *Ann. Stat.* 15:234–51

Tyler DE. 2010. A note on multivariate location and scatter statistics for sparse data sets. *Stat. Probab. Lett.* 80:1409–13

Tyler DE, Critchley F, Dümbgen L, Oja H. 2009. Invariant coordinate selection. *J. R. Stat. Soc. Series B Stat. Methodol.* 71:549–92

Venables WN, Ripley BD. 2002. Modern applied statistics with S. New York: Springer, 4th ed.

Virta J, Li B, Nordhausen K, Oja H. 2016a. Independent component analysis for tensor-valued data. *Preprint available as arXiv:1602.00879*

Virta J, Li B, Nordhausen K, Oja H. 2016b. JADE for tensor-valued observations. *Preprint available as arXiv:1603.05406*

Virta J, Nordhausen K. 2017. Blind source separation of tensor-valued time series. *Preprint available as arXiv:1703.10381*

Virta J, Nordhausen K, Oja H. 2016. Projection pursuit for non-gaussian independent components. *Preprint available as arXiv:1612.05445*

Virta J, Taskinen S, Nordhausen K. 2016. Applying fully tensorial ICA to fMRI data, In *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*

Visuri S, Koivunen V, Oja H. 2000. Sign and rank covariance matrices. *J. Stat. Plann. Inf.* 91:557–75

Weisberg S. 2002. Dimension reduction regression in R. *J. Stat. Softw.* 7:1–22

Wiesel A. 2012. Geodesic convexity and covariance estimation. *IEEE Trans. Signal Process.* 60:6182–89

Yao F, Müller HG, Wang JL. 2005a. Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* 100:577–90

Yao F, Müller HG, Wang JL. 2005b. Functional linear regression analysis for longitudinal data. *Ann. Stat.* 33:2873–03