A Mahalanobis Multivariate Quantile Function

Robert Serfling[1]

University of Texas at Dallas

July 2008
revised December 2008

[1]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: `serfling@utdallas.edu`. Website: `www.utdallas.edu/~serfling`. Support by NSF Grants DMS-0103698 and DMS-0805786 is gratefully acknowledged.

Abstract

The Mahalanobis distance is pervasive throughout multivariate statistical analysis. Here it receives a further role, in formulating a new affine equivariant and mathematically tractable multivariate quantile function with favorable properties. Besides having sample versions that are robust, computationally easy, and asymptotically normal, this "Mahalanobis quantile function" also provides two special benefits. Its associated "outlyingness" contours, unlike those of the "usual" Mahalanobis outlyingness function, are not restricted to be elliptical. And it provides a rigorous foundation for understanding the transformation-retransformation (TR) method used in practice for constructing affine equivariant versions of the sample spatial quantile function. Indeed, the Mahalanobis quantile function has a TR representation in terms of the spatial quantile function. This yields, for example, that the "TR sample spatial median" estimates not the population spatial median, but rather the population Mahalanobis median. This clarification actually strengthens, rather than weakens, the motivation for the TR approach. Two major tools, both of independent interest as well, are developed and applied: a general formulation of affine equivariance for multivariate quantile functions, and a notion of "weak covariance functional" that connects with the functionals used in the TR approach. Variations on the definition of "Mahalanobis" quantiles are also discussed.

AMS 2000 Subject Classification: Primary 62G99 Secondary 62H99.

Key words and phrases: Mahalanobis distance; multivariate analysis; quantile functions; covariance functionals; equivariance; nonparametric methods; robustness.

1 Introduction

For distributions F on $\mathbb{R}^d$, Mahalanobis (1936) introduced standardization of points $\boldsymbol{x}$ in $\mathbb{R}^d$ by

$$\|\boldsymbol{\Sigma}(F)^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}(F))\|, \tag{1}$$

with $\boldsymbol{\mu}(F)$ and $\boldsymbol{\Sigma}(F)$ location and covariance functionals of F, respectively, and $\|\cdot\|$ the usual Euclidean distance. The *Mahalanobis distance* (1) has become pervasive in multivariate statistical analysis and data mining.

A leading application of this distance is to measure *outlyingness* of a point $\boldsymbol{x}$ relative to a distribution F. In practice, robust choices of $\boldsymbol{\mu}(F)$ and $\boldsymbol{\Sigma}(F)$ are used. The approach has mathematical tractability and intuitive appeal. A shortcoming, however, is that the contours of equal outlyingness generated by (1) are necessarily *ellipsoidal*, regardless of whether F is elliptically symmetric. Here an alternative way to use the Mahalanobis distance, that allows the contours to follow the actual shape of F and yields certain other benefits, is introduced and explored. More fundamentally, we define a (new) *Mahalanobis quantile function* on $\mathbb{R}^d$, from which one can then pass to associated *depth, outlyingness, and centered rank functions.* (As a preliminary, these four functions and their interrelations are discussed in Section 2.1.)

The "center", i.e., the associated *median*, of the Mahalanobis quantile function that we formulate has already been considered (Isogai, 1985, Rao, 1988). It is the minimizer $\boldsymbol{\theta}$ of the *expected Mahalanobis distance*,

$$E\|\boldsymbol{\Sigma}(F)^{-1/2}(\boldsymbol{X} - \boldsymbol{\theta})\|. \tag{2}$$

Comparatively, the so-called *spatial median* minimizes the *expected Euclidean distance*,

$$E\|\boldsymbol{X} - \boldsymbol{\theta}\|, \tag{3}$$

and has a long literature (reviewed in Small, 1990). Although the Mahalanobis median is fully affine equivariant while the spatial median is only orthogonally equivariant, the latter has received some preference (e.g., Chakraborty, Chaudhuri, and Oja, 1998), primarily on the grounds that the coordinate system resulting from standardization as in (2) lacks a "simple and natural geometric interpretation". On the other hand, standardization is a fundamental step used throughout statistical analysis, motivated by the principle that deviations should be interpreted in a relative sense. Thus (2) indeed has the same supporting arguments that have made (1) a basic tool.

Furthermore, in practice the sample version used by the advocates of (3) is corrected to be fully affine equivariant by substituting a sample analogue of (2) for the sample analogue of (3). For a data set $\mathbb{X}_n = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ in $\mathbb{R}^d$, the *transformation-retransformation (TR)* estimator is obtained by first transforming the data $\mathbb{X}_n$ to $\mathbb{Y}_n = S(\mathbb{X}_n)^{-1/2}\mathbb{X}_n$ using a special data-based functional $S(\mathbb{X}_n)$, then minimizing the sample analogue of (3) based on $\mathbb{Y}_n$, and finally "retransforming" that result through multiplication by $S(\mathbb{X}_n)^{1/2}$. As seen in Section 3.3, however, if one asks what the "TR sample spatial median" actually estimates, the answer is: *not* the population spatial median, but rather the corresponding population *Mahalanobis* median.

Similarly, relative to the *spatial quantile function* of Chaudhuri (1996) (discussed as a preliminary in Section 2.2), an affine equivariant "TR sample spatial quantile function" has been developed by Chakraborty (2001). It is shown in Section 3.3 to estimate not the spatial quantile function, but rather a *Mahalanobis* quantile function.

Thus here we study the "Mahalanobis quantile function" in its own right. In comparison with the direct use of (1), it is seen to yield more flexible outlyingness contours. Relative to the practical use of (3), and indeed of the whole spatial quantile function, the Mahalanobis quantile function is seen to provide a rigorous foundation for the TR procedure, strengthening the motivation for it and broadening the scope and technique of its application.

As a key concept used in our development, we formulate a general and flexible definition of *affine equivariance for multivariate quantiles.* This is nontrivial, because, in general, quantile functions of affinely transformed distributions must include a suitable re-indexing such that simultaneously the outlyingness function is affine invariant. Of independent interest, this preliminary topic is treated in Section 2.3.

As another concept instrumental to our development, and also of independent interest, we introduce a notion of *weak covariance functional.* With this we clarify that the TR approach is equivalent to, and best understood as, simply standardization by a (weak) covariance functional. This provides a unified view of the rather *ad hoc* implementations of the TR approach and a model-based formulation with respect to which TR approaches are simply the sample versions. Weak covariance functionals and their connection with "TR functionals" are treated as a preliminary, in Section 2.4.

The four key preliminaries covered in Section 2 have been described above. Section 3 presents our main development, introducing the *Mahalanobis quantile function* and treating its primary ramifications. It covers formulation, a TR representation, equivariance, sample versions, computation, and choice of weak covariance functional for standardization. Further results and remarks are provided in Section 4: additional features and properties, a Bahadur-Kiefer representation for sample versions along with consistency and asymptotic normality, "information" about F contained in the spatial and the Mahalanobis quantile functions, robustness, and variations on formulation of "Mahalanobis" quantiles.

2 Preliminaries

For key preliminaries needed for the formulation and treatment of our Mahalanobis quantile function are treated in Sections 2.1–2.4, respectively. The treatments of *affine equivariance for multivariate quantiles* (Section 2.3) and *weak covariance functionals* (Section 2.4) are of general interest as well.

2.1 Multivariate depth, outlyingness, quantile, and rank functions

. For a univariate distribution F, the quantile function is $F^{-1}(p) = \inf\{x : F(x) \geq p\}$, $0 < p < 1$. As a preliminary to extending to the multivariate case, where a natural linear order is lacking, we orient to a "center" by re-indexing to the open interval $(-1, 1)$ via

$u = 2p - 1$ and representing the quantile function as $Q(u, F) = F^{-1}\left(\frac{1+u}{2}\right)$, $-1 < u > 1$. Each point $x \in \mathbb{R}$ has a quantile representation $x = Q(u, F)$ for some choice of u. The *median* is $Q(0, F)$. For $u \neq 0$, the index u indicates through its sign *the direction of x from the median* and through its magnitude $|u|$ the *outlyingness of x from the median.* For $|u| = c \in (0, 1)$, the "contour" $\left\{F^{-1}\left(\frac{1-c}{2}\right), F^{-1}\left(\frac{1+c}{2}\right)\right\}$ demarks the upper and lower tails of equal probability weight $\frac{1-c}{2}$. Then $|u| = c$ is also the *probability weight* of the enclosed "central region".

In general, for a distribution F on $\mathbb{R}^d$, an associated *quantile function* indexed by $\boldsymbol{u}$ in the unit ball $\mathbb{B}^{d-1}(\mathbf{0})$ in $\mathbb{R}^d$ should attach to each point $\boldsymbol{x}$ a *quantile representation* $\boldsymbol{Q}(\boldsymbol{u}, F)$ and generate *nested* contours

$$\{\boldsymbol{Q}(\boldsymbol{u}, F) : \|\boldsymbol{u}\| = c\},\ 0 \leq c < 1.$$

For $\boldsymbol{u} = \mathbf{0}$, the most central point $\boldsymbol{Q}(\mathbf{0}, F)$ is interpreted as a *d-dimensional median.* For $\boldsymbol{u} \neq \mathbf{0}$, the index $\boldsymbol{u}$ represents *direction from the center* in some sense, and the magnitude $\|\boldsymbol{u}\|$ represents an *outlyingness parameter*, higher values corresponding to more extreme points. The contours thus represent equivalence classes of points of equal outlyingness. (But in general $\|\boldsymbol{u}\|$ need not be the probability weight of the enclosed central regions.)

Three closely related functions have special meanings and roles.

(i) *Centered rank function.* The quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\mathbf{0})$, has an *inverse*, given at each point $\boldsymbol{x} \in \mathbb{R}^d$ by the point $\boldsymbol{u}$ in $\mathbb{B}^{d-1}(\mathbf{0})$ for which $\boldsymbol{x}$ has a quantile interpretation as $\boldsymbol{Q}(\boldsymbol{u}, F)$, i.e., by the solution $\boldsymbol{u}$ of the equation $\boldsymbol{x} = \boldsymbol{Q}(\boldsymbol{u}, F)$. The solutions $\boldsymbol{u}$ define a *centered rank function* $\boldsymbol{R}(\boldsymbol{x}, F)$, $\boldsymbol{x} \in \mathbb{R}^d$, taking values in $\mathbb{B}^{d-1}(\mathbf{0})$ with the origin assigned as rank of the multivariate median $\boldsymbol{Q}(\mathbf{0}, F)$. Thus $\boldsymbol{R}(\boldsymbol{x}, F) = \mathbf{0}$ for $\boldsymbol{x} = \boldsymbol{Q}(\mathbf{0}, F)$ and $\boldsymbol{R}(\boldsymbol{x}, F)$ gives a "directional rank" in $\mathbb{B}^{d-1}(\mathbf{0})$ for other $\boldsymbol{x}$.

(ii) *Outlyingness function.* The magnitude $\|\boldsymbol{R}(\boldsymbol{x}, F)\|$ defines an *outlyingness function* $O(\boldsymbol{x}, F)$, $\boldsymbol{x} \in \mathbb{R}^d$, giving a *center-inward ordering* of points $\boldsymbol{x}$ in $\mathbb{R}^d$.

(iii) *Depth function.* A corresponding *depth function* $D(\boldsymbol{x}, F) = 1 - O(\boldsymbol{x}, F)$ provides a *center-outward ordering* of points $\boldsymbol{x}$ in $\mathbb{R}^d$, higher depth corresponding to higher centrality.

In fact, *depth, outlyingness, quantiles, and ranks in $\mathbb{R}^d$ are all equivalent*:

- $\boldsymbol{Q}(\boldsymbol{u}, F)$ and $\boldsymbol{R}(\boldsymbol{x}, F)$ are equivalent (inversely).
- $D(\boldsymbol{x}, F)$ and $O(\boldsymbol{x}, F)$ are equivalent (inversely).
- These are linked by
  - a) $O(\boldsymbol{x}, F) = \|\boldsymbol{R}(\boldsymbol{x}, F)\|$ $(= \|\boldsymbol{u}\|)$,
  - b) $D(\boldsymbol{x}, F)$ induces a corresponding $\boldsymbol{Q}(\boldsymbol{u}, F)$.

Thus each of D, O, $\boldsymbol{Q}$, and $\boldsymbol{R}$ can generate the others and we may use them interchangeably. (However, as noted in Section 4.1, the "$\boldsymbol{Q}$" generated by "D" can differ in its indexing from that "$\boldsymbol{Q}$" which generates "D", although the contours will still be the same.)

Example 2.1 Depth-induced quantile functions. *For $D(\boldsymbol{x}, F)$ having nested contours enclosing the "median" $\boldsymbol{M}_F$ and bounding "central regions" $\{\boldsymbol{x} : D(\boldsymbol{x}, F) \geq \alpha\}$, $\alpha > 0$, the depth contours induce $\boldsymbol{Q}(\boldsymbol{u}, F)$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\mathbf{0})$, endowing each $\boldsymbol{x} \in \mathbb{R}^d$ with a quantile representation, as follows. For $\boldsymbol{x} = \boldsymbol{M}_F$, denote it by $\boldsymbol{Q}(\mathbf{0}, F)$. For $\boldsymbol{x} \neq \boldsymbol{M}_F$, denote it by $\boldsymbol{Q}(\boldsymbol{u}, F)$ with $\boldsymbol{u} = p\boldsymbol{v}$, where p is the probability weight of the central region with $\boldsymbol{x}$ on its boundary and $\boldsymbol{v}$ is the unit vector toward $\boldsymbol{x}$ from $\boldsymbol{M}_F$. In this case, $\boldsymbol{u} = \boldsymbol{R}(\boldsymbol{x}, F)$ indicates direction toward $\boldsymbol{x} = \boldsymbol{Q}(\boldsymbol{u}, F)$ from $\boldsymbol{M}_F$, and the outlyingness parameter $\|\boldsymbol{u}\| = \|\boldsymbol{R}(\boldsymbol{x}, F)\|$ is the probability weight of the central region with $\boldsymbol{Q}(\boldsymbol{u}, F)$ on its boundary.* □

2.2 Equivariance of multivariate quantile functions

Quantile functions on $\mathbb{R}^d$ are desirably *equivariant*, and *outlyingness functions* should be *invariant*. That is, the new quantile representation of a point $\boldsymbol{x}$ after affine transformation should agree with the original representation similarly transformed, and its outlyingness measure should remain unchanged. The following definition formalizes these requirements.

Definition 2.1 *An $\mathbb{R}^d$-valued quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\mathbf{0})$, is affine equivariant if, for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$,*

$$\boldsymbol{Q}(\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}}), F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}(\boldsymbol{u}, F_{\boldsymbol{X}}) + \boldsymbol{b},\ \ \boldsymbol{u} \in \mathbb{B}^{d-1}(\mathbf{0}), \tag{4}$$

with a re-indexing $\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}})$ which is a vector-valued function of $\boldsymbol{u}$, $\boldsymbol{A}$, and $F_{\boldsymbol{X}}$ taking values in $\mathbb{B}^{d-1}(\mathbf{0})$ and satisfying

$$\|\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}})\| = \|\boldsymbol{u}\|,\ \ \boldsymbol{u} \in \mathbb{B}^{d-1}(\mathbf{0}). \tag{5}$$

In particular, $\boldsymbol{Q}(\mathbf{0}, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}(\mathbf{0}, F_{\boldsymbol{X}}) + \boldsymbol{b}$. □

In some cases the condition (5) is achieved through $\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}})$ being an *orthogonal* linear transformation of $\boldsymbol{u}$.

Example 2.2 The univariate case. *For the univariate quantile function in center-outward notation, $Q(u, F) = F^{-1}(\frac{1+u}{2})$, $-1 < u < 1$, as discussed in Section 2.1 above, the usual translation and scale equivariance takes the form*

$$Q(\mathrm{sgn}(a)u, F_{aX+b}) = a\,Q(u, F_X) + b,\ \ -1 < u < 1,$$

for all $a, b \in \mathbb{R}$, which is (4) with $\tilde{a}(u, a, F_X) = \mathrm{sgn}(a)u$, satisfying (5). (We could equivalently choose $\tilde{a}(u, a, F_X) = -\mathrm{sgn}(a)u$.) □

Remark 2.1 *The equivariance of $\boldsymbol{Q}(\boldsymbol{u}, F)$ yields equivariance and invariance properties for the related functions. By the definition of $\boldsymbol{R}(\boldsymbol{x}, F)$, (4) immediately yields equivariance of the centered rank function in the following sense:*

$$\boldsymbol{R}(\boldsymbol{y}, F_{\boldsymbol{Y}}) = \widetilde{\boldsymbol{A}}\left(\boldsymbol{R}(\boldsymbol{x}, F_{\boldsymbol{X}}), \boldsymbol{A}, F_{\boldsymbol{X}}\right). \tag{6}$$

In turn, the relation $O(\boldsymbol{x}, F) = \|\boldsymbol{R}(\boldsymbol{x}, F)\|$ yields through (5) and (6) invariance of the outlyingness function, $O(\boldsymbol{y}, F_{\boldsymbol{Y}}) = O(\boldsymbol{x}, F_{\boldsymbol{X}})$, and likewise invariance of the depth function, $D(\boldsymbol{y}, F_{\boldsymbol{Y}}) = D(\boldsymbol{x}, F_{\boldsymbol{X}})$. □

Example 2.3 $\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}})$ **for depth-induced quantile functions**. *Suppose that a quantile function* $\boldsymbol{Q}(\boldsymbol{u}, F)$ *is constructed as in Example 2.1 and satisfies (4). Then (in obvious notation)* $\boldsymbol{M}_{\boldsymbol{Y}} = \boldsymbol{A}\,\boldsymbol{M}_{\boldsymbol{X}} + \boldsymbol{b}$*, from which it follows that the unnormalized direction vector toward* $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ *from* $\boldsymbol{M}_{\boldsymbol{Y}}$ *is given by* $\boldsymbol{y} - \boldsymbol{M}_{\boldsymbol{Y}} = \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{M}_{\boldsymbol{X}})$*. Therefore, for some constant* c_0*, we have* $\boldsymbol{R}(\boldsymbol{y}, F_{\boldsymbol{Y}}) = c_0 \boldsymbol{A}\boldsymbol{R}(\boldsymbol{x}, F_{\boldsymbol{X}})$*, or, equivalently,*

$$\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}}) = c_0 \boldsymbol{A}\boldsymbol{u}.$$

Then the requirement (5) is satisfied if and only if $|c_0| = \|\boldsymbol{u}\|/\|\boldsymbol{A}\boldsymbol{u}\|$*, yielding*

$$\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}}) = \pm \frac{\|\boldsymbol{u}\|}{\|\boldsymbol{A}\boldsymbol{u}\|} \boldsymbol{A}\boldsymbol{u}, \tag{7}$$

for either choice of sign. In the univariate case, (7) reduces to $\pm\mathrm{sgn}(a)u$*, in agreement with Example 2.2. Note that for* $\boldsymbol{A}$ *orthogonal (7) becomes simply*

$$\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}}) = \pm \boldsymbol{A}\boldsymbol{u}. \tag{8}$$

□

2.3 The Spatial Quantile Function

For univariate Z with $E|Z| < \infty$, the univariate pth quantile for $0 < p < 1$ may be characterized as any value θ minimizing

$$E\{|Z - \theta| + (2p - 1)(Z - \theta)\} \tag{9}$$

(Ferguson, 1967, p. 51). In the center-outward notation via $u = 2p - 1$, and defining $\Phi(u, t) = |t| + ut$, $-1 < u < 1$, we may equivalently obtain θ by minimizing

$$E\{\Phi(u, Z - \theta) - \Phi(u, Z)\}, \tag{10}$$

where subtraction of $\Phi(u, Z)$ eliminates the need of a moment assumption on Z. As a multivariate extension, d-dimensional "spatial" or "geometric" quantiles were introduced by Dudley and Koltchinskii (1992) and Chaudhuri (1996). Following the latter, we extend the index set to the open unit ball $\mathbb{B}^{d-1}(\boldsymbol{0})$ and minimize a generalized form of (10). Specifically, for random vector $\boldsymbol{X}$ having cdf F on $\mathbb{R}^d$, and for $\boldsymbol{u}$ in $\mathbb{B}^{d-1}(\boldsymbol{0})$, the $\boldsymbol{u}$*th spatial quantile* $\boldsymbol{Q}_S(\boldsymbol{u}, F)$ is given by $\boldsymbol{\theta}$ minimizing

$$E\{\Phi(\boldsymbol{u}, \boldsymbol{X} - \boldsymbol{\theta}) - \Phi(\boldsymbol{u}, \boldsymbol{X})\}, \tag{11}$$

where $\Phi(\boldsymbol{u}, \boldsymbol{t}) = \|\boldsymbol{t}\| + \boldsymbol{u}'\boldsymbol{t}$. In particular, $\boldsymbol{Q}_S(\boldsymbol{0}, F)$ is the well-known *spatial median.*

Equivalently, in terms of the *spatial sign function* (or *unit vector function*),

$$\boldsymbol{S}(\boldsymbol{x}) = \begin{cases} \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, & \boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{x} \neq \boldsymbol{0}, \\ \boldsymbol{0}, & \boldsymbol{x} = \boldsymbol{0}, \end{cases}$$

the quantile $\boldsymbol{Q}_S(\boldsymbol{u}, F)$ may be represented as the solution $\boldsymbol{x}$ of

$$E\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\} = \boldsymbol{u}. \tag{12}$$

By (12), $\boldsymbol{Q}_S(\boldsymbol{u}, F)$ is obtained by inverting the map $\boldsymbol{x} \mapsto E\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\}$, from which it is seen that spatial quantiles are a special case of the "M-quantiles" introduced by Breckling and Chambers (1988) and also treated by Koltchinskii (1997) and Breckling, Kokic and Lübke (2001). Solving (12) for $\boldsymbol{u}$ yields the *spatial centered rank function*,

$$\boldsymbol{R}_S(\boldsymbol{x}, F) = E\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\}, \tag{13}$$

the *expected direction* to $\boldsymbol{x}$ from a random point $\boldsymbol{X} \sim F$. (Compare with the interpretation of the *depth-induced* $\boldsymbol{R}(\boldsymbol{x}, F)$ as *direction from the median*, as seen in Example 2.1.) Möttönen and Oja (1995) apply the function $\boldsymbol{R}_S(\boldsymbol{x}, F)$ in hypothesis testing, and Vardi and Zhang (2000) treat the related *spatial depth function*, $D_S(\boldsymbol{x}, F) = 1 - \|\boldsymbol{R}_S(\boldsymbol{x}, F)\|$, whose inverse $O_S(\boldsymbol{x}, F) = \|\boldsymbol{R}_S(\boldsymbol{x}, F)\|$ $(= \|\boldsymbol{u}\|$ relative to equation (12)) represents a *spatial outlyingness function.* Koltchinskii (1994a,b, 1997) develops a Bahadur-Kiefer represention for the sample spatial quantile function and derives other theoretical properties of $\boldsymbol{Q}_S(\boldsymbol{u}, F)$. Serfling (2004) treats further properties of $\boldsymbol{Q}_S(\boldsymbol{u}, F)$ and introduces related nonparametric multivariate descriptive measures for location, spread, skewness, and kurtosis,

A well-known limitation of the spatial quantile function is its *orthogonal*, rather than fully affine, equivariance. Indeed, as pointed out by Chaudhuri (1996, Fact 2.2.1), for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any *orthogonal* $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$,

$$\boldsymbol{Q}_S(\boldsymbol{A}\boldsymbol{u}, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}_S(\boldsymbol{u}, F_{\boldsymbol{X}}) + \boldsymbol{b},\ \boldsymbol{u} \in \mathbb{B}^{d-1}, \tag{14}$$

which corresponds to Definition 2.1 restricted to *orthogonal* $\boldsymbol{A}$ and thus with $\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}}) = \boldsymbol{A}\boldsymbol{u}$. Consequently, a point $\boldsymbol{x}$ labeled a (spatial) "outlier" or "nonoutlier" would have the same classification after orthogonal transformation to a new coordinate system but not necessarily after transformation by heterogeneous scale changes.

2.4 Weak Covariance Functionals

A symmetric positive definite $d \times d$ matrix-valued functional $\boldsymbol{C}(F)$ defined on distributions F on $\mathbb{R}^d$ is called a *covariance functional* if it satisfies *covariance equivariance*:

$$\boldsymbol{C}(F_{\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}}) = \boldsymbol{A}\boldsymbol{C}(F_{\boldsymbol{X}})\boldsymbol{A}'$$

for all vectors $\boldsymbol{b}$ and all nonsingular $d \times d$ A (Rousseeuw and Leroy, 1987, and Lopuhaä and Rousseeuw, 1991). Here we introduce a somewhat broader notion that suffices for most purposes involving covariance equivariance, yet offers useful additional flexibility.

Definition 2.2 Weak covariance functionals, and weak covariance equivariance. *A symmetric positive definite $d \times d$ matrix-valued functional $\boldsymbol{C}(F)$ is called a weak covariance functional if it satisfies weak covariance equivariance:*

$$\boldsymbol{C}(F_{\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}}) = k(\boldsymbol{A}, F_{\boldsymbol{X}})\boldsymbol{A}\boldsymbol{C}(F_{\boldsymbol{X}})\boldsymbol{A}' \tag{15}$$

for any nonsingular $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$, with $k(\boldsymbol{A}, F_{\boldsymbol{X}})$ a positive scalar function of $\boldsymbol{A}$ and $F_{\boldsymbol{X}}$. □

The strict version is recovered with $k(\boldsymbol{A}, F_{\boldsymbol{X}}) \equiv 1$. The sample version of (15) for a data set $\mathbb{X}_n$ may be expressed as

$$\boldsymbol{C}_n(\boldsymbol{A}\mathbb{X}_n + \boldsymbol{b}) = k(\boldsymbol{A}, \mathbb{X}_n)\boldsymbol{A}\boldsymbol{C}_n(\mathbb{X}_n)\boldsymbol{A}'. \tag{16}$$

The following general lemma gives an important property of a key matrix that arises in using covariance functionals.

Lemma 2.1 *For any weak covariance functional $\boldsymbol{C}(F)$, and for any nonsingular $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$, the matrix*

$$\widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}}) = (\boldsymbol{A}\,\boldsymbol{C}(F_{\boldsymbol{X}})\,\boldsymbol{A}')^{1/2}(\boldsymbol{A}')^{-1}\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}, \tag{17}$$

is orthogonal.

Proof. With $\widetilde{\boldsymbol{A}} = \widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}})$, we have,

$$\begin{aligned}
(\widetilde{\boldsymbol{A}})'\widetilde{\boldsymbol{A}} &= \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{A}^{-1}(\boldsymbol{A}\,\boldsymbol{C}(F_{\boldsymbol{X}})\,\boldsymbol{A}')^{1/2}(\boldsymbol{A}\,\boldsymbol{C}(F_{\boldsymbol{X}})\,\boldsymbol{A}')^{1/2}(\boldsymbol{A}')^{-1}\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2} \\
&= \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{A}^{-1}(\boldsymbol{A}\,\boldsymbol{C}(F_{\boldsymbol{X}})\,\boldsymbol{A}')(\boldsymbol{A}')^{-1}\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2} \\
&= \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{C}(F_{\boldsymbol{X}})\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2} \\
&= \boldsymbol{I}_d.
\end{aligned}$$

□

By (15), an alternative expression for $\widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}})$ is

$$\widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}}) = k(\boldsymbol{A}, F_{\boldsymbol{X}})^{-1/2}\boldsymbol{C}(F_{\boldsymbol{A}\boldsymbol{X}+\boldsymbol{b}})^{1/2}(\boldsymbol{A}')^{-1}\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}. \tag{18}$$

For a weak covariance functional given by Tyler (1987) that we discuss in Section 3.4, the sample version of Lemma 2.1 using expression (18) is noted and utilized by Randles (2000).

Remark 2.2 Two special cases. *In some typical cases of $\boldsymbol{A}$, the matrix $\widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}})$ in Lemma 2.1 is quite simple.*

(i) For $\boldsymbol{A}$ symmetric, $\widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}}) = \boldsymbol{I}_d$. For proof, check and apply the fact that, for any weak covariance functional $\boldsymbol{C}(F)$ and $\boldsymbol{A}$ symmetric,

$$(\boldsymbol{A}\,\boldsymbol{C}(F_{\boldsymbol{X}})\,\boldsymbol{A}')^{1/2} = \boldsymbol{A}\,\boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}.$$

For example, coordinatewise scale changes correspond to transformation of $\boldsymbol{X}$ by a diagonal matrix $\boldsymbol{A}$.

(ii) For $\boldsymbol{A}$ orthogonal, $\widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}}) = \boldsymbol{A}$. For proof, check and apply the fact that, for any weak covariance functional $\boldsymbol{C}(F)$ and $\boldsymbol{A}$ orthogonal,

$$(\boldsymbol{A}\,\boldsymbol{C}(F_{\boldsymbol{X}})\,\boldsymbol{A}')^{1/2} = \boldsymbol{A}\,\boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}\,\boldsymbol{A}'.$$

This will be relevant in Section 3.2 in discussing statistical procedures that are orthogonally, but not fully affine, equivariant or invariant. □

Remark 2.3 Standardization of a standardized variable. *Consider standardization of X by any weak covariance functional $\boldsymbol{C}(F_{\boldsymbol{X}})$: $\boldsymbol{Y} = \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}$.*

(i) Using (15) with $\boldsymbol{X}$ replaced by $\boldsymbol{Y}$, $\boldsymbol{A} = \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}$, and $\boldsymbol{b} = \boldsymbol{0}$, we immediately obtain

$$\boldsymbol{C}(F_{\boldsymbol{Y}}) = k^{-1}\left(\boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}\right)\boldsymbol{I}_d. \tag{19}$$

That is, standardization of the (already standardized) $\boldsymbol{Y}$ by $\boldsymbol{C}(F_{\boldsymbol{Y}})^{-1/2}$ simply multiplies $\boldsymbol{Y}$ by

$$k^{1/2}\left(\boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}\right).$$

(ii) Also, using Remark 2.2(i), the matrix

$$\widetilde{\boldsymbol{A}}\left(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}\right)$$

of Lemma 2.1 is seen to be simply $\boldsymbol{I}_d$. □

Remark 2.4 The identity matrix as a covariance functional. *With $\boldsymbol{C}(F) \equiv \boldsymbol{I}_d$, condition (15) with $k(\boldsymbol{A}, F_{\boldsymbol{X}}) \equiv 1$ is fulfilled for all orthogonal $\boldsymbol{A}$. That is, the identity matrix $\boldsymbol{I}_d$ is a (strict) covariance functional if only orthogonal covariance equivariance is required.* □

Connection with "Transformation-Retransformation" (TR) Approaches

Various *ad hoc* "transformation-retransformation" (TR) approaches have been introduced as a means of modifying certain sample testing or estimation procedures in order to achieve full affine invariance or equivariance. See Randles (2000) for broad discussion. The TR approach transforms the data to a new coordinate system using a particular kind of data-based nonsingular $d \times d$ matrix transformation $\boldsymbol{M}(\mathbb{X}_n)$,

$$\boldsymbol{x} \mapsto \boldsymbol{M}(\mathbb{X}_n)\,\boldsymbol{x},$$

then carries out the procedure on the transformed data and retransforms that result back to the original coordinate system. A structural requirement for such an $\boldsymbol{M}(\mathbb{X}_n)$, expressed in terms of the population analogue functional $\boldsymbol{M}(F)$, is

$$\boldsymbol{A}'\boldsymbol{M}(F_{\boldsymbol{AX}+\boldsymbol{b}})'\boldsymbol{M}(F_{\boldsymbol{AX}+\boldsymbol{b}})\boldsymbol{A} = k_0(\boldsymbol{A}, F_{\boldsymbol{X}})\boldsymbol{M}(F_{\boldsymbol{X}})'\boldsymbol{M}(F_{\boldsymbol{X}}), \tag{20}$$

with $k_0(\boldsymbol{A}, F_{\boldsymbol{X}})$ a positive scalar function of $F_{\boldsymbol{X}}$ and $\boldsymbol{A}$. Equivalently,

$$\boldsymbol{M}(F_{\boldsymbol{AX}+\boldsymbol{b}})'\boldsymbol{M}(F_{\boldsymbol{AX}+\boldsymbol{b}}) = k_0(\boldsymbol{A}, F_{\boldsymbol{X}})(\boldsymbol{A}')^{-1}\boldsymbol{M}(F_{\boldsymbol{X}})'\boldsymbol{M}(F_{\boldsymbol{X}})(\boldsymbol{A})^{-1}.$$

The following result clarifies that the design or selection of a TR functional is merely an indirect but equivalent way to select a (weak) covariance functional, for which an extensive literature already provides many choices designed to meet various desired criteria of robustness and computational efficiency. The proof is straightforward.

Lemma 2.2 *Every TR functional $\boldsymbol{M}(F)$ is equivalent to a weak covariance functional, and conversely:*

(i) Given $\boldsymbol{M}(F)$ satisfying (20), the functional $\boldsymbol{C}(F) = (\boldsymbol{M}(F)'\boldsymbol{M}(F))^{-1}$ satisfies (15).

(ii) Given $\boldsymbol{C}(F)$ satisfying (15), the functional $\boldsymbol{M}(F) = \boldsymbol{C}(F)^{-1/2}$ satisfies (20).

3 A Mahalanobis Quantile Function

For $\boldsymbol{X}$ having cdf F on $\mathbb{R}^d$, and for a given weak covariance functional $\boldsymbol{C}(\cdot)$, the corresponding *Mahalanobis quantile function* $\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}})$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\boldsymbol{0})$, is defined at $\boldsymbol{u}$ as the vector $\boldsymbol{\theta}$ minimizing

$$E\left\{\Phi\left(\boldsymbol{u}, \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{X}-\boldsymbol{\theta})\right) - \Phi\left(\boldsymbol{u}, \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}\right)\right\}, \tag{21}$$

analogous to the definition of spatial quantile but using a standardized deviation. For $\boldsymbol{u} = \boldsymbol{0}$, we have the *Mahalanobis median*, which minimizes

$$E\left\{\left\|\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{X}-\boldsymbol{\theta})\right\| - \left\|\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}\right\|\right\},$$

as discussed already in Section 1. The Mahalanobis quantile function has an explicit and productive connection with the spatial quantile function. This "TR representation" is treated in Section 3.1. Equivariance properties are developed in Section 3.2. Sample versions are discussed in Sections 3.3 and 3.5, and the choice of $\boldsymbol{C}(F)$ in Section 3.4.

3.1 Transformation-Retransformation (TR) Representation

Note that (21) is simply (11) with $\boldsymbol{X}$ and $\boldsymbol{\theta}$ replaced by $\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}$ and $\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{\theta}$, respectively. Therefore, if $\boldsymbol{\theta}$ minimizes (21), then $\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{\theta}$ minimizes (11) with $\boldsymbol{X}$ replaced by $\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}$. This relationship may be characterized as a *"transformation-retransformation" representation* for the Mahalanobis quantile function in terms of the spatial quantile function:

$$\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) = \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}\,\boldsymbol{Q}_S(\boldsymbol{u}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}). \tag{22}$$

For $\boldsymbol{u} = \boldsymbol{0}$, the TR representation (22) states that

$$(\textit{Mahalanobis}\text{ median of }\boldsymbol{X}) \;=\; \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2} \times \left(\textit{spatial}\text{ median of }\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}\right). \tag{23}$$

For helpful perspective, consider the *univariate case*,

$$(\text{median of } X) \;=\; \sigma \times (\text{median of } \sigma^{-1}X), \tag{24}$$

which simply expresses the *scale equivariance* of the univariate median. However, in passing to the multivariate *spatial* median in place of the univariate median, we cannot assert (23) with "spatial median" on the left-hand side, due to the lack of full affine equivariance of the spatial median. Consequently, in extending (24) to (23), the right-hand side needs a *new name* and a *new interpretation*. This is supplied by the term "Mahalanobis median", interpreted as the point from which expected Mahalanobis distance is minimized. For $\boldsymbol{u} \neq \boldsymbol{0}$, the TR representation (22) asserts

$$(\textit{Mahalanobis } \boldsymbol{u}\text{th quantile of } \boldsymbol{X}) \; = \; \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2} \times \left(\textit{spatial } \boldsymbol{u}\text{th quantile of } \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}\right),$$

which similarly extends a univariate relationship,

$$(u\text{th quantile of } X) \; = \; \sigma \times (u\text{th quantile of } \sigma^{-1}X).$$

Also, in view of this discussion, it is clear that use of the same index $\boldsymbol{u}$ on both sides of (22) is consistent with the univariate case and quite natural.

The TR representation (22) for the Mahalanobis quantile function readily yields similar representatons for the Mahalanobis centered rank, depth, and outlyingness functions. The definition of $\boldsymbol{R}_M(\boldsymbol{x}, F_{\boldsymbol{X}})$ via $\boldsymbol{x} = \boldsymbol{Q}_M(\boldsymbol{R}_M(\boldsymbol{x}, F_{\boldsymbol{X}}), F_{\boldsymbol{X}})$, along with (12), yields

$$\begin{aligned} \boldsymbol{R}_M(\boldsymbol{x}, F_{\boldsymbol{X}}) &= \boldsymbol{R}_S\left(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{x}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}\right) \\ &= E\left\{\boldsymbol{S}\left(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{x} - \boldsymbol{X})\right)\right\}. \end{aligned} \tag{25}$$

From this immediately follows

$$\begin{aligned} O_M(\boldsymbol{x}, F_{\boldsymbol{X}}) &= O_S\left(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{x}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}\right) \\ &= \left\|E\left\{\boldsymbol{S}\left(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{x} - \boldsymbol{X})\right)\right\}\right\| \end{aligned} \tag{26}$$

and $D_M(\boldsymbol{x}, F_{\boldsymbol{X}}) = 1 - O_M(\boldsymbol{x}, F_{\boldsymbol{X}})$.

3.2 Equivariance and Invariance Properties

The following result establishes that, with a suitable re-indexing, the Mahalanobis quantile function is *fully affine equivariant*.

Lemma 3.1 *For* $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ *with any nonsingular* $d \times d$ $\boldsymbol{A}$ *and any* $\boldsymbol{b}$*,*

$$\boldsymbol{Q}_M(\widetilde{\boldsymbol{A}}\boldsymbol{u}, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) + \boldsymbol{b}, \tag{27}$$

with $\widetilde{\boldsymbol{A}} = \widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}})$ *as given in (17) of Lemma 2.1 (and hence orthogonal).*

That is, $\boldsymbol{Q}_M(\boldsymbol{u}, F)$ satisfies Definition 2.1 with

$$\widetilde{\boldsymbol{A}}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}}) = \widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}})\,\boldsymbol{u}. \tag{28}$$

From Remark 2.2, we note that $\widetilde{\boldsymbol{A}}\boldsymbol{u}$ in (27) is $\boldsymbol{u}$ for $\boldsymbol{A}$ *symmetric* and $\boldsymbol{A}\boldsymbol{u}$ for $\boldsymbol{A}$ *orthogonal.*

Indeed, with $\boldsymbol{C}(F) \equiv \boldsymbol{I}_d$ and a restriction to $\boldsymbol{A}$ *orthogonal*, the Mahalanobis quantile function is just the spatial quantile function and $\boldsymbol{I}_d$ is a covariance functional as noted in Remark 2.4. Then (27) yields the orthogonal equivariance of the spatial quantile function stated earlier in (14).

As per Remark 2.1, the equivariance of $\boldsymbol{Q}_M(\boldsymbol{u}, F)$ yields corresponding equivariance and invariance properties for the related functions $\boldsymbol{R}_M(\boldsymbol{x}, F)$, $O_M(\boldsymbol{x}, F)$, and $D_M(\boldsymbol{x}, F)$. In particular, (6) yields

$$\boldsymbol{R}(\boldsymbol{y}, F_{\boldsymbol{Y}}) = \widetilde{\boldsymbol{A}}(\boldsymbol{A}, F_{\boldsymbol{X}})\, \boldsymbol{R}(\boldsymbol{x}, F_{\boldsymbol{X}}). \tag{29}$$

We note that (28) is different from the form (7) arising with depth-induced quantile functions as discussed in Example 2.3 (although these agree for the case of $\boldsymbol{A}$ *orthogonal*).

Proof of Lemma 3.1. Let $\boldsymbol{\theta}$ minimize (21) and put $\boldsymbol{\eta} = \boldsymbol{A\theta} + \boldsymbol{b}$ and $\boldsymbol{Y} = \boldsymbol{Ax} + \boldsymbol{b}$, with $\boldsymbol{A}$ nonsingular. It suffices for (27) to show that $\boldsymbol{\eta}$ minimizes

$$E\left\{\Phi\left(\widetilde{\boldsymbol{A}}\boldsymbol{u}, \boldsymbol{C}(F_{\boldsymbol{Y}})^{-1/2}(\boldsymbol{Y} - \boldsymbol{\eta})\right) - \Phi\left(\widetilde{\boldsymbol{A}}\boldsymbol{u}, \boldsymbol{C}(F_{\boldsymbol{Y}})^{-1/2}\boldsymbol{Y}\right)\right\}. \tag{30}$$

Using $\boldsymbol{Y} - \boldsymbol{\eta} = \boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\theta})$, the weak covariance equivariance condition (15), and nonsingularity of $\boldsymbol{A}$, it is quickly checked that

$$\|\boldsymbol{C}(F_{\boldsymbol{Y}})^{-1/2}(\boldsymbol{Y} - \boldsymbol{\eta})\| = k(\boldsymbol{A}, F_{\boldsymbol{X}})^{-1/2}\, \|\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{X} - \boldsymbol{\theta})\|.$$

Also, again applying (15), we obtain

$$(\widetilde{\boldsymbol{A}}\boldsymbol{u})'\boldsymbol{C}(F_{\boldsymbol{Y}})^{-1/2}(\boldsymbol{Y} - \boldsymbol{\eta}) = k(\boldsymbol{A}, F_{\boldsymbol{X}})^{-1/2}\, \boldsymbol{u}'\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{X} - \boldsymbol{\theta}).$$

Hence

$$\Phi\left(\widetilde{\boldsymbol{A}}\boldsymbol{u}, \boldsymbol{C}(F_{\boldsymbol{Y}})^{-1/2}(\boldsymbol{Y} - \boldsymbol{\eta})\right) = k(\boldsymbol{A}, F_{\boldsymbol{X}})^{-1/2}\, \Phi\left(\boldsymbol{u}, \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{X} - \boldsymbol{\theta})\right).$$

Similarly,

$$\Phi\left(\widetilde{\boldsymbol{A}}\boldsymbol{u}, \boldsymbol{C}(F_{\boldsymbol{Y}})^{-1/2}\boldsymbol{Y}\right) = k(\boldsymbol{A}, F_{\boldsymbol{X}})^{-1/2}\, \Phi\left(\boldsymbol{u}, \boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}\right).$$

Consequently, with $\boldsymbol{\eta} = \boldsymbol{A\theta} + \boldsymbol{b}$ the quantity in (30) is equal to $k(\boldsymbol{A}, F_{\boldsymbol{X}})^{-1/2}$ times that in (21), and so both quantities are simultaneously minimized by $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, respectively. This completes the proof. □

Remark 3.1 Connection between TR representation and equivariance. *Starting with* $\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}})$ *on the right-hand side of (27) and taking* $\boldsymbol{A} = \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}$ *and* $\boldsymbol{b} = \boldsymbol{0}$, *and putting* $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X} + \boldsymbol{b} = \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X} = \boldsymbol{X}$, *and using* $\widetilde{\boldsymbol{A}}\boldsymbol{u} = \boldsymbol{u}$ *applying symmetry of* $\boldsymbol{A} = \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}$ *as per Remark 2.2, we obtain,*

$$\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) = \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}\, \boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}). \tag{31}$$

Comparing (31) with the TR representation (22), we obtain

$$\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}) = \boldsymbol{Q}_S(\boldsymbol{u}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}), \tag{32}$$

that is, the Mahalanobis and spatial quantile functions of the already standardized variable $\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}$ *agree. This is essentially explained by Remark 2.3.* □

3.3 Sample Version: a TR Sample Spatial Quantile Function

The *sample Mahalanobis quantile function* based on a given weak covariance functional $\boldsymbol{C}(F)$ may be expressed, by the TR representation (22) and recalling the notation of (16) for sample versions, as

$$\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n) = \boldsymbol{C}_n(\mathbb{X}_n)^{1/2}\,\boldsymbol{Q}_S(\boldsymbol{u}, \boldsymbol{C}_n(\mathbb{X}_n)^{-1/2}\mathbb{X}_n). \tag{33}$$

By Lemma 3.1 applied to sample versions, *full affine equivariance* follows for $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$.

By (33), the sample Mahalanobis quantile function thus may be characterized as a *TR sample spatial quantile function* based on the sample analogue of the TR functional $\boldsymbol{M}(F) = \boldsymbol{C}(F)^{-1/2}$. In this spirit, we could use the notation "$\boldsymbol{Q}_S^{(\mathrm{TR})}(\boldsymbol{u}, \mathbb{X}_n)$" for the left-hand side of (33). Hence the use of a TR sample spatial quantile function as a correction to the lack of full affine equivariance of the sample spatial quantile function is simply a substitution of the sample *Mahalanobis* quantile function for the sample *spatial* quantile function. This is tantamount, therefore, to adopting the *Mahalanobis* quantile function in the first place as the target population quantile function to be estimated, a way of thinking that we recommend.

In particular, the *sample Mahalanobis median* $\boldsymbol{Q}_M(\boldsymbol{0}, \mathbb{X}_n)$ may be viewed as a *TR sample spatial median*, "$\boldsymbol{Q}_S^{(\mathrm{TR})}(\boldsymbol{0}, \mathbb{X}_n)$". Previously, Chakraborty, Chaudhuri, and Oja (1998) have introduced and investigated a TR sample spatial median based upon a particular TR functional they construct that makes the estimator possess favorable relative efficiency in elliptically symmetric models. Of course, in a broader view, any weak covariance functional $\boldsymbol{C}(F)$ may be selected from the many possibilities in the literature, and in Section 3.4 we consider some relevant criteria for such a choice. The foregoing also clarifies that the TR sample spatial median estimates the population *Mahalanobis* median

$$\boldsymbol{Q}_M(\boldsymbol{0}, F_{\boldsymbol{X}}) = \boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}\,\boldsymbol{Q}_S\left(\boldsymbol{0}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}\right)$$

which minimizes (21), rather than the spatial median $\boldsymbol{Q}_S(\boldsymbol{0}, F_{\boldsymbol{X}})$ which minimizes (11). Since the spatial median is not fully affine equivariant, $\boldsymbol{Q}_M(\boldsymbol{0}, F_{\boldsymbol{X}})$ and $\boldsymbol{Q}_S(\boldsymbol{0}, F_{\boldsymbol{X}})$ are different location parameters of $F_{\boldsymbol{X}}$ (and only $\boldsymbol{Q}_M(\boldsymbol{0}, F_{\boldsymbol{X}})$ is fully affine equivariant).

Chakraborty (2001) extends the approach of Chakraborty, Chaudhuri, and Oja (1998), employing the same TR functional, to construct an associated TR sample spatial quantile function given by

$$\boldsymbol{Q}_S^{(\mathrm{C})}(\boldsymbol{u}, \mathbb{X}_n) = \boldsymbol{C}_n(\mathbb{X}_n)^{1/2}\,\boldsymbol{Q}_S(\boldsymbol{\nu}(\boldsymbol{u}, \mathbb{X}_n), \boldsymbol{C}_n(\mathbb{X}_n)^{-1/2}\mathbb{X}_n), \tag{34}$$

where

$$\boldsymbol{\nu}(\boldsymbol{u}, \mathbb{X}_n) = \frac{\|\boldsymbol{u}\|}{\|\boldsymbol{C}_n(\mathbb{X}_n)^{-1/2}\boldsymbol{u}\|}\boldsymbol{C}_n(\mathbb{X}_n)^{-1/2}\boldsymbol{u}.$$

The re-indexing $\boldsymbol{\nu}(\boldsymbol{u}, \mathbb{X}_n)$ has population analogue

$$\boldsymbol{\nu}(\boldsymbol{u}, F_{\boldsymbol{X}}) = \frac{\|\boldsymbol{u}\|}{\|\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{u}\|}\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{u}.$$

Thus, in view of (33), $\boldsymbol{Q}_S^{(\mathrm{C})}(\boldsymbol{u}, \mathbb{X}_n)$ estimates a re-indexed Mahalanobis quantile function, $\boldsymbol{Q}_M(\boldsymbol{\nu}(\boldsymbol{u}, F_{\boldsymbol{X}}), F_{\boldsymbol{X}})$, rather than $\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}})$. It is clear from the discussion in Section 3.1, however, that such a re-indexing is not required for construction of a proper "TR sample spatial quantile function". On the other hand, this particular re-indexing causes no harm because it satisfies

$$\|\boldsymbol{\nu}(\boldsymbol{u}, F_{\boldsymbol{X}})\| = \|\boldsymbol{u}\|,$$

so that the outlyingness parameters associated with the representations of a point $\boldsymbol{x}$ by $\boldsymbol{Q}_M(\boldsymbol{\nu}(\boldsymbol{u}, F_{\boldsymbol{X}}), F_{\boldsymbol{X}})$ and $\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}})$, respectively, and likewise with the representations by $\boldsymbol{Q}_S^{(\mathrm{TR})}(\boldsymbol{\nu}(\boldsymbol{u}, \mathbb{X}_n), \mathbb{X}_n)$ and $\boldsymbol{Q}_S^{(\mathrm{TR})}(\boldsymbol{u}, \mathbb{X}_n)$, respectively, are the same.

3.4 Choice of Weak Covariance Functional in Defining $\boldsymbol{Q}_M(\cdot, F)$

For practical application, we desire a weak covariance functional $\boldsymbol{C}(F)$ for which the sample version $\boldsymbol{C}(\mathbb{X}_n)$ is both *robust* and *computationally efficient*. By Lemma 2.2, covariance and TR functionals may be discussed interchangeably. The selection of a functional involves a *trade-off* between robustness and computational ease. Below we consider two cases at opposite ends of the spectrum of choices.

The Tyler (1987) scatter estimator. Of particular interest is an M-estimate of scatter introduced by Tyler (1987) as "most robust" in the sense of minimizing over a class of competing estimators the maximum asymptotic variance over a class of elliptically symmetric distributions. An iterative algorithm using Cholesky factorizations to compute the sample version quickly in any practical dimension is also provided. Randles (2000) applies this scatter measure as a TR functional in designing a new multivariate location test, showing that (20) is satisfied with $k_0(\boldsymbol{A}, F_{\boldsymbol{X}}) \equiv d^2$. The relevant TR functional $\boldsymbol{M}_0(F)$ is the upper triangular $d \times d$ matrix with positive diagonal elements and "1" in the upper left cell satisfying

$$E\left\{ \frac{\boldsymbol{M}_0(F_{\boldsymbol{X}})\boldsymbol{X}}{\|\boldsymbol{M}_0(F_{\boldsymbol{X}})\boldsymbol{X}\|} \left(\frac{\boldsymbol{M}_0(F_{\boldsymbol{X}})\boldsymbol{X}}{\|\boldsymbol{M}_0(F_{\boldsymbol{X}})\boldsymbol{X}\|} \right)' \right\} = d^{-1}\boldsymbol{I}_d. \tag{35}$$

The sample version is unique if $n > d(d-1)$, and the n transformed data points lie on axes corresponding to unit vectors approximately "equally spaced".

It is informative to re-express (35) in terms of the equivalent weak covariance functional (via $\boldsymbol{M}_0(F) = \boldsymbol{C}_0(F)^{-1/2}$) and the sign function:

$$E\left\{ \boldsymbol{S}(\boldsymbol{C}_0(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}) \boldsymbol{S}(\boldsymbol{C}_0(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X})' \right\} = d^{-1}\boldsymbol{I}_d. \tag{36}$$

Then $\boldsymbol{C}_0(F)$ is seen to be a weak covariance functional generating a scale standardization of the data such that the transformed observations have *expected sign covariance matrix* = $d^{-1}\boldsymbol{I}_d$, reflecting uncorrelated components.

The Tyler (1987) estimator assumes, however, a known location parameter. In hypothesis testing problems as in Randles (2000), this parameter is specified by the null hypothesis. For other inference situations, this issue may be eliminated by using a symmetrized version of

$\boldsymbol{C}_0(F)$ given by Dümbgen (1998): $\boldsymbol{C}_{0s}(F) = \boldsymbol{C}_0(F \ominus F)$, where $F \ominus F$ denotes the distribution of $\boldsymbol{X}_1 - \boldsymbol{X}_2$ for two independent observations $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ on F.

Robustness of $\boldsymbol{C}_0(\mathbb{X}_n)$ and $\boldsymbol{C}_{0s}(\mathbb{X}_n)$ is of interest more generally than under an elliptical symmetry assumption. Breakdown points for both estimators are treated in Dümbgen and Tyler (2005), with the following results. For $\boldsymbol{C}_0(\mathbb{X}_n)$ the finite sample breakdown point (BP) is $(\lceil n/d \rceil - 1)/n \sim 1/d$, this limit being the maximum possible for multivariate M-estimates of scatter. For $\boldsymbol{C}_{0s}(\mathbb{X}_n)$, a BP of $1 - \sqrt{1 - 1/d}$, taking values between $1/(2d)$ and $1/d$, is obtained for a contamination model. Dümbgen and Tyler (2005) also discuss in detail the nature of the contamination that can cause breakdown of these estimators.

High breakdown point scatter estimators. Of course, estimators with much higher breakdown points are often desired. The Minimum Covariance Determinant (MCD) estimator of Rousseeuw (1985), attains a BP of $\lfloor (n - d + 1)/2 \rfloor /n \sim 1/2$. Computational efficiency is sacrificed, however, to gain such a high BP while retaining full affine covariance equivariance. On the other hand, "Fast-MCD", a fast algorithm constructed by Rousseeuw and Van Driessen (1999), approximates the MCD and is implemented in the R packages *MASS*, *rrcov*, and *robustbase*, for example, as well as in other software packages. Other well-known covariance functionals also attain the preceding high BP, again at the expense of computational complexity.

3.5 Computation of Sample Versions

By (25), once $\boldsymbol{C}_n(\mathbb{X}_n)$ is computed for a data set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ (which may or not be intensive, depending on the choice), computation of the $\boldsymbol{u}$th sample Mahalanobis quantile $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$ may be carried out in straightforward fashion by solving

$$-\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{S}\left(\boldsymbol{C}_n(\mathbb{X}_n)^{-1/2}(\boldsymbol{x} - \boldsymbol{X}_i)\right) = \boldsymbol{u}. \tag{37}$$

See Chaudhuri (1996) for discussion with respect to spatial quantiles and Chakraborty (2001) with respect to TR spatial quantiles. By comparison, some notions of multivariate quantiles have computationally intensive sample versions.

4 Further Results and Remarks

4.1 Additional Features and Properties of $\boldsymbol{Q}_M(\cdot, F_{\boldsymbol{X}})$

Equivalent depth-based expression for $\boldsymbol{Q}_M(\cdot, F_{\boldsymbol{X}})$

As seen in Section 3.1, $\boldsymbol{Q}_M(\cdot, F_{\boldsymbol{X}})$ generates an associated depth function via

$$\begin{aligned} \boldsymbol{Q}_M(\cdot, F_{\boldsymbol{X}}) \;&\longrightarrow\; \boldsymbol{R}_M(\boldsymbol{x}, F_{\boldsymbol{X}}) \longrightarrow O_M(\boldsymbol{x}, F_{\boldsymbol{X}}) \\ &\longrightarrow\; D_M(\boldsymbol{x}, F_{\boldsymbol{X}}) = 1 - \left\| E\left\{ \boldsymbol{S}\left(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{x} - \boldsymbol{X})\right)\right\}\right\|. \end{aligned}$$

On the other hand, as seen in Example 2.1, $D_M(\cdot, F_{\boldsymbol{X}})$ itself generates an associated quantile function, say $\widetilde{\boldsymbol{Q}}_M(\cdot, F_{\boldsymbol{X}})$, where $\widetilde{\boldsymbol{Q}}_M(\tilde{\boldsymbol{u}}, F_{\boldsymbol{X}})$ is the quantile representation of a point $\boldsymbol{x}$ lying on the boundary of some depth-based "central region" $\{\boldsymbol{x} : D_M(\boldsymbol{x}, F_{\boldsymbol{X}}) \geq \alpha\}$ and in the direction $\boldsymbol{u}$ from the median $\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{Q}_M(\boldsymbol{0}, F_{\boldsymbol{X}}) = \widetilde{\boldsymbol{Q}}_M(\boldsymbol{0}, F_{\boldsymbol{X}})$. Here $\tilde{\boldsymbol{u}} = p\boldsymbol{v}$, where p is the probability weight of the central region with $\boldsymbol{x}$ on its boundary and $\boldsymbol{v}$ is the unit vector toward $\boldsymbol{x}$ from $\boldsymbol{M}_{\boldsymbol{X}}$. For this version of Mahalanobis quantile function, the outlyingness parameter $\|\tilde{\boldsymbol{u}}\|$ is the probability weight of the central region with $\boldsymbol{x} = \widetilde{\boldsymbol{Q}}_M(\tilde{\boldsymbol{u}}, F_{\boldsymbol{X}})$ on its boundary. While $\boldsymbol{Q}_M(\cdot, F_{\boldsymbol{X}})$ and $\widetilde{\boldsymbol{Q}}_M(\cdot, F_{\boldsymbol{X}})$ generate the same contours, but indexed differently, the correspondence

$$\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) = \widetilde{\boldsymbol{Q}}_M(\tilde{\boldsymbol{u}}, F_{\boldsymbol{X}})$$

is nontrivial to characterize explicitly.

Central regions and re-indexing by probability weight

We call

$$C_M(r, F) = \{\boldsymbol{Q}_M(\boldsymbol{u}, F) : \|\boldsymbol{u}\| \leq r\}$$

the rth *central region* corresponding to $\boldsymbol{Q}_M(\boldsymbol{u}, F)$. Affine equivariance of $C(r, F)$ follows from that of $\boldsymbol{Q}_M(\boldsymbol{u}, F)$. As noted above, the central regions can equivalently be indexed by their probability weights, since these regions are ordered and increase with respect to the "outlyingness" parameter r that describes their boundaries, i.e., $r < r'$ implies $C_M(r, F) \subset C_M(r', F)$ and the respective probability weights p_r increase with r. The correspondence

$$\{\boldsymbol{Q}_M(\boldsymbol{u}, F) : \|\boldsymbol{u}\| \leq r\} = \{\widetilde{\boldsymbol{Q}}_M(\tilde{\boldsymbol{u}}, F) : \|\tilde{\boldsymbol{u}}\| \leq p_r\}$$

may be described by a mapping $\psi_F : r \mapsto p_r \in [0, 1)$, with inverse $\psi_F^{-1} : p \mapsto r_p$ (thus $p_r = \psi_F(r)$ and $r_p = \psi_F^{-1}(p)$), but characterization of this mapping is complicated.

Volume functionals and scale curves

A (real-valued) *volume functional* corresponding to $\boldsymbol{Q}_M(\boldsymbol{u}, F)$ is defined by

$$v_M(r, F) = \text{volume}\,(C_M(r, F)),\ \ 0 \leq r < 1.$$

For each r, $v_M(r, F)$ provides a dispersion measure, as noted by Chaudhuri (1996) for the spatial quantile function. It is affine invariant, and $v_M(r, F)^{1/d}$ is affine equivariant. As an increasing function of the variable r, $v_M(r, F)$ characterizes the dispersion of F in terms of expansion of the central regions $C_M(r, F)$.

Analogous to the scale curve introduced by Liu, Parelius, and Singh (1999) in connection with *depth-based* central regions indexed by their *probability weight*, the volume functional $v_M(r, F)$ may likewise be plotted as a "scale curve" over $0 \leq r < 1$, thus providing a convenient two-dimensional device for viewing or comparing multivariate distributions of any dimension. Illustrations of the depth-based scale curves are included in Liu, Parelius, and Singh (1999) and, for elliptical distributions along with detailed elucidation and inference approaches, in Hettmansperger, Oja, and Visuri (1999). The latter suggest and illustrate in the bivariate case a *PP*-plot of the empirical cdf's of the elliptical areas determined by the data in each sample. These ideas may be exploited for the Mahalanobis scale curve as well. Alternatively, two multivariate distributions F and G may be compared via a *spread-spread plot*, the graph of $v_M(\cdot, G)\, v_M(\cdot, F)^{-1}$, as introduced for the univariate case in Balanda and MacGillivray (1990). Besides having intrinsic appeal as just described, such volume functionals play key roles in defining skewness and kurtosis measures, as discussed in Serfling (2004).

Skew-symmetry of $\boldsymbol{Q}_M(\cdot, F_{\boldsymbol{X}})$ for F centrally symmetric

Let F be *centrally symmetric* about $\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{Q}_M(\boldsymbol{0}, F_{\boldsymbol{X}})$, that is, $\boldsymbol{X} - \boldsymbol{M}_{\boldsymbol{X}}$ and $\boldsymbol{M}_{\boldsymbol{X}} - \boldsymbol{X}$ are identically distributed. Then the Mahalanobis quantile function $\boldsymbol{Q}_M(\cdot, F_{\boldsymbol{X}})$ is *skew-symmetric*:

$$\boldsymbol{Q}_M(-\boldsymbol{u}, F_{\boldsymbol{X}}) - \boldsymbol{M}_{\boldsymbol{X}} = -[\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) - \boldsymbol{M}_{\boldsymbol{X}}],\ \boldsymbol{u} \in \mathbb{B}^{d-1}(\boldsymbol{0}). \tag{38}$$

Proof. We apply the equivariance (27) with $\boldsymbol{A} = -\boldsymbol{I}_d = \widetilde{\boldsymbol{A}}$. Thus

$$\begin{aligned}
\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) - \boldsymbol{M}_{\boldsymbol{X}} &= \boldsymbol{Q}_M\left(\boldsymbol{u}, F_{\boldsymbol{X}-\boldsymbol{M}_{\boldsymbol{X}}}\right) \\
&= -\boldsymbol{Q}_M\left(-\boldsymbol{u}, F_{\boldsymbol{M}_{\boldsymbol{X}}-\boldsymbol{X}}\right) \\
&= -\boldsymbol{Q}_M\left(-\boldsymbol{u}, F_{\boldsymbol{X}-\boldsymbol{M}_{\boldsymbol{X}}}\right) \\
&= -[\boldsymbol{Q}_M(-\boldsymbol{u}, F_{\boldsymbol{X}}) - \boldsymbol{M}_{\boldsymbol{X}}].
\end{aligned}$$

□

For the spatial quantile function, this result is derived in Koltchinskii (1997) p. 448 (see also Serfling, 2004).

When F is centrally symmetric, the above skew-symmetry yields that the central regions $C_M(r, F)$ have the nice property of being *symmetric* sets, in the sense that for each point $\boldsymbol{x}$ in $C_M(r, F)$ its reflection about the point of symmetry $\boldsymbol{M}_F$ is also in $C(r, F)$.

Mahalanobis rank covariance matrix

A useful tool based on the Mahalanobis centered rank function is the corresponding *rank covariance matrix* (RCM) of $F_{\boldsymbol{X}}$,

$$E_F\left(\boldsymbol{R}_M(\boldsymbol{X}, F_{\boldsymbol{X}})\boldsymbol{R}_M(\boldsymbol{X}, F_{\boldsymbol{X}})'\right),$$

that is, the covariance matrix of the *rank* of $\boldsymbol{X}$ rather than of $\boldsymbol{X}$ itself. While the covariance of $\boldsymbol{X}$ has already been addressed in some sense via $\boldsymbol{C}(F)$, the *Mahalanobis RCM* permits reduction to rank methods while retaining full affine equivariance. The Mahalanobis RCM of $\boldsymbol{X}$ is, of course, just the *spatial* RCM of $\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}$. Rank covariance matrices as robust alternatives to the classical covariance matrix are applied in a variety of ways (see Visuri, Koivunen, and Oja, 2000, and Visuri, Ollila, Koivunen, Möttönen, and Oja, 2003, for example).

4.2 Asymptotic Behavior of Sample Versions

Results for the sample Mahalanobis median

For $\boldsymbol{C}(F)$ given by the usual covariance $\boldsymbol{\Sigma}(F)$, and assuming that $\boldsymbol{Q}_M(\boldsymbol{0}, F_{\boldsymbol{X}})$ uniquely minimizes $E\|\boldsymbol{\Sigma}(F)^{-1/2}(\boldsymbol{X} - \boldsymbol{\theta})\|$ with $\boldsymbol{\theta}$ confined to a compact domain within $\mathbb{R}^d$, Isogai (1985) establishes *convergence in probability* of $\boldsymbol{Q}_M(\boldsymbol{0}, \mathbb{X}_n)$ to $\boldsymbol{Q}_M(\boldsymbol{0}, F_{\boldsymbol{X}})$. In fact, a more general objective function $E\psi(\|\boldsymbol{\Sigma}(F)^{-1/2}(\boldsymbol{X}-\boldsymbol{\theta})\|)$ is permitted, for ψ strictly increasing and satisfying certain regularity conditions. Under additional regularity conditions and following standard methods for M-estimators, *asymptotic normality* is also established. These results easily extend to the general case of weak covariance functional $\boldsymbol{C}(F)$.

Results for a vector of several sample Mahalanobis quantiles

The above approach extends in straightforward fashion to obtain the joint consistency and asymptotic normality of $\boldsymbol{Q}_M(\boldsymbol{u}_1, F_{\boldsymbol{X}}), \ldots, \boldsymbol{Q}_M(\boldsymbol{u}_m, F_{\boldsymbol{X}})$. However, we look at this below in the more encompassing and illuminating context of a Bahadur-Kiefer represenation for the sample Mahalanobis quantile function.

Bahadur-Kiefer representation for the sample Mahalanobis quantile function

The sample spatial quantile function. The classical Bahadur (1966) representation for univariate sample quantiles is extended to multivariate sample *spatial* quantiles $\boldsymbol{Q}_S(\boldsymbol{u}, \mathbb{X}_n)$ pointwise by Chaudhuri (1996) and uniformly by Koltchinskii (1994a,b). In these results the estimation error $\boldsymbol{Q}_S(\boldsymbol{u}, \mathbb{X}_n) - \boldsymbol{Q}_S(\boldsymbol{u}, F)$ is approximated by a convenient *linear term* with a suitably negligible remainder $\boldsymbol{R}_n(\boldsymbol{u})$, as follows:

$$\begin{aligned}\boldsymbol{Q}_S(\boldsymbol{u}, \mathbb{X}_n) &- \boldsymbol{Q}_S(\boldsymbol{u}, F) \\ &= -[\boldsymbol{D}_1(\boldsymbol{Q}_S(\boldsymbol{u}, F))]^{-1}\frac{1}{n}\sum[\boldsymbol{S}(\boldsymbol{Q}_S(\boldsymbol{u}, F) - \boldsymbol{X}_i) - \boldsymbol{u}] + \boldsymbol{R}_n(\boldsymbol{u}), \qquad (39)\end{aligned}$$

where, with $\partial \boldsymbol{v}(\boldsymbol{x})/\partial \boldsymbol{x}$ the matrix derivative of a vector $\boldsymbol{v}(\boldsymbol{x})$,

$$\begin{aligned}
\boldsymbol{D}_1(\boldsymbol{x}) &= E\left\{\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{S}(\boldsymbol{x}-\boldsymbol{X})\right\} \\
&= E\left\{\frac{1}{\|\boldsymbol{x}-\boldsymbol{X}\|}\left[\boldsymbol{I}_d - \frac{1}{\|\boldsymbol{x}-\boldsymbol{X}\|^2}(\boldsymbol{x}-\boldsymbol{X})(\boldsymbol{x}-\boldsymbol{X})'\right]\right\} \\
&= E\{\boldsymbol{D}_2(\boldsymbol{x}-\boldsymbol{X})\}, \qquad (40)
\end{aligned}$$

with $\boldsymbol{D}_2(\boldsymbol{x})$ the $d \times d$ Hessian or second order derivative of the function $\|\boldsymbol{x}\|$ (the gradient or first order derivative being the sign function $\boldsymbol{S}(\boldsymbol{x})$), i.e.,

$$\boldsymbol{D}_2(\boldsymbol{x}) = \left\{\frac{1}{\|\boldsymbol{x}\|}\left[\boldsymbol{I}_d - \frac{1}{\|\boldsymbol{x}\|^2}\boldsymbol{x}\boldsymbol{x}'\right]\right\}.$$

Note that $\boldsymbol{D}_1(\boldsymbol{x})$ is $d \times d$ symmetric and, unless F is supported by a straight line in $\mathbb{R}^d$, positive definite. (For details, see Chaudhuri, 1992, Lemma 5.3.)

The *Bahadur-Kiefer representation for sample spatial quantiles* asserts an almost sure uniform convergence rate for the remainder term $\boldsymbol{R}_n(\boldsymbol{u})$ in (39). Let the inverse of $\boldsymbol{Q}_S(\boldsymbol{u}, F)$, i.e., the spatial centered rank function $\boldsymbol{R}_S(\boldsymbol{x}, F)$, be continuously differentiable in an open set $\mathbb{V}$ in $\mathbb{R}^d$ with $\boldsymbol{D}_1(\boldsymbol{x})$ locally Lipschitz in $\mathbb{V}$. For any compact $K \subset \boldsymbol{R}_S(\mathbb{V}, F) \subset \mathbb{B}^{d-1}(\boldsymbol{0})$, define

$$\Delta_n(K) = \sup_{\boldsymbol{u} \in K} \|\boldsymbol{R}_n(\boldsymbol{u})\|.$$

Then, under twice continuous differentiability of $\boldsymbol{R}_S(\boldsymbol{x}, F)$ in $\mathbb{V}$, we have almost surely

$$\Delta_n(K) = \begin{cases} O(n^{-1}\log n) & \text{for } d \geq 3, \\ O(n^{-1}(\log n)^{3/2}) & \text{for } d = 2, \\ O(n^{-3/4}\log n) & \text{for } d = 1. \end{cases} \qquad (41)$$

Further details, expressed in the above formulation, are available in Zhou and Serfling (2008), where also extension to sample spatial *U-quantiles* is developed. (These are the sample spatial quantiles of the set of evaluations of a *vector-valued* kernel $\boldsymbol{h}(x_1, \ldots, x_m)$ over a sample of observations $X_1, \ldots, X_m$ in some space $\mathcal{X}$.)

The sample Mahalanobis quantile function. Let us now consider extension of the above result to the sample *Mahalanobis* quantile function. Unfortunately, a quick approach by direct use of the TR representation (22) in conjunction with (39) leads to a result just short of being sharp enough for practical applications. Rather, one needs to define an appropriate analogue of (39) and follow the same steps of proof (omitted here – see Koltchinskii, 1994a,b, or Zhou and Serfling, 2008). For this purpose, the estimation error $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n) - \boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}})$ is approximated as follows:

$$\begin{aligned}
&\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n) - \boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) \\
&\quad = -\boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}[\widetilde{\boldsymbol{D}}_1(\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}))]^{-1}\frac{1}{n}\sum\left[\boldsymbol{S}(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) - \boldsymbol{X}_i)) - \boldsymbol{u}\right] \\
&\qquad + \widetilde{\boldsymbol{R}}_n(\boldsymbol{u}), \qquad (42)
\end{aligned}$$

where instead of (40) we use

$$\widetilde{\boldsymbol{D}}_1(\boldsymbol{x}) = E\left\{\boldsymbol{D}_2(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{x}-\boldsymbol{X}))\right\}.$$

Let $\boldsymbol{R}_M(\boldsymbol{x}, F_{\boldsymbol{X}})$ $(= \boldsymbol{R}_S(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{x}, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}))$ be continuously differentiable in an open set $\mathbb{V}$ in $\mathbb{R}^d$ with $\widetilde{\boldsymbol{D}}_1(\boldsymbol{x})$ locally Lipschitz in $\mathbb{V}$. The following result gives the *Bahadur-Kiefer representation for sample Mahalanobis quantiles*

Lemma 4.1 *For any compact $K \subset \boldsymbol{R}_M(\mathbb{V}, F) \subset \mathbb{B}^{d-1}(\boldsymbol{0})$, with $\widetilde{\boldsymbol{R}}_n(\boldsymbol{u})$ defined by (42),*

$$\widetilde{\Delta}_n(K) = \sup_{\boldsymbol{u}\in K} \|\widetilde{\boldsymbol{R}}_n(\boldsymbol{u})\|$$

satisfies, almost surely, (41) with $\Delta_n(K)$ replaced by $\widetilde{\Delta}_n(K)$.

Following the treatment of Zhou and Serfling (2008), extension of this result to *sample Mahalanobis U-quantiles* is straightforward.

In view of the connection with TR spatial quantiles discussed in Section 3.1, the above result may be interpreted as a Bahadur-Kiefer representation for sample TR spatial quantiles, given any selection of weak covariance functional $\boldsymbol{C}(\cdot)$. Chakraborty (2001) provides, for a particular TR functional and a different indexing (as already mentioned in Section 3.1), the (pointwise) Bahadur representation implied by Lemma 4.1, with a slightly weaker rate in the case $d = 2$. We note, however, that Chakraborty (2001) also treats the interesting more general setting of TR ℓ_p quantiles for $p \geq 1$, within which the TR spatial quantiles are the case $p = 2$.

Applications of Lemma 4.1. Note that the leading right-hand term in (42) serves as a uniform approximation to the estimation error $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n) - \boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}})$ and is a *sample mean* in structure. Through this feature, Lemma 4.1 is quite useful in several general lines of application:

- *Joint asymptotic normality of $\boldsymbol{Q}_M(\boldsymbol{u}_1, \mathbb{X}_n), \ldots, \boldsymbol{Q}_M(\boldsymbol{u}_m, \mathbb{X}_n)$.* This follows by straightforward application of the asymptotic multivariate normality of a vector sample mean.

- *Linkage between estimators and test statistics.* A natural test statistic for the null hypothesis that $F_{\boldsymbol{X}}$ has Mahalanobis median $\boldsymbol{\theta}_0$ (i.e., that $Q_M(\boldsymbol{0}, F) = \boldsymbol{\theta}_0$) is given by the leading right-hand term in (42) evaluated at $\boldsymbol{u} = \boldsymbol{0}$ and with $\boldsymbol{C}_n(\mathbb{X}_n)$ substituted for $\boldsymbol{C}(F_{\boldsymbol{X}})$, i.e., (equivalently) by

$$T = \sum_{i=1}^{n} \boldsymbol{S}(\boldsymbol{C}_n(\mathbb{X}_n)^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta}_0)),$$

  to which classical multivariate central limit theory may be applied. (This is a particular multivariate generalization of the univariate *sign test.*) Equivalently, T is a test statistic for the hypothesis that the *spatial* median of $F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}$ is $\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{\theta}_0$. Through the linkage (42) between the *test statistic* T for the Mahalanobis median and the *estimator* $\boldsymbol{Q}_M(\boldsymbol{0}, \mathbb{X}_n)$ of that parameter, the performance characteristics of these two statistics are equivalent.

- *Almost sure convergence and law of iterated logarithm for* $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$. Through (42) and Lemma 4.1, the almost sure behavior of $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$ follows by classical results.
- *The influence function of* $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$. The leading right-hand term in (42) also serves as a (vector-valued) type of influence function for $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$:

  $$\begin{aligned}&\mathrm{IF}(\boldsymbol{x}; \boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}), F_{\boldsymbol{X}}) = \\ &\quad -\boldsymbol{C}(F_{\boldsymbol{X}})^{1/2}[\widetilde{\boldsymbol{D}}_1(\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}))]^{-1}\left[\boldsymbol{S}(\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}) - \boldsymbol{x})) - \boldsymbol{u}\right], \ \boldsymbol{x} \in \mathbb{R}^d.\end{aligned}$$

  Favorably, $\|\mathrm{IF}(\cdot; \boldsymbol{Q}_M(\boldsymbol{u}, F_{\boldsymbol{X}}), F_{\boldsymbol{X}})\|$ is *bounded*.

4.3 Determination of F, given choice of $\boldsymbol{C}(\cdot)$

An important property of the spatial quantile function $\boldsymbol{Q}(\cdot, F)$ is that it characterizes the associated distribution: $Q_S(\cdot, F) = Q_S(\cdot, G)$ implies $F = G$ (Koltchinskii, 1997, Cor. 2.9). Consequently, for any choice of weak covariance functional $\boldsymbol{C}(\cdot)$, the spatial quantile function determines the Mahalanobis quantile function:

$$\begin{aligned}\{\boldsymbol{C}(\cdot), \boldsymbol{Q}_S(\cdot, F_{\boldsymbol{X}})\} &\longrightarrow \{\boldsymbol{C}(\cdot), F_{\boldsymbol{X}}\} \\ &\longrightarrow \left\{\boldsymbol{C}(F_{\boldsymbol{X}}), F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}\right\} \\ &\longrightarrow \left\{\boldsymbol{C}(F_{\boldsymbol{X}}), \boldsymbol{Q}_S\left(\cdot, F_{\boldsymbol{C}(F_{\boldsymbol{X}})^{-1/2}\boldsymbol{X}}\right)\right\} \\ &\longrightarrow \boldsymbol{Q}_M\left(\cdot, F_{\boldsymbol{X}}\right).\end{aligned}$$

The reverse path does not hold, however. This shows the *fundamental importance of the spatial quantile function*. Despite a lack of full affine equivariance, it contains all of the "information" about F, whereas the Mahalanobis quantile function does not (and need not).

4.4 Breakdown Point of $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$

As is well-known, the finite sample breakdown point (BP) of the sample *spatial* median $\boldsymbol{Q}_S(\boldsymbol{0}, \mathbb{X}_n)$ is 1/2 (Kemperman, 1987). It is not difficult, using this fact, to obtain for $\boldsymbol{Q}_S(\boldsymbol{u}, \mathbb{X}_n)$ the BP of $(1 - \|\boldsymbol{u}\|)/2$. It is then straightforward to obtain the BP for the sample *Mahalanobis* median $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$, giving the following result.

Lemma 4.2 *For the sample Mahalanobis median* $\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)$ *based on the weak covariance functional* $\boldsymbol{C}(\cdot)$*, we have*

$$\mathrm{BP}(\boldsymbol{Q}_M(\boldsymbol{u}, \mathbb{X}_n)) = \min\left\{\mathrm{BP}(\boldsymbol{C}_n(\mathbb{X}_n)), \frac{1 - \|\boldsymbol{u}\|}{2}\right\}.$$

Choices of $\boldsymbol{C}(\cdot)$ with high BP have been discussed in Section 3.4. We note from (37) an attractive qualitative robustness property of the Mahalanobis quantile $Q_M(\boldsymbol{u}, \mathbb{X}_n)$: its value remains unchanged if the points $\boldsymbol{X}_i$ are moved outward along the rays joining them with $Q_M(\boldsymbol{u}, \mathbb{X}_n)$.

4.5 Variations on Mahalanobis Outlyingness, Depth, and Quantile Functions

The "usual" Mahalanobis approach

Let us express in terms of a general weak covariance function the popular Mahalanobis outlyingness function based on (1) and normalized to take values in $(0,1)$:

$$O_1(\boldsymbol{x}, F) = \frac{\|\boldsymbol{C}(F)^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}(F))\|}{1 + \|\boldsymbol{C}(F)^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}(F))\|}, \ \boldsymbol{x} \in \mathbb{R}^d. \tag{43}$$

To distinguish from the *Mahalanobis quantile outlyingness* $O_M(\boldsymbol{x}, F)$ that we have introduced in association with $\boldsymbol{Q}(\boldsymbol{u}, F)$, let us call this the *Mahalanobis distance outlyingness*. As mentioned earlier, it has favorable mathematical tractability and intuitive appeal. However, as noted earlier in Section 1, its contours of equal outlyingness are necessarily *ellipsoidal*. The same is true for the contours of the associated depth and quantile functions defined as in Example 2.1.

When the distribution F may be assumed elliptically symmetric, this approach using robust $\boldsymbol{C}_n(\mathbb{X}_n)$ and $\boldsymbol{\mu}_n(\mathbb{X}_n)$ is certainly as strong as any. One compelling reason is that $O_1(\boldsymbol{x}, F)$ and $\boldsymbol{Q}_1(\boldsymbol{u}, F)$ have breakdown points

$$\min\{\mathrm{BP}(\boldsymbol{C}_n(\mathbb{X}_n)), \mathrm{BP}(\boldsymbol{\mu}_n(\mathbb{X}_n)\}$$

determined just by $\boldsymbol{C}_n(\mathbb{X}_n)$ and $\boldsymbol{\mu}_n(\mathbb{X}_n)$.

The Mahalanobis approach of the present paper

If, on the other hand, an assumption of elliptical symmetry is not justified, then the approach using $\boldsymbol{Q}_M(\boldsymbol{u}, F)$ (or the TR spatial quantile function) becomes an attractive possibility. The contours need not be elliptical, the "center" is well-defined in a "Mahalanobis way" as the point minimizing expected Mahalanobis distance, mathematical tractability and intuitive appeal remain strong, and affine equivariance holds. Also, the sample quantile function is computationally easy (Section 3.5) and satisfies a convenient asymptotic theory (Section 4.3). On the other hand, the BP suffers somewhat, depending not only on $\boldsymbol{C}_n(\mathbb{X}_n)$ but also on the "threshold" of the quantile or outlyingness. This is understandable, however, because the quantile function $\boldsymbol{Q}_1(\cdot, F)$ is completely determined by the parameters $\boldsymbol{C}(F)$ and $\boldsymbol{\mu}(F)$, whereas the quantile function $\boldsymbol{Q}_M(\cdot, F)$ involves not only $\boldsymbol{C}(F)$ but also the spatial quantile function $\boldsymbol{Q}_S(\cdot, F)$ as an infinite dimensional parameter. This makes the estimation problem challenging and more susceptible to breakdown. The assumption of elliptical symmetry is avoided but at a price. Let us call $O_M(\boldsymbol{x}, F)$

Still another Mahalanobis approach

A variation on $O_1(\boldsymbol{x}, F)$ above is

$$O_2(\boldsymbol{x}, F) = \frac{E\{\|\boldsymbol{C}(F)^{-1/2}(\boldsymbol{x} - \boldsymbol{X})\|\}}{1 + E\{\|\boldsymbol{C}(F)^{-1/2}(\boldsymbol{x} - \boldsymbol{X})\|\}}, \ \boldsymbol{x} \in \mathbb{R}^d. \quad (44)$$

This outlyingness function corresponds to a depth function listed among Type B versions in Zuo and Serfling (2000). It is affine invariant and the associated quantile function $\boldsymbol{Q}_2(\cdot, F)$ affine equivariant, and the contours need not be elliptical.

Clearly, here the associated median (minimizing outlyingness) is the same as $\boldsymbol{Q}_M(\boldsymbol{0}, F)$, the Mahalanobis median of Isogai (1985) and of the present paper. Otherwise, however, $\boldsymbol{Q}_M(\boldsymbol{u}, F)$ and $\boldsymbol{Q}_2(\boldsymbol{u}, F)$ do not agree. Indeed, $\boldsymbol{Q}_2(\boldsymbol{u}, F)$ seems worthy of investigation.

4.6 Brief comments on Applications

An extensive treatment of applications of the Mahalanobis quantile function is beyond the scope of the present paper, whose main purposes are a theoretical study of this tool and its properties and a rigorous clarification of the so-called TR spatial quantile approach. Along the way we have indicated a few applications, and further possibilities may be explored through a number of general sources. Applications of depth-based approaches are described in detail in Liu, Parelius, and Singh (1999). Chakraborty (2001) provides several techniques using the TR sample spatial quantile approach. Serfling (2004) treats further applications of the spatial quantile function, and these can be emulated using the Mahalanobis quantile function.

Acknowledgment

The author thanks Xin Dang very appreciatively for pointing out a significant error in a previous version. Also, support under National Science Foundation Grants DMS-0103698 and DMS-0805786 and NSA Grant H98230-08-1-0106 is gratefully acknowledged.

References

[1] Breckling, J. and Chambers, R. (1988). *M*-quantiles. *Biometrika* **75** 761–771.

[2] Breckling, J., Kokic, P. and Lübke, O. (2001). A note on multivariate *M*-quantiles. *Statistics & Probability Letters* **55** 39–44.

[3] Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Annals of the Institute of Statistical Mathematics* **53** 380–403.

[4] Chakraborty, B., Chaudhuri, P., and Oja, H. (1998). Operating transformation and retransformation 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. (1996. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association* **91** 862–872.

[7] Dudley, R. M. and Koltchinskii, V. I. (1992). The spatial quantiles. Unpublished manuscript.

[8] Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Annals of the Institute of Statistical Mathematics* **50** 471–491.

[9] Dümbgen, L. and Tyler, D. E. (2005). On the breakdown properties of some multivariate M-functionals. *Scandinavian Journal of Statistics* **32** 247–264.

[10] Hettmansperger, T. P., Oja, H. and Visuri, S. (1999). Discussion of Liu, Parelius and Singh (1999). *Annals of Statistics* **27** 845–854.

[11] Isogai, T. (1985). Some extension of Haldane's multivariate median and its application. *Annals of Institute of Statistical Mathematics* **37** 289–301.

[12] Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical Data Analysis Based On the L_1-Norm and Related Methods* (Y. Dodge, ed.), pp. 217–230, North-Holland.

[13] Koltchinskii, V. (1994a). Bahadur-Kiefer approximation for spatial quantiles. In: Hoffman-Jørgensen, J., Kuelbs, J., Marcus, M. (Eds.), *Probability in Banach Spaces 9*, Birkhaüser, 401–415.

[14] Koltchinskii, V. (1994b). Nonlinear transformations of empirical processes: functional inverses and Bahadur-Kiefer representations. In: *Proceedings of 6th International Vilnius Conference on Probability and Mathematical Statistics*, VSP-TEV, Netherlands, 423–445.

[15] Koltchinskii, V. (1997). M-estimation, convexity and quantiles. *Annals of Statistics* **25** 435–477.

[16] Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Annals of Statistics* **27** 783–858.

[17] Lopuhaä, H. P. and Rousseeuw, J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Annals of Statistics* **19** 229-248.

[18] Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India* **12** 49–55.

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

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

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

[22] Rousseeuw P. (1985). Multivariate estimation with high breakdown point. In: W. Grossmann, G. Pflug, and W. Wertz (Eds.), *Mathematical Statistics and Applications, Vol. B*, Reidel Publishing, Dordrecht, 283–297.

[23] Rousseeuw, P. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41** 212–223.

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

[25] Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference* **123** 259–278.

[26] Small, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canadian Journal of Statistics* **15** 31–39.

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

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

[29] Visuri, S., Koivunen, V., and Oja, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference* **91** 557–575.

[30] Visuri, E., Ollila, V., Koivunen, J., Möttönen, and Oja, H. (2003). Affine equivariant multivariate rank methods. *Journal of Statistical Planning and Inference* **114** 161–185.

[31] Zhou, W. and Serfling, R. (2008). Multivariate spatial U-quantiles: a Bahadur-Kiefer representation, a Theil-Sen estimator for multiple regression, and a robust dispersion estimator. *Journal of Statistical Planning and Inference* **138** 1660–1678.

[32] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics* **31** 1460–1490.