IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 182–193

DOI: 10.1214/10-IMSCOLL718

Asymptotic theory of the spatial median

Jyrki Möttönen[1,*], **Klaus Nordhausen**[2,3] **and Hannu Oja**[2]

University of Helsinki, University of Tampere and Tampere University Hospital

Abstract: In this paper we review, prove and collect the results on the limiting behavior of the regular spatial median and its affine equivariant modification, the transformation retransformation spatial median. Estimation of the limiting covariance matrix of the spatial median is discussed as well. Some algorithms for the computation of the regular spatial median and its different modifications are described. The theory is illustrated with two examples.

1. Introduction

For a set of p-variate data points $\mathbf{y}_1, \dots, \mathbf{y}_n$, there are several versions of multivariate median and related multivariate sign test proposed and studied in the literature. For some reviews, see Small [23], Chaudhuri and Sengupta [6] and Niinimaa and Oja [17]. The so called spatial median which minimizes the sum $\sum_{i=1}^n |\mathbf{y}_i - \boldsymbol{\mu}|$ with a Euclidean norm $|\cdot|$ has a very long history, Gini and Galvani [8] and Haldane [10] for example have independently considered the spatial median as a generalization of the univariate median. Gover [9] used the term mediancenter. Brown [3] has developed many of the properties of the spatial median. This minimization problem is also sometimes known as the Fermat-Weber location problem, see Vardi and Zhang [25]. Taking the gradient of the objective function, one sees that if $\hat{\boldsymbol{\mu}}$ solves the equation $\sum_{i=1}^n \{\mathbf{U}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})\} = \mathbf{0}$ with spatial sign $\mathbf{U}(\mathbf{y}) = |\mathbf{y}|^{-1}\mathbf{y}$, then $\hat{\boldsymbol{\mu}}$ is the observed spatial median. The spatial sign test for $H_0 : \boldsymbol{\mu} = \mathbf{0}$ based on the sum of spatial signs, $\sum_{i=1}^n \mathbf{U}(\mathbf{y}_i)$ was considered by Möttönen and Oja [14], for example.

The spatial median is unique, if the dimension of the data cloud is greater than one, see Milasevic and Ducharme [13]. The so called Weiszfeld algorithm for the computation of the spatial median has a simple iteration step, namely $\mu \leftarrow \mu + \{\sum_{i=1}^n |\mathbf{y}_i - \boldsymbol{\mu}|^{-1}\}^{-1} \sum_{i=1}^n \{\mathbf{U}(\mathbf{y}_i - \boldsymbol{\mu})\}$. The algorithm may fail sometimes, however, but a slightly modified algorithm which converges quickly and monotonically is described by Vardi and Zhang [25].

One drawback of the spatial median (the spatial sign test) is the lack of equivariance (invariance) under affine transformations of the data. The performance of the spatial median as well as the spatial sign test then may be poor compared to affine equivariant and invariant procedures if there is a significant deviance from

*Corresponding author
[1]Department of Social Research, University of Helsinki P.O.Box 68, 00014 University of Helsinki, Finland, e-mail: jyrki.mottonen@helsinki.fi
[2]Tampere School of Public Health, University of Tampere, 33014 University of Tampere, Finland, e-mail: klaus.nordhausen@uta.fi; hannu.oja@uta.fi
[3]Department of Internal Medicine, Tampere University Hospital, P.O.Box 2000, 33521 Tampere, Finland
AMS 2000 subject classifications: Primary 62H12; secondary 62G05, 62G20.
Keywords and phrases: asymptotic normality, Hettmansperger–Randles estimate, multivariate location, spatial sign, spatial sign test, transformation retransformation.

a spherical symmetry. Chakraborty et al. [4] proposed and investigated an affine equivariant modification of the spatial median constructed using an adaptive transformation and retransformation (TR) procedure. An affine invariant modification of the spatial sign test was also proposed. Randles [19] used Tyler's transformation [24] to construct an affine invariant modification of the spatial sign test. Later Hettmansperger and Randles [19] proposed an equivariant modification of the spatial median, again based on Tyler's transformation; this estimate is known as the Hettmansperger–Randles (HR) estimate.

In this paper we review and collect the results on the limiting behavior of the regular spatial median and its affine equivariant modification, the transformation retransformation spatial median. In Section 2 some auxiliary results and tools for asymptotic studies are given. Asymptotic theory for the regular spatial median is reviewed in Section 3. Estimation of the limiting covariance matrix of the spatial median is discussed in Section 4. Section 5 considers the transformation retransformation spatial median. The paper ends with some discussion on the algorithms for the computation of the spatial median in Section 6 and two examples in Section 7. Many of the results can be collected from Arcones [1], Bai et al. [2], Brown [3], Chakraborty et al.[4], Chaudhuri [5], Möttönen et al. [14] and Rao [20]. See also Nevalainen et al. [16] for the spatial median in the case of cluster correlated data. For the proofs in this paper it is crucial that the dimension $p > 1$. For the properties of the univariate median, see Section 2.3 in Serfling [22], for example.

2. Auxiliary results

Let $\mathbf{y} \neq \mathbf{0}$ and $\boldsymbol{\mu}$ be any p-vectors, $p > 1$. Write also $r = |\mathbf{y}|$ and $\mathbf{u} = |\mathbf{y}|^{-1}\mathbf{y}$.

Then accuracies of different (constant, linear and quadratic) approximations of function $\boldsymbol{\mu} \to |\mathbf{y} - \boldsymbol{\mu}|$ around the origin are given by

(A1) $||\mathbf{y} - \boldsymbol{\mu}| - |\mathbf{y}|| \leq |\boldsymbol{\mu}|$,
(A2) $||\mathbf{y} - \boldsymbol{\mu}| - |\mathbf{y}| + \mathbf{u}'\boldsymbol{\mu}| \leq 2r^{-1}|\boldsymbol{\mu}|^2$ and
(A3) $\left||\mathbf{y} - \boldsymbol{\mu}| - |\mathbf{y}| + \mathbf{u}'\boldsymbol{\mu} - \boldsymbol{\mu}'(2r)^{-1}[\mathbf{I}_p - \mathbf{u}\mathbf{u}']\boldsymbol{\mu}\right| \leq C_1 r^{-1-\delta}|\boldsymbol{\mu}|^{2+\delta}$ for all $0<\delta<1$,

where C_1 does not depend on $\mathbf{y}$ or $\boldsymbol{\mu}$.

Similarly, the accuracies of constant and linear approximations of unit vector $|\mathbf{y} - \boldsymbol{\mu}|^{-1}(\mathbf{y} - \boldsymbol{\mu})$ around the origin are given by

(B1) $\left|\frac{\mathbf{y}-\boldsymbol{\mu}}{|\mathbf{y}-\boldsymbol{\mu}|} - \frac{\mathbf{y}}{|\mathbf{y}|}\right| \leq 2r^{-1}|\boldsymbol{\mu}|$ and
(B2) $\left|\frac{\mathbf{y}-\boldsymbol{\mu}}{|\mathbf{y}-\boldsymbol{\mu}|} - \frac{\mathbf{y}}{|\mathbf{y}|} - \frac{1}{r}[\mathbf{I}_p - \mathbf{u}\mathbf{u}']\boldsymbol{\mu}\right| \leq C_2 r^{-1-\delta}|\boldsymbol{\mu}|^{1+\delta}$ for all $0 < \delta < 1$,

where C_2 does not depend on $\mathbf{y}$ or $\boldsymbol{\mu}$.

For these and similar results, see Arcones [1] and Bai et al. [2].

Lemma 1. *Assume that the density function $f(\mathbf{y})$ of the p-variate continuous random vector $\mathbf{y}$ is bounded. If $p > 1$ then $E\{|\mathbf{y}|^{-\alpha}\}$ exists for all $0 \leq \alpha < 2$.*

The following key result for convex processes is Lemma 4.2 in Davis et al. [7] and Theorem 1 in Arcones [1].

Theorem 1. *Let $G_n(\boldsymbol{\mu})$, $\boldsymbol{\mu} \in \mathbb{R}^p$, be a sequence of convex stochastic processes, and let $G(\boldsymbol{\mu})$ be a convex (limit) process in the sense that the finite dimensional distributions of $G_n(\boldsymbol{\mu})$ converge to those of $G(\boldsymbol{\mu})$. Let $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2, \ldots$ be random variables* *such that*

$$G(\hat{\boldsymbol{\mu}}) = \inf_{\boldsymbol{\mu}} G(\boldsymbol{\mu}) \quad \textit{and} \quad G_n(\hat{\boldsymbol{\mu}}_n) = \inf_{\boldsymbol{\mu}} G_n(\boldsymbol{\mu}), \quad n = 1, 2, \ldots$$

Then $\hat{\boldsymbol{\mu}}_n \to_d \hat{\boldsymbol{\mu}}$.

3. Spatial median

Let $\mathbf{y}$ be a p-variate random vector with cdf F, $p > 1$. The **spatial median** of F minimizes the objective function

$$D(\boldsymbol{\mu}) = E\{|\mathbf{y} - \boldsymbol{\mu}| - |\mathbf{y}|\}.$$

Note that no moment assumptions are needed in the definition as $||\mathbf{y}-\boldsymbol{\mu}|-|\mathbf{y}|| \leq |\boldsymbol{\mu}|$ but for the asymptotic theory we assume that

(C1) The p-variate density function f of $\mathbf{y}$ is continuous and bounded.
(C2) The spatial median of the distribution of $\mathbf{y}$ is zero and unique.

We next define vector and matrix valued functions

$$\mathbf{U}(\mathbf{y}) = \frac{\mathbf{y}}{|\mathbf{y}|}, \quad \mathbf{A}(\mathbf{y}) = \frac{1}{|\mathbf{y}|}\left[\mathbf{I}_p - \frac{\mathbf{y}\mathbf{y}'}{|\mathbf{y}|^2}\right], \quad \text{and} \quad \mathbf{B}(\mathbf{y}) = \frac{\mathbf{y}\mathbf{y}'}{|\mathbf{y}|^2}$$

for $\mathbf{y} \neq \mathbf{0}$ and, by convention, $\mathbf{U}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{A}(\mathbf{0}) = \mathbf{B}(\mathbf{0}) = \mathbf{0}$. We write also $\mathbf{A} = E\{\mathbf{A}(\mathbf{y})\}$ and $\mathbf{B} = E\{\mathbf{B}(\mathbf{y})\}$.

The expectation defining $\mathbf{B}$ clearly exists and is bounded ($|\mathbf{B}(\mathbf{y})|^2 = tr(\mathbf{B}(\mathbf{y})'\mathbf{B}(\mathbf{y}))$ $= 1$). Our assumption implies that $E(|\mathbf{y}|^{-1}) < \infty$ and therefore also $\mathbf{A}$ exists and is bounded. Auxiliary result (A3) in Section 2 then implies

Lemma 2. *Under assumptions (C1) and (C2),* $D(\boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + o(|\boldsymbol{\mu}|^2)$.

See also Lemma 19 in Arcones [1].

Let $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)'$ be a random sample from a p-variate distribution F. Write

$$D_n(\boldsymbol{\mu}) = \text{ave}\{|\mathbf{y}_i - \boldsymbol{\mu}| - |\mathbf{y}_i|\}.$$

The function $D_n(\boldsymbol{\mu})$ as well as $D(\boldsymbol{\mu})$ are convex and bounded. Boundedness follows from (A1). The *sample spatial median* $\hat{\boldsymbol{\mu}}$ is defined as

$$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(\mathbf{Y}) = \arg\min \; D_n(\boldsymbol{\mu}).$$

The estimate $\hat{\boldsymbol{\mu}}$ is unique if the observations do not fall on a line. Under assumption (C1) $\hat{\boldsymbol{\mu}}$ is unique with probability one. As $D(\boldsymbol{\mu})$ is the limiting process of $D_n(\boldsymbol{\mu})$, Theorem 1 implies that $\hat{\boldsymbol{\mu}} \to_P \mathbf{0}$.

The statistic $\mathbf{T}_n = \mathbf{T}(\mathbf{Y}) = \text{ave}\{\mathbf{U}(\mathbf{y}_i)\}$ is the *spatial sign test statistic* for testing the null hypothesis that the spatial median is zero. As $\boldsymbol{\mu}$ is assumed to be a zero vector, the multivariate central limit theorem implies that

Lemma 3. $\sqrt{n}\mathbf{T}_n \to_d N_p(\mathbf{0}, \mathbf{B})$.

The approximation (A3) in Section 1 implies that

$$\left|\sum_{i=1}^{n}\{|\mathbf{y}_i - n^{-1/2}\boldsymbol{\mu}| - |\mathbf{y}_i|\} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{y}_i'}{|\mathbf{y}_i|}\boldsymbol{\mu} - \boldsymbol{\mu}'\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2|\mathbf{y}_i|}\left[\mathbf{I}_p - \frac{\mathbf{y}_i\mathbf{y}_i'}{|\mathbf{y}_i|^2}\right]\right]\boldsymbol{\mu}\right|$$

$$\leq \frac{C_1}{n^{(2+\delta)/2}} \sum_{i=1}^{n} \frac{|\boldsymbol{\mu}|^{2+\delta}}{r_i^{1+\delta}} \to_P 0, \quad \text{for all } \boldsymbol{\mu},$$

and we get

Lemma 4. *Under assumptions (C1) and (C2),*

$$nD_n(n^{-1/2}\boldsymbol{\mu}) - \left(\sqrt{n}\mathbf{T}_n - \frac{1}{2}\mathbf{A}\boldsymbol{\mu}\right)' \boldsymbol{\mu} \ \to_P 0.$$

Now apply Theorem 1 with $G_n(\boldsymbol{\mu}) = nD_n(n^{-1/2}\boldsymbol{\mu})$ and $G(\boldsymbol{\mu}) = \left(\mathbf{z} - \frac{1}{2}\mathbf{A}\boldsymbol{\mu}\right)' \boldsymbol{\mu}$ where $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{B})$. We then obtain ($\mathbf{A}$ is positive definite)

Theorem 2. *Under assumptions (C1) and (C2),* $\sqrt{n}\hat{\boldsymbol{\mu}} \ \to_d \ N_p(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$.

It is well known that, if $E(\mathbf{y}_i) = \mathbf{0}$ and the second moments exist, also $\sqrt{n}\bar{\mathbf{y}} \to_d N_p(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{y}_i$. The asymptotic relative efficiency of the spatial median with respect to the sample mean is then given by $\det(\boldsymbol{\Sigma})/\det\left(\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right)$. The spatial median has good efficiency properties even in the multivariate normal model. Möttönen et al [15] for example calculated the asymptotic relative efficiencies $e(p, \nu)$ of the multivariate spatial median with respect to the mean vector in the p-variate $t_{\nu,p}$ distribution case ($t_{\infty,p}$ is the p-variate normal distribution). In the 3-variate and 10-variate cases, for example, the asymptotic relative efficiencies are

$$\begin{array}{lll} e(3,3) = 2.162, & e(3,10) = 1.009, & e(3,\infty) = 0.849, \\ e(10,3) = 2.422, & e(10,10) = 1.131, & e(10,\infty) = 0.951. \end{array}$$

4. Estimation of the covariance matrix of the spatial median

For a practical use of the normal approximation of the distribution of $\hat{\boldsymbol{\mu}}$ one naturally needs an estimate for the asymptotic covariance matrix $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$. We estimate $\mathbf{A}$ and $\mathbf{B}$ separately. Recall that we assume that the true value $\boldsymbol{\mu} = \mathbf{0}$. Write, as before,

$$\mathbf{A}(\mathbf{y}) = \frac{1}{|\mathbf{y}|}\left(\mathbf{I}_p - \frac{\mathbf{y}\mathbf{y}'}{|\mathbf{y}|^2}\right) \quad \text{and} \quad \mathbf{B}(\mathbf{y}) = \frac{\mathbf{y}\mathbf{y}'}{|\mathbf{y}|^2}.$$

Then write $\hat{\mathbf{A}} = \mathbf{A}(\mathbf{Y}) = \text{ave}\{\mathbf{A}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})\}$ and $\hat{\mathbf{B}} = \mathbf{B}(\mathbf{Y}) = \text{ave}\{\mathbf{B}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})\}$. We will show that, under our assumptions, $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ converge in probability to the population values $\mathbf{A} = E\{\mathbf{A}(\mathbf{y}_i)\}$ and $\mathbf{B} = E\{\mathbf{B}(\mathbf{y}_i)\}$, respectively:

Theorem 3. *Under assumptions (C1) and (C2),* $\hat{\mathbf{A}} \to_P \mathbf{A}$ *and* $\hat{\mathbf{B}} \to_P \mathbf{B}$.

Proof We thus assume that the true spatial median $\boldsymbol{\mu} = \mathbf{0}$. By Theorem 2, $\sqrt{n}\hat{\boldsymbol{\mu}} = O_p(1)$. Write $\tilde{\mathbf{A}} = \text{ave}\{\mathbf{A}(\mathbf{y}_i)\}$ and $\tilde{\mathbf{B}} = \text{ave}\{\mathbf{B}(\mathbf{y}_i)\}$. Then by the law of large numbers $\tilde{\mathbf{A}} \to_P \mathbf{A}$ and $\tilde{\mathbf{B}} \to_P \mathbf{B}$. Our auxiliary result (B1) implies that

$$\left| \frac{(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'}{|\mathbf{y} - \boldsymbol{\mu}|^2} - \frac{\mathbf{y}\mathbf{y}'}{|\mathbf{y}|^2} \right| \leq 4\frac{|\boldsymbol{\mu}|}{|\mathbf{y}|}, \quad \forall \ \mathbf{y} \neq 0, \boldsymbol{\mu},$$

and therefore by Slutsky's theorem $|\hat{\mathbf{B}} - \tilde{\mathbf{B}}| \leq \frac{1}{n}\sum_{i=1}^{n}\{4|\hat{\boldsymbol{\mu}}|/|\mathbf{y}_i|\} \to_P 0$. As $\tilde{\mathbf{B}} \to_P \mathbf{B}$, also $\hat{\mathbf{B}} \to_P \mathbf{B}$.

We now prove that $\hat{\mathbf{A}} \to_P \mathbf{A}$. We play with three positive constants, "large" δ_1, "small" δ_2 and "small" δ_3. For a moment, we assume that $|\hat{\boldsymbol{\mu}}| < \delta_1/\sqrt{n}$. (This is true with a probability that can be made close to one with large δ_1.) Next we write $I_{1i} = I\left\{ |\mathbf{y}_i - \hat{\boldsymbol{\mu}}| < \frac{\delta_2}{\sqrt{n}} \right\}$, $I_{2i} = I\left\{ \frac{\delta_2}{\sqrt{n}} \leq |\mathbf{y}_i - \hat{\boldsymbol{\mu}}| < \delta_3 \right\}$ and $I_{3i} = I\left\{ |\mathbf{y}_i - \hat{\boldsymbol{\mu}}| \geq \delta_3 \right\}$. Then

$$
\begin{aligned}
\tilde{\mathbf{A}} - \hat{\mathbf{A}} &= \frac{1}{n}\sum_{i=1}^{n} (\mathbf{A}(\mathbf{y}_i) - \mathbf{A}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})) \\
&= \frac{1}{n}\sum_{i=1}^{n} (I_{1i} \cdot [\mathbf{A}(\mathbf{y}_i) - \mathbf{A}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})]) + \frac{1}{n}\sum_{i=1}^{n} (I_{2i} \cdot [\mathbf{A}(\mathbf{y}_i) - \mathbf{A}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})]) \\
&+ \frac{1}{n}\sum_{i=1}^{n} (I_{3i} \cdot [\mathbf{A}(\mathbf{y}_i) - \mathbf{A}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})]) .
\end{aligned}
$$

The first average is zero with probability

$$
P(I_{11} = \ldots = I_{1n} = 0) \geq \left(1 - \frac{\delta_2^p c_p M}{n^{p/2}}\right)^n \geq \left(1 - \frac{\delta_2^2 c_p M}{n}\right)^n \to e^{-c_p M \delta_2^2},
$$

where $M = \sup_{\mathbf{y}} f(\mathbf{y}) < \infty$ and c_p is the volume of the p-variate unit ball. (The first average is thus zero with a probability that can be made close to one with small choices of $\delta_2 > 0$.) For the second average, one gets

$$
\frac{1}{n}\sum_{i=1}^{n} |I_{2i} \cdot [\mathbf{A}(\mathbf{y}_i) - \mathbf{A}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})]| \quad \leq \quad \frac{1}{n}\sum_{i=1}^{n} \frac{6 I_{2i} |\hat{\boldsymbol{\mu}}|}{|\mathbf{y}_i - \hat{\boldsymbol{\mu}}||\mathbf{y}_i|} \leq \frac{1}{n}\sum_{i=1}^{n} \frac{6 I_{2i} \delta_1}{\delta_2 |\mathbf{y}_i|}
$$

which converges to a constant which can be made as close to zero as one wishes with small $\delta_3 > 0$. Finally, also the third average

$$
\frac{1}{n}\sum_{i=1}^{n} |I_{3i} \cdot [\mathbf{A}(\mathbf{y}_i) - \mathbf{A}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})]| \quad \leq \quad \frac{1}{n}\sum_{i=1}^{n} \frac{6 I_{3i} |\hat{\boldsymbol{\mu}}|}{|\mathbf{y}_i - \hat{\boldsymbol{\mu}}||\mathbf{y}_i|} \leq \frac{1}{n\sqrt{n}}\sum_{i=1}^{n} \frac{6 I_{3i} \delta_1}{\delta_3 |\mathbf{y}_i|}
$$

converges to zero in probability for all choices of δ_1 and δ_3, and the proof follows.□

Theorems 2 and 3 thus suggest that the distribution of $\hat{\boldsymbol{\mu}}$ can be approximated by $N_p\left(\boldsymbol{\mu}, \frac{1}{n}\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}\right)$. Approximate 95 % confidence ellipsoids for $\boldsymbol{\mu}$ are given by $\left\{\boldsymbol{\mu} \;:\; n(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})'\hat{\mathbf{A}}\hat{\mathbf{B}}^{-1}\hat{\mathbf{A}}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \leq \chi^2_{p,.95}\right\}$, where $\chi^2_{p,.95}$ is the 95 % quantile of a chi square distribution with p degrees of freedom. Also, by Slutsky's theorem, under the null hypothesis $H_0 : \;\; \boldsymbol{\mu} = \mathbf{0}$ the squared version of the test statistic $Q^2 = n\mathbf{T}_n'\hat{\mathbf{B}}^{-1}\mathbf{T}_n \;\to_d\; \chi^2_p$.

5. Transformation retransformation spatial median

Shifting the data cloud, naturally shifts the spatial median by the same constant, that is, $\hat{\boldsymbol{\mu}}(\mathbf{1}_n\mathbf{a}' + \mathbf{Y}) = \mathbf{a} + \hat{\boldsymbol{\mu}}(\mathbf{Y})$, It is also easy to see that rotating the data cloud also rotates the spatial median correspondingly, that is, $\hat{\boldsymbol{\mu}}(\mathbf{Y}\mathbf{O}') = \mathbf{O}\hat{\boldsymbol{\mu}}(\mathbf{Y})$, for all orthogonal $p \times p$ matrices $\mathbf{O}$. Unfortunately, the estimate is not equivariant under heterogeneous rescaling of the components, and therefore not fully affine equivariant.

A fully affine equivariant version of the spatial median can be found using the so called transformation retransformation estimation technique. First, a positive definite $p \times p$ scatter matrix $\mathbf{S} = \mathbf{S}(\mathbf{Y})$ is a matrix valued sample statistic which is affine equivariant in the sense

$$\mathbf{S}(\mathbf{1}_n \mathbf{a}' + \mathbf{Y}\mathbf{B}') = \mathbf{B}\mathbf{S}(\mathbf{Y})\mathbf{B}'$$

for all p-vectors $\mathbf{a}$ and all nonsingular $p \times p$ matrices $\mathbf{B}$. Let $\mathbf{S}^{-1/2}$ be any matrix which satisfies $\mathbf{S}^{-1/2}\mathbf{S}(\mathbf{S}^{-1/2})' = \mathbf{I}_p$. The procedure is then as follows.

1. Take any scatter matrix $\mathbf{S} = \mathbf{S}(\mathbf{Y})$.
2. Transform the data matrix: $\mathbf{Y}(\mathbf{S}^{-1/2})'$.
3. Find the spatial median for the standardized data matrix $\hat{\boldsymbol{\mu}}(\mathbf{Y}(\mathbf{S}^{-1/2})')$.
4. Retransform the estimate: $\tilde{\boldsymbol{\mu}}(\mathbf{Y}) = \mathbf{S}^{1/2}\hat{\boldsymbol{\mu}}(\mathbf{Y}(\mathbf{S}^{-1/2})')$.

This median $\tilde{\boldsymbol{\mu}}(\mathbf{Y})$ utilizing "data driven" transformation $\mathbf{S}^{-1/2}$ is known as the *transformation retransformation (TR) spatial median.* (See Chakraborty et al. [4] for other type of data driven transformations.) Then the affine equivariance follows:

Theorem 4. *Let* $\mathbf{S} = \mathbf{S}(\mathbf{Y})$ *be any scatter matrix. Then the transformation retransformation spatial median* $\tilde{\boldsymbol{\mu}}(\mathbf{Y}) = \mathbf{S}^{1/2}\hat{\boldsymbol{\mu}}(\mathbf{Y}(\mathbf{S}^{-1/2})')$ *is affine equivariant, that is,* $\tilde{\boldsymbol{\mu}}(\mathbf{1}_n\mathbf{a}' + \mathbf{Y}\mathbf{B}') = \mathbf{a} + \mathbf{B}\tilde{\boldsymbol{\mu}}(\mathbf{Y})$.

The proof easily follows from the facts that the regular spatial median is shift and orthogonally equivariant and that $(\mathbf{S}(\mathbf{1}_n\mathbf{a}' + \mathbf{Y}\mathbf{B}'))^{-1/2} = \mathbf{O}(\mathbf{S}(\mathbf{Y}))^{-1/2}$ for some orthogonal matrix $\mathbf{O}$.

In the following we assume (without loss of generality) that the population value of $\mathbf{S}$ is $\mathbf{I}_p$, and that $\mathbf{S} = \mathbf{S}(\mathbf{Y})$ is a root-n consistent estimate of $\mathbf{I}_p$. We write $\boldsymbol{\Delta} = \sqrt{n}(\mathbf{S}^{-1/2} - \mathbf{I}) = O_p(1)$ and $\mathbf{Y}^* = \mathbf{Y}(\mathbf{S}^{-1/2})'$. Then we have the following result for the test statistic.

Lemma 5. *Let* $\mathbf{Y}$ *be a random sample from a symmetric distribution satisfying (C1) and (C2). (By a symmetry we mean that* $-\mathbf{y}_i$ *and* $\mathbf{y}_i$ *have the same distribution.) Assume also that scatter matrix* $\mathbf{S} = \mathbf{S}(\mathbf{Y})$ *satisfies* $\sqrt{n}(\mathbf{S} - \mathbf{I}_p) = O_p(1)$*. Then* $\sqrt{n}(\mathbf{T}(\mathbf{Y}^*) - \mathbf{T}(\mathbf{Y})) \to_P \mathbf{0}$*.*

Proof Our assumptions imply that also $\boldsymbol{\Delta} = \sqrt{n}(\mathbf{S}^{-1/2} - \mathbf{I}_p) = O_p(1)$. Thus $\mathbf{S}^{-1/2} = \mathbf{I}_p + n^{-1/2}\boldsymbol{\Delta}$ where $\boldsymbol{\Delta}$ is bounded in probability. Using auxiliary result (B2) in Section 2 we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{U}(\mathbf{S}^{-1/2}\mathbf{y}_i) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{U}_i = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\Delta} - \mathbf{U}_i'\boldsymbol{\Delta}\mathbf{U}_i)\mathbf{U}_i + o_P(1)$$

where $\mathbf{U}_i = \mathbf{U}(\mathbf{y}_i)$, $i = 1, \ldots, n$. For $|\boldsymbol{\Delta}| < M$, the second term in the expansion converges uniformly in probability to zero due to its linearity with respect to the elements of $\boldsymbol{\Delta}$ and due to the symmetry of the distribution of $\mathbf{U}_i$. ($E(\mathbf{U}_i) = E(\mathbf{U}_i'\boldsymbol{\Delta}\mathbf{U}_i\mathbf{U}_i) = \mathbf{0}$) . Therefore $n^{-1/2}\sum_{i=1}^{n}\mathbf{U}(\mathbf{S}^{-1/2}\mathbf{y}_i) - n^{-1/2}\sum_{i=1}^{n}\mathbf{U}_i \to_P 0$ and the proof follows. $\square$

We also have to show that $\mathbf{A}(\mathbf{Y}^*)$ and $\mathbf{A}(\mathbf{Y})$ both converge to $\mathbf{A}$, and similarly with $\mathbf{B}(\mathbf{Y}^*)$ and $\mathbf{B}(\mathbf{Y})$:

Lemma 6. *Let* $\mathbf{Y}$ *be a random sample from a distribution satisfying (C1) and (C2). Assume also that scatter matrix* $\mathbf{S} = \mathbf{S}(\mathbf{Y})$ *satisfies* $\sqrt{n}(\mathbf{S} - \mathbf{I}_p) = O_p(1)$*.* $\mathbf{A}(\mathbf{Y}^*) - \mathbf{A}(\mathbf{Y}) \to_P \mathbf{0}$ *and* $\mathbf{B}(\mathbf{Y}^*) - \mathbf{B}(\mathbf{Y}) \to_P \mathbf{0}$*.*

Proof Again $\mathbf{S}^{-1/2} = \mathbf{I}_p + n^{-1/2}\boldsymbol{\Delta}$ where $\boldsymbol{\Delta} = O_p(1)$. Suppose that $\boldsymbol{\Delta} \le M$. ($P(\boldsymbol{\Delta} \le M) \to 1$ as $M \to \infty$.) Write $\mathbf{y}_i^* = (\mathbf{I}_p - n^{-1/2}\boldsymbol{\Delta})\mathbf{y}_i$. Then

$$\left| \frac{\mathbf{y}_i^*\mathbf{y}_i^{*\prime}}{|\mathbf{y}_i^*|^2} - \frac{\mathbf{y}_i\mathbf{y}_i^\prime}{|\mathbf{y}_i|^2} \right| \le \frac{1}{\sqrt{n}}|\boldsymbol{\Delta}| \quad \text{and} \quad \left| \frac{1}{|\mathbf{y}_i^*|} - \frac{1}{|\mathbf{y}_i|} \right| \le \frac{|\mathbf{I}_p - (\mathbf{I}_p - n^{-1/2}\boldsymbol{\Delta})^{-1}|}{|\mathbf{y}_i|}.$$

The first inequality gives $|\mathbf{B}(\mathbf{Y}^*) - \mathbf{B}(\mathbf{Y})| \le \frac{1}{\sqrt{n}}|\boldsymbol{\Delta}| \to 0$. The two inequalities together imply that

$$\left| \frac{1}{|\mathbf{y}_i|}\left(\mathbf{I}_p - \frac{\mathbf{y}_i\mathbf{y}_i^\prime}{|\mathbf{y}_i|^2}\right) - \frac{1}{|\mathbf{y}_i^*|}\left(\mathbf{I}_p - \frac{\mathbf{y}_i^*\mathbf{y}_i^{*\prime}}{|\mathbf{y}_i^*|^2}\right) \right| \quad \le \quad \frac{1}{|\mathbf{y}_i|}\left(\frac{3M}{\sqrt{n}} + o(n^{-1/2})\right).$$

Then $|\mathbf{A}(\mathbf{Y}^*) - \mathbf{A}(\mathbf{Y})| \le \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{|\mathbf{y}_i|}\left(\frac{3M}{\sqrt{n}} + o(n^{-1/2})\right)\right] \to_P 0$. □

Using Lemmas 5 and 6 and the auxiliary results in Section 2 we then get

Theorem 5. *Let* $\mathbf{Y}$ *be a random sample from a symmetric distribution satisfying (C1) and (C2). Assume also that scatter matrix* $\mathbf{S} = \mathbf{S}(\mathbf{Y})$ *satisfies* $\sqrt{n}(\mathbf{S} - \mathbf{I}_p) = O_p(1)$. *Then* $\sqrt{n}\tilde{\boldsymbol{\mu}}(\mathbf{Y})$ *and* $\sqrt{n}\hat{\boldsymbol{\mu}}(\mathbf{Y})$ *have the same limiting distribution.*

Proof Write again $\mathbf{S}^{-1/2} = \mathbf{I}_p + n^{-1/2}\boldsymbol{\Delta}$, and $\mathbf{y}_i^* = (\mathbf{I}_p - n^{-1/2}\boldsymbol{\Delta})\mathbf{y}_i$, $i = 1, \ldots, n$, and $\mathbf{Y}^* = (\mathbf{y}_1^*, \ldots, \mathbf{y}_n^*)^\prime$. Then our auxiliary results imply that that

$$\left| \sum_{i=1}^n \{|\mathbf{y}_i^* - n^{-1/2}\boldsymbol{\mu}| - |\mathbf{y}_i^*|\} - \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\mathbf{y}_i^{*\prime}}{|\mathbf{y}_i^*|}\boldsymbol{\mu} - \boldsymbol{\mu}^\prime \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{2|\mathbf{y}_i^*|}\left[\mathbf{I}_p - \frac{\mathbf{y}_i^*\mathbf{y}_i^{*\prime}}{|\mathbf{y}_i^*|^2}\right]\right]\boldsymbol{\mu} \right|$$

$$\le \frac{C_1}{n^{(2+\delta)/2}} \sum_{i=1}^n \frac{|\boldsymbol{\mu}|^{2+\delta}|(\mathbf{I}_p - n^{-1/2}\boldsymbol{\Delta})^{-1}|^{1+\delta}}{|\mathbf{y}_i^*|^{1+\delta}} \to_P 0$$

Thus Lemmas 5 and 6 together with Theorem 1 imply that $\sqrt{n}\hat{\boldsymbol{\mu}}(\mathbf{Y}^*)$ and $\sqrt{n}\hat{\boldsymbol{\mu}}(\mathbf{Y})$ have the same limiting distribution. As $\sqrt{n}\tilde{\boldsymbol{\mu}}(\mathbf{Y}) = \mathbf{S}^{1/2}\sqrt{n}\hat{\boldsymbol{\mu}}(\mathbf{Y}^*)$, the result follows from Slutsky's theorem. □

Based on the results above, the distribution of $\tilde{\boldsymbol{\mu}}$ can in the symmetric case be approximated by $N_p\left(\boldsymbol{\mu}, \widehat{Cov}(\tilde{\boldsymbol{\mu}})\right)$, where $\widehat{Cov}(\tilde{\boldsymbol{\mu}}) = \frac{1}{n}\mathbf{S}^{1/2}\hat{\mathbf{A}}_S^{-1}\hat{\mathbf{B}}_S\hat{\mathbf{A}}_S^{-1}(\mathbf{S}^{1/2})^\prime$ with $\hat{\mathbf{A}}_S = \mathrm{ave}\left\{\frac{1}{|\mathbf{e}_i|^2}\left(\mathbf{I}_p - \frac{\mathbf{e}_i\mathbf{e}_i^\prime}{|\mathbf{e}_i|^2}\right)\right\}$ and $\hat{\mathbf{B}}_S = \mathrm{ave}\left\{\frac{\mathbf{e}_i\mathbf{e}_i^\prime}{|\mathbf{e}_i|^2}\right\}$ calculated from the standardized residuals $\mathbf{e}_i = \mathbf{S}^{-1/2}(\mathbf{y}_i - \tilde{\boldsymbol{\mu}})$, $i = 1, \ldots, n$.

The stochastic convergence and the limiting normality of the spatial median did not require any moment assumptions. Therefore, for the transformation, a scatter matrix with weak assumptions should be used as well. It is an appealing idea to link also the spatial median with the Tyler's transformation. This was proposed by Hettmansperger and Randles [11]:

Definition 1. *Let* $\hat{\boldsymbol{\mu}}$ *be a p-vector and* $\mathbf{S} > 0$ *a symmetric* $p \times p$ *matrix, and define* $\hat{\mathbf{e}}_i = \mathbf{S}^{-1/2}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})$, $i = 1, \ldots, n$. *The Hettmansperger–Randles (HR) estimate of location and scatter are the values of* $\hat{\boldsymbol{\mu}}$ *and* $\mathbf{S}$ *which simultaneously satisfy*

$$\mathrm{ave}\{\mathbf{U}(\hat{\mathbf{e}}_i)\} = \mathbf{0} \quad \textit{and} \quad p\ \mathrm{ave}\{\mathbf{U}(\hat{\mathbf{e}}_i)\mathbf{U}(\hat{\mathbf{e}}_i)^\prime\} = \mathbf{I}_p.$$

Note that the HR estimate is not a TR estimate as the location vector and scatter matrix are in fact estimated simultaneously. This pair of estimates was first mentioned in Tyler [24]. Hettmansperger and Randles [11] developed the properties of these estimates. They showed that the HR estimate has a bounded influence function and a positive breakdown point. The distribution of the HR location estimate can be approximated by

$$N_p\left(\mathbf{0}, \frac{1}{np}\mathbf{S}^{1/2}\hat{\mathbf{A}}_S^{-2}\mathbf{S}^{1/2}\right)$$

where $\hat{\mathbf{A}}_S = \text{ave}(\mathbf{A}(\mathbf{S}^{-1/2}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})))$ and $\mathbf{S}$ is Tyler's scatter matrix.

6. Computation of the spatial median

The spatial median can often be computed using the following two steps:

Step 1: $\mathbf{e}_i \leftarrow \mathbf{y}_i - \boldsymbol{\mu}, \quad i = 1, \ldots, n$
Step 2: $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + \left(\sum_{i=1}^n |\mathbf{e}_i|^{-1}\right)^{-1} \sum_{i=1}^n \mathbf{U}(\mathbf{e}_i)$

provided an initial estimate for $\boldsymbol{\mu}$.

The above algorithm may fail in case of ties or when an estimate falls on a data point. Assume then that the distinct data points are $\mathbf{y}_1, \ldots, \mathbf{y}_m$ with multiplicities $w_1, \ldots, w_m$ $(w_1 + \ldots + w_m = n)$. The algorithm by Vardi and Zhang [25] then uses the steps:

Step 1: $\mathbf{e}_i \leftarrow \mathbf{y}_i - \boldsymbol{\mu}, \quad i = 1, \ldots, m$
Step 2: $c \leftarrow (\sum_{\mathbf{e}_i = \mathbf{0}} w_i)/|\sum_{\mathbf{e}_i \neq \mathbf{0}} w_i \mathbf{U}(\mathbf{e}_i)|$
Step 3: $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + \max(0, 1 - c) \left(\sum_{\mathbf{e}_i \neq \mathbf{0}} w_i |\mathbf{e}_i|^{-1}\right)^{-1} \sum_{\mathbf{e}_i \neq \mathbf{0}} w_i \mathbf{U}(\mathbf{e}_i)$

Furthermore many other approaches can be used to solve this non-smooth optimization problem. For example Hössjer and Croux [12] suggest a steepest descent algorithm combined with stephalving and discuss also some other algorithms. We prefer however the above algorithm since it seems efficient and can be easily combined with the HR approach with the following steps:

Step 1: $\mathbf{e}_i \leftarrow \mathbf{S}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}), \quad i = 1, \ldots, n$
Step 2: $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + \left[\sum_{i=1}^n \{|\mathbf{e}_i|^{-1}\}\right]^{-1} \mathbf{S}^{1/2} \sum_{i=1}^n \{\mathbf{U}(\mathbf{e}_i)\}$
Step 3: $\mathbf{S} \leftarrow (p/n)\, \mathbf{S}^{1/2} \sum_{i=1}^n \{\mathbf{U}(\mathbf{e}_i)\mathbf{U}(\mathbf{e}_i)'\}\, \mathbf{S}^{1/2}.$

There are actually two ways to implement the algorithm. The first one is just to repeat these three steps 1, 2 and 3 until convergence. The second one is first (i) to repeat steps 1 and 2 until convergence, and then (ii) repeat steps 1 and 3 until convergence. Finally (i) and (ii) are repeated until convergence. The second version is sometimes considered faster and more stable, see Hettmansperger and Randles[11] and the references therein.

Both versions of the algorithm are easy to implement and the computation is fast even in high dimensions. Unfortunately, there is no proof for the convergence of the algorithms so far, although in practice they always seem to work. There is no proof for the existence or uniqueness of the HR estimate either. In practice, this is not a problem, however. One can start with any initial root-n consistent estimates, then repeat the above steps for location and scatter, and stop after k iterations. If, in the spherical case around the origin, the initial location and shape estimates, say $\hat{\boldsymbol{\mu}}$ and $\mathbf{S}$ are root-n consistent, that is,

$$\sqrt{n}\hat{\boldsymbol{\mu}} = O_P(1) \quad \text{and} \quad \sqrt{n}(\mathbf{S} - \mathbf{I}_p) = O_P(1)$$

and $tr(\mathbf{S}) = p$ then the k-step estimate using the single loop version of the above algorithm (obtained after k iterations) satisfies

$$\sqrt{n}\hat{\boldsymbol{\mu}}_k = \left(\frac{1}{p}\right)^k \sqrt{n}\hat{\boldsymbol{\mu}} + \left[1 - \left(\frac{1}{p}\right)^k\right] \frac{1}{E(r_i^{-1})} \frac{p}{p-1} \sqrt{n}\ \text{ave}\{\mathbf{u}_i\} + o_P(1)$$

and

$$\begin{aligned}\sqrt{n}(\mathbf{S}_k - \mathbf{I}_p) &= \left(\frac{2}{p+2}\right)^k \sqrt{n}(\mathbf{S} - \mathbf{I}_p) \\ &+ \left[1 - \left(\frac{2}{p+2}\right)^k\right] \frac{p+2}{p} \sqrt{n}\,(p \cdot \text{ave}\{\mathbf{u}_i\mathbf{u}_i'\} - \mathbf{I}_p) + o_P(1).\end{aligned}$$

Asymptotically, the k-step estimate behaves as a linear combination of the initial pair of estimates and Hettmansperger–Randles estimate. The larger k, the more similar is the distribution to that of the HR estimate. More work is needed, however, to carefully consider the properties of this k-step HR-estimate.

7. Examples

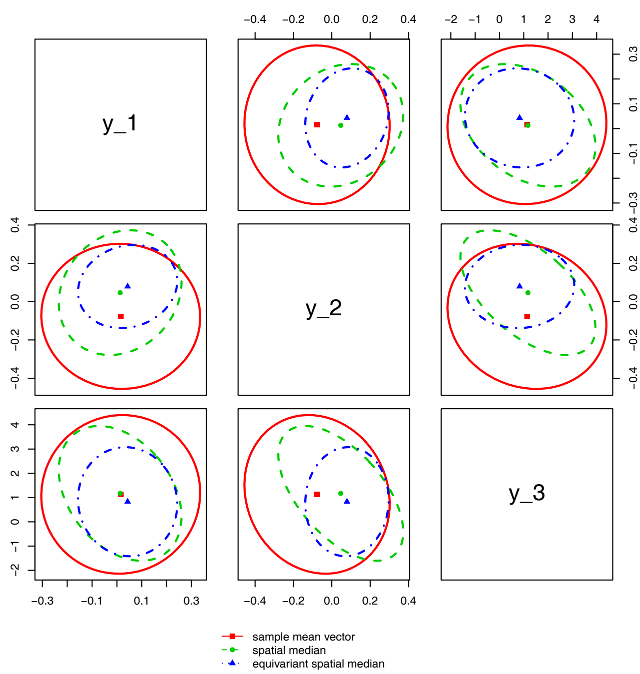

Fig 1. *The sample mean vector, the spatial median and the HR location estimate with corresponding bivariate 95% confidence ellipsoids for a simulated dataset from a non-spherical 3-variate t_3 distribution.*

In this section we compare the mean vector, the regular spatial median, and the HR location estimate for simulated and real datasets. First, the simulated data with

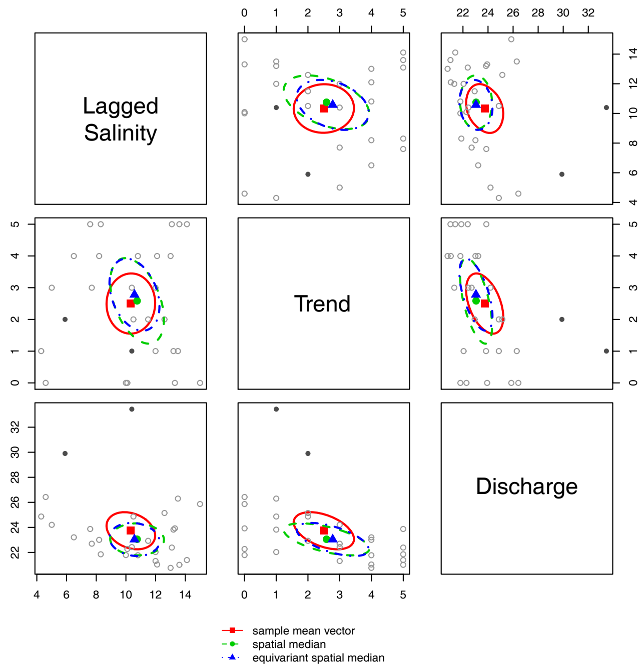

FIG 2. *Salinity data with the sample mean vector, the spatial median and the HR location estimate with corresponding bivariate 95% confidence ellipsoids. Two outliers are marked with a darker colour.*

sample size $n = 200$ was generated from a 3-variate spherical t distribution with 3 degrees of freedom. In the case of a spherical distribution, the regular spatial median and the affine equivariant HR location estimate are behaving in a very similar way. To illustrate the differences between these two estimates in a non-spherical case, the third component was multiplied by 10. The three location estimates with their bivariate 95% confidence ellipsoids are presented in Figure 1. The mean vector is less accurate due to the heavy tails of the distribution. For non-spherical data, the equivariant HR location estimate is more efficient than the spatial median as seen in the Figure. If the measurement units for the components are the same, however, as in the case of the repeated measures, and heterogeneous rescaling is not natural, then of course the spatial median may be preferable.

To illustrate the robustness properties of the three estimates we consider the three variables “Lagged Salinity”, “Trend” and “Discharge” in the Salinity dataset discussed in Rousseeuw and Leroy [21]. There are two clearly visible outliers among the 28 observations. As seen from Figure 2, the mean vector and the corresponding confidence ellipsoid are clearly affected by these outliers. The HR estimate seems a bit more accurate than the spatial median due to the different scales of the marginal variables. Estimation of the spatial median and HR estimate and their covariances is implemented in the R package MNM [18].

Acknowledgements. We thank the two referees for their valuable comments on the earlier version of the paper.

References

[1] Arcones, M. A. (1998). Asymptotic theory for M-estimators over a convex kernel. *Econometric Theory* **14** 387–422.

[2] Bai, Z. D., Chen, X. R., Miao, B. Q., and Rao, C. R. (1990). Asymptotic theory of least distances estimate in multivariate linear models. *Statistics* **21** 503–519.

[3] Brown, B. M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society, Series B* **45** 25–30.

[4] Chakraborty, B, Chaudhuri, P., and Oja, H. (1998). Operating transformation re-transformation on spatial median and angle test. *Statistica Sinica* **8** 767–784.

[5] Chaudhuri, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Annals of Statistics* **20** 897–916.

[6] Chaudhuri, P. and Sengupta, D. (1993). Sign tests in multidimension: Inference based on the geometry of data cloud. *Journal of the American Statistical Society* **88** 1363–1370.

[7] Davis, R. A., Knight, K., and Liu, J. (1992). M-estimation for autoregression with infinite variance. *Stochastic Processes and Their Applications* **40** 145–180.

[8] Gini, C. and Galvani, L. (1929). Di talune estensioni dei concetti di media ai caratteri qualitative. *Metron* **8**.

[9] Gower, J. S. (1974). The mediancentre. *Applied Statistics* **2** 466–470.

[10] Haldane, J. B. S. (1948). Note on the median of the multivariate distributions. *Biometrika* **35** 414–415.

[11] Hettmansperger, T. P. and Randles, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860.

[12] Hössjer, O. and Croux, C. (1995). Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter. *Journal of Nonparametric Statistics* **4** 293–308.

[13] Milasevic, P. and Ducharme, G. R. (1987). Uniqueness of the spatial median. *Annals of Statistics* **15** 1332–1333.

[14] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics* **5** 201–213.

[15] Möttönen, J., Oja, H., and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Annals of Statistics* **25** 542–552.

[16] Nevalainen, J., Larocque, D., and Oja, H. (2007). On the multivariate spatial median for clustered data. *Canadian Journal of Statistics* **35** 215-231.

[17] Niinimaa, A. and Oja, H. (1999). Multivariate median. In: *Encyclopedia of Statistical Sciences (Update Volume 3)*. Eds. by Kotz, S., Johnson, N. L. and Read, C. P., Wiley.

[18] Nordhausen, K. Möttönen, J., and Oja, H. (2009). MNM: Multivariate Nonparametric Methods. An Approach Based on Spatial Signs and Ranks. R package version 0.95-1.

[19] Randles, R. H. (2000). A simpler, affine equivariant multivariate, distribution-free sign test. *Journal of the American Statistical Association* **95** 1263–1268.

[20] Rao, C. R. (1988). Methodology based on the L_1-norm in statistical inference. *Sankhyā* Ser. A **50** 289–313.

[21] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* John Wiley & Sons, New York.

[22] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[23] Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Review* **58** 263–277.

[24] Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Annals of Statistics* **15** 234–251.

[25] Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *The Proceedings of the National Academy of Sciences USA (PNAS)* **97** 1423–1426.