On the Breakdown Properties of Some Multivariate M-Functionals

LUTZ DÜMBGEN
Department of Statistics
University of Bern

DAVID E. TYLER
Department of Statistics
Rutgers, The State University of New Jersey

Running title: *Breakdown of Multivariate M-Functionals*

ABSTRACT. For probability distributions on $\mathbb{R}^q$, a detailed study of the breakdown properties of some multivariate M-functionals related to Tyler's (1987a) "distribution-free" M-functional of scatter is given. These include a symmetrized version of Tyler's M-functional of scatter, and the multivariate t M-functionals of location and scatter. It is shown that for "smooth" distributions, the (contamination) breakdown point of Tyler's M-functional of scatter and of its symmetrized version are $1/q$ and $1 - \sqrt{1 - 1/q}$, respectively. For the multivariate t M-functional which arises from the maximum likelihood estimate for the parameters of an elliptical t distribution on $\nu \geq 1$ degrees of freedom the breakdown point at smooth distributions is $1/(q + \nu)$. Breakdown points are also obtained for general distributions, including empirical distributions. Finally, the sources of breakdown are investigated. It turns out that breakdown can only be caused by contaminating distributions that are concentrated near low-dimensional subspaces.

Keywords and phrases: breakdown, coplanar contamination, M-estimates, M-functionals, scatter matrix, symmetrization, t-distributions, tight contamination.

1 Introduction

Affine equivariant M-estimates of multivariate location and scatter were first proposed by Maronna (1976) as robust alternatives to the sample mean vector and covariance matrix. One critical feature of these estimates, which was noted by Maronna (1976), is their relatively low breakdown point in higher dimensions. Maronna (1976) obtains upper bounds for the breakdown points of the M-estimates and notes that none have a breakdown point greater than $1/(q + 1)$, where q represents the dimension of the data. Stahel (1981) obtains a general bound of $1/q$ for a slightly more general class of M-estimates.

Subsequently, affine equivariant high breakdown point estimates of multivariate location and scatter have been introduced, i.e. estimates with breakdown points near $1/2$ regardless of q. These include projection-based estimates, cf. Stahel (1981), Donoho (1982), Maronna, Stahel & Yohai (1992) and Tyler (1994), the minimum volume ellipsoid and the minimum covariance determinant estimates, cf. Rousseeuw (1986), S-estimates, cf. Davies (1987), constrained M-estimates, cf. Kent & Tyler (1996), and multivariate MM-estimates, cf. Tatsuoka & Tyler (2002) and Tyler (2003).

All of the known high breakdown point estimates are computationally intensive, and only approximate algorithms for small values of q are feasible. On the other hand, the multivariate M-estimates are computationally feasible even for very large values of q since they can be formulated in terms of convex optimization problems. In particular, they can be readily calculated via simple reweighting algorithms. Moreover, for small values of q, the upper bound of $1/q$ is not unreasonably low for many applications.

Aside from the upper bounds reported by Maronna (1976) and Stahel (1981), little work has been published on the breakdown points of the multivariate M-estimates. A more detailed study of their breakdown properties is warranted. In this paper, we study the breakdown point problem of some M-estimates related to Tyler's (1987a) "distribution-free" M-estimate of scatter. These include the multivariate t M-estimates of location and scatter. Since the breakdown point concept represents a worst case scenario, one may gain further insight into the properties of an estimate by also understanding what causes an estimate to break down as well as the behavior of the estimate outside of its worst case. We address these issues for the M-estimates considered in this paper.

At first we investigate Tyler's (1987a) M-estimate of scatter. This scatter statistic also arises within the context of directional data, see Tyler (1987b) and Kent & Tyler (1988), and has been independently proposed for "shape from texture" problems in computer vision by Blake & Marinos (1990). It has also been used for standardizing the design matrix for GM-estimates and P-estimates of regression, see Martin, Yohai & Zamar (1989) and Maronna & Yohai (1993), respectively. More recently, it has been applied in the development non-parametric multivariate methods, see e.g. Randles (2000), Hallin & Paindaveine (2002a, 2002b) and Hettmansperger & Randles (2003). We show in section 2 that this M-estimate of scatter obtains the maximal possible breakdown point of $1/q$ for multivariate M-estimates. Moreover, we note that breakdown can only be caused by contaminating distributions that are concentrated near low-dimensional subspaces. In accordance with Tyler (1986) we call this "coplanar contamination". If we restrict attention to contaminating distributions which are not concentrated near subspaces with dimension less than r, then the breakdown point becomes r/q. These results on breakdown follow readily from known results on the existence of the estimate of the scatter matrix.

Tyler's (1987a) M-estimate of scatter presumes a given "center". This presumption can be avoided by using a symmetrized version of the estimate as suggested by Dümbgen (1998). For this symmetrized version we show in section 3 that the breakdown point is $1 - \sqrt{1 - 1/q}$, which lies strictly between $1/(2q)$ and $1/q$. Breakdown in this case is caused by rather specific types of contaminating distributions concentrated near low-dimensional subspaces and at infinity which we refer to as "coplanar contamination at infinity". If this type of contamination is not considered, and we consider only "tight" contamination, then we show that the breakdown point becomes $\sqrt{1/q}$. Thus, the symmetrized M-estimate is less vulnerable to "inliers" than the original M-estimate.

Finally, in section 4, we consider the multivariate t M-estimates of location and scatter. These are the class of M-estimates corresponding to the maximum likelihood estimates derived for the location-scatter class of elliptical t distributions. For integer degrees of freedom, $\nu \geq 1$, we show that the breakdown point equals $1/(q + \nu)$ and examine the cause of breakdown. These are the first results on the exact breakdown point for any simultaneous M-estimate of multivariate location and scatter. The derivation of these breakdown points follows from a simple relationship between the multivariate t M-estimates and Tyler's (1987a) "distribution-free" M-estimate of scatter. Some

3

comments suggesting how one might approach studying in more detail the breakdown points of multivariate M-estimates in general are given at the end of section 4.

Proofs and some technical results are deferred to an appendix.

2 Tyler's M-estimate of scatter

Throughout let P represent a nondegenerate probability distribution on $\mathbb{R}^q$, and let $\boldsymbol{x}$ be a random vector with distribution P. In order to emphasize the latter fact, we sometimes write $\mathcal{L}_P(h(\boldsymbol{x}))$ and $E_P h(\boldsymbol{x})$ for the distribution and expected value, respectively, of any function $h(\boldsymbol{x})$. Tyler (1987a) introduced a "distribution-free" M-estimate of scatter, which in its functional form is defined as a solution $M \in \mathbb{M}^+$ to the equation

$$E_P\Big(\frac{M^{-1/2}\boldsymbol{x}\boldsymbol{x}^\top M^{-1/2}}{\boldsymbol{x}^\top M^{-1}\boldsymbol{x}}\,\Big|\,\boldsymbol{x} \neq 0\Big) \;=\; I. \tag{1}$$

Here $\mathbb{M}^+$ stands for the set of symmetric, positive definite matrices in $\mathbb{R}^{q \times q}$. Note that the left hand side of (1) remains unchanged if M is replaced with cM for any scalar $c > 0$. If there is a unique matrix $M \in \mathbb{M}^+$ satisfying (1) and $\det(M) = 1$, then this solution is denoted by $\Sigma(P)$. Otherwise we set arbitrarily $\Sigma(P) := 0$. Proposition 1 below provides necessary and sufficient conditions on P such that $\Sigma(P) \in \mathbb{M}^+$. The requirement $\det(M) = 1$ could be replaced by other conditions, e.g. $\operatorname{trace}(M) = q$; see also section 2.1.

If P_n represents an empirical probability distribution on $\mathbb{R}^q$, then $\Sigma(P_n)$ is an M-estimate of scatter. Otherwise, we refer to $\Sigma(P)$ as an M-functional of scatter. This M-estimate or M-functional is for scatter only, that is it is defined about a fixed "center", which in the above definition is taken to be the origin. It is useful as long as only the "shape" of $\Sigma(P)$ is important. That means, we are interested only in functions of $\Sigma(P)$ which are invariant under scalar multiplication. This is the case when one in interested in, for example, correlations, multiple correlations, partial correlations, principal component directions, the ratio of principal component roots, or canonical correlations and vectors.

The phrase "distribution-free" arises from the observation that $\Sigma(P)$ depends on $\boldsymbol{x} \sim P$, only through the distribution $\mathcal{L}_P(\boldsymbol{x}/\|\boldsymbol{x}\| \,|\, \boldsymbol{x} \neq 0)$, which lies on the compact unit sphere $\mathbf{S}^{q-1}$ in $\mathbb{R}^q$.

Here $\|\cdot\|$ denotes the Euclidean norm. If P is an elliptically symmetric distribution centered at the origin, i.e. one with density of the form

$$f(x) = \det(\Sigma)^{-1/2} g(x^\top \Sigma^{-1} x), \tag{2}$$

then the distribution of $x/\|x\|$ does not depend on the function g. Consequently, if P_n is the empirical distribution based upon i.i.d. observations from P, then the distribution of $\Sigma(P_n)$ does not depend on the function g. See Tyler (1987a) for details. Further properties of Tyler's M-estimate of scatter have been studied by Maronna & Yohai (1990) and by Adrover (1999). Both papers demonstrate that it has very good bias-robustness properties. The former also shows that its breakdown point at any elliptical distribution under point-mass contamination is $1/q$, while the latter shows that its breakdown point at any elliptical distribution under arbitrary contamination is between $1/(q+1)$ and $1/q$.

A quantity describing the robustness of the functional $\Sigma(\cdot)$ at P is its contamination breakdown point (cf. Huber, 1981). This is defined to be the supremum $\epsilon(P)$ of all $\epsilon \in [0,1]$ such that the maximum "bias" over $\mathcal{U}(P, \epsilon)$, denoted $b(\epsilon; P)$, is finite. Here $\mathcal{U}(P, \epsilon)$ denotes the contamination neighborhood

$$\mathcal{U}(P, \epsilon) := \{(1-\epsilon)P + \epsilon H : H \text{ some distribution on } \mathbb{R}^q\}$$

of P. A linear invariant measure of the maximum bias over $\mathcal{U}(P, \epsilon)$ can be taken to be

$$b(P, \epsilon) := \sup_{Q \in \mathcal{U}(P,\epsilon)} \max\left\{ \lambda_1 \left(\Sigma(P)^{-1} \Sigma(Q) \right), \lambda_q^{-1} \left(\Sigma(P)^{-1} \Sigma(Q) \right) \right\} \tag{3}$$

with $0^{-1} := \infty$, where $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_q(A)$ denote the ordered eigenvalues of a $q \times q$ matrix A with real eigenvalues. Thus breakdown of $\Sigma(\cdot)$ occurs at P if either $\lambda_1(\Sigma(Q))$ can be made arbitrarily large, or $\lambda_q(\Sigma(Q))$ can be made arbitrarily small over $Q \in \mathcal{U}(P, \epsilon)$.

Our results on breakdown depend essentially upon the following existence and continuity properties for $\Sigma(P)$ given in Kent & Tyler (1988) and Dümbgen (1998).

Proposition 1 *Let $\mathcal{V}$ be the set of linear subspaces V of $\mathbb{R}^q$ with $1 \leq \dim(V) < q$, and suppose that $P\{0\} = 0$.*

(a) There is a unique solution $M \in \mathbb{M}^+$ of (1) with $\det(M) = 1$ if, and only if,

$$P(V) < \dim(V)/q \quad \text{for all } V \in \mathcal{V}. \tag{4}$$

5

(b) If (1) holds for some matrix $M \in \mathbb{M}^+$ but $P(V) = \dim(V)/q$ for some space $V \in \mathcal{V}$, then there is a second space $W \in \mathcal{V}$ such that $V \cap W = \{0\}$ and $P(V \cup W) = 1$.

(c) If $P(V) > \dim(V)/q$ for some $V \in \mathcal{V}$, then there is no matrix $M \in \mathbb{M}^+$ satisfying (1).

Proposition 2 *If $P\{0\} = 0$ and Condition (4) holds, then $\Sigma(Q) \to \Sigma(P)$ as $Q \to P$.*

Throughout this paper, convergence of probability distributions is meant to be weak convergence.

If P is "smooth" in the sense that

$$P(V) = 0 \quad \text{for any } V \in \mathcal{V}, \tag{5}$$

then it follows from the above existence and continuity propositions for $\Sigma(P)$ that $\epsilon(P) = 1/q$. To see this, note that if the contaminating distribution H lies in a subspace of dimension one and $H\{0\} = 0$, then by Proposition 1, $\Sigma(Q)$ equals 0 whenever $\epsilon \geq 1/q$. Hence the breakdown point is bounded above by $1/q$. On the other hand, if $\epsilon < 1/q$, then by Proposition 1.a, $\Sigma(Q) \in \mathbb{M}^+$ for all $Q \in \mathcal{U}(P, \epsilon)$. Furthermore, Proposition 2 implies that $b(\epsilon; P)$ must be finite. For $\Sigma(Q)$ depends only on $\mathcal{L}_Q(\boldsymbol{x}/\|\boldsymbol{x}\| \,|\, \boldsymbol{x} \neq 0)$, which has compact support. Thus the corresponding family of distributions $\mathcal{L}_Q(\boldsymbol{x}/\|\boldsymbol{x}\| \,|\, \boldsymbol{x} \neq 0)$ is compact with respect to weak convergence. This entails that the breakdown point is bounded from below by $1/q$, whence $\epsilon(P) = 1/q$.

We now give a general expression for $\epsilon(P)$ which applies to any P, and in particular to empirical distributions. We also investigate the case $\epsilon = \epsilon(P)$ in more detail. For this case it turns out that for any sequence of distributions $Q_k = (1 - \epsilon)P + \epsilon H_k \in \mathcal{U}(P, \epsilon)$ with $\Sigma(Q_k) \in \mathbb{M}^+$, the "bias" goes to to infinity if, and only if, the distributions H_k are concentrated near suitable linear subspaces of $\mathbb{R}^q$.

Some additional notation is first needed. Define

$$\beta(P) := \min_{V \in \mathcal{V}} \frac{[\dim(V)/q - P(V)]^+}{1 - P(V)} \in \left[0, \frac{1}{q}\right]$$

with $0/0 := 0$. It follows from Lemma 12 in the appendix that this minimum is well-defined. Condition (4) is equivalent to $\beta(P) > 0$. Denote the set of all $V \in \mathcal{V}$ such that $[\dim(V)/q - P(V)]^+$ equals $\beta(P)$ by $\mathcal{V}(P)$. Another useful abbreviation is

$$\Pi P := \frac{1}{2}\left(\mathcal{L}_P\left(\boldsymbol{x}/\|\boldsymbol{x}\| \,\Big|\, \boldsymbol{x} \neq 0\right) + \mathcal{L}\left(-\boldsymbol{x}/\|\boldsymbol{x}\| \,\Big|\, \boldsymbol{x} \neq 0\right)\right).$$

6

This is a symmetric distribution on the unit sphere $\mathbf{S}^{q-1}$ of $\mathbb{R}^q$. Note that $\Sigma(P) = \Sigma(\Pi P)$.

Theorem 3 *Suppose Condition (4) holds. Let $P = P\{0\}\delta_0 + (1 - P\{0\})P_o$, where δ_x denotes the Dirac measure at $x \in \mathbb{R}^q$ and P_o is a distribution on $\mathbb{R}^q \setminus \{0\}$. Then*

(a)

$$
\epsilon(P) = \begin{cases} \dfrac{(1 - P\{0\})\beta(P_o)}{1 - P\{0\}\beta(P_o)} & \text{in general,} \\[2mm] \beta(P) & \text{if } P\{0\} = 0, \\[2mm] 1/q & \text{if } P \text{ satisfies (5).} \end{cases}
$$

(b) Let $\epsilon = \epsilon(P)$. For any $Q = (1 - \epsilon)P + \epsilon H$ in $\mathcal{U}(P, \epsilon)$, $\Sigma(Q)$ equals 0 if, and only if, $H\{0\} = 0$ and $H(V) = 1$ for some $V \in \mathcal{V}(P_o)$. Moreover, for $k \geq 1$ let $Q_k = (1 - \epsilon)P + \epsilon H_k \in \mathcal{U}(P, \epsilon)$ be such that $\Sigma(Q_k) \in \mathbb{M}^+$. Then $\lim_{k \to \infty} \lambda_1\left(\Sigma(Q_k)\right) = \infty$ or $\lim_{k \to \infty} \lambda_q\left(\Sigma(Q_k)\right) = 0$ if, and only if, the following two conditions are satisfied:

(i) $\lim_{k \to \infty} H_k\{0\} = 0$, and

(ii) any cluster point $\widetilde{H}$ of $(\Pi H_k)_k$ is supported by some $V \in \mathcal{V}(P_o)$.

Now we address the question of what happens to the functional if we consider only contaminations which are not concentrated near any subspace of dimension less than or equal to r. That is, suppose we replace $\mathcal{U}(P, \epsilon)$ in the definition of the breakdown point with

$$
\mathcal{U}(P, \epsilon \mid \mathcal{H}_r) := \left\{(1 - \epsilon)P + \epsilon H : H \in \mathcal{H}_r\right\},
$$

where $\mathcal{H}_r$ is any collection of distributions on $\mathbb{R}^q$ with the following property: Any cluster point $\widetilde{H}$ of the family $\{\Pi H : H \in \mathcal{H}_r\}$ satisfies $\widetilde{H}(V) = 0$ for any $V \in \mathcal{V}$ with $\dim(V) < r$. Denote the resulting breakdown point for $\Sigma(P)$ by $\epsilon(P \mid \mathcal{H}_r)$, which we note is linearly invariant. We then have for smooth P, $\epsilon(P \mid \mathcal{H}_r) = r/q$.

The proof for this result is analogous to that given for $\epsilon(P)$. In general, define

$$
\beta_r(P) := \min_{V \in \mathcal{V}_r} \frac{[\dim(V)/q - P(V)]^+}{1 - P(V)} \in \left[0, \frac{r}{q}\right],
$$

where $\mathcal{V}_r$ is the set of linear subspaces V of $\mathbb{R}^q$ with $r \leq \dim(V) < q$.

Theorem 4 *Suppose that $\Sigma(P) \in \mathbb{M}^+$. Let $P = P\{0\}\delta_0 + (1 - P\{0\})P_o$, where δ_x denotes Dirac measure in $x \in \mathbb{R}^q$ and P_o is a distribution on $\mathbb{R}^q \setminus \{0\}$. Then*

$$
\epsilon(P \mid \mathcal{H}_r) = \begin{cases} \dfrac{(1 - P\{0\})\beta_r(P_o)}{1 - P\{0\}\beta_r(P_o)} & \text{in general,} \\[2ex] \beta_r(P) & \text{if } P\{0\} = 0, \\[1ex] r/q & \text{if } P \text{ satisfies (5).} \end{cases}
$$

2.1   *Estimation of "shape" and "scale".*

There are different ways to introduce a scaling factor. For instance, in case of $\Sigma(P) \in \mathbb{M}^+$ let $\bar{\Sigma}(P) := \sigma(P)^2 \Sigma(P)$ with a scaling factor $\sigma(P) > 0$ such that

$$
\text{Median}_P\big(\boldsymbol{x}^\top \bar{\Sigma}(P)^{-1}\boldsymbol{x} \,\big|\, \boldsymbol{x} \neq 0\big) \; = \; q.
$$

Then $\bar{\Sigma}(P)$ defines a linear equivariant functional of scatter.

The breakdown point $\epsilon(P)$ remains the same if we would replace $\Sigma(\cdot)$ with $\bar{\Sigma}(\cdot)$. This follows essentially from the observation that for $0 < \epsilon < \epsilon(P) \leq 1/2$ and any distribution $Q \in \mathcal{U}(P, \epsilon)$,

$$
\frac{\text{Median}_Q(\boldsymbol{x}^T \Sigma(Q)^{-1}\boldsymbol{x} \,|\, \boldsymbol{x} \neq 0)}{\text{Median}_Q(\boldsymbol{x}^T \Sigma(P)^{-1}\boldsymbol{x} \,|\, \boldsymbol{x} \neq 0)} \; \in \; \Big[b(P, \epsilon)^{-1}, b(P, \epsilon)\Big].
$$

Moreover, by definition of contamination neighborhoods, $\text{Median}_Q(\boldsymbol{x}^T \Sigma(P)^{-1}\boldsymbol{x} \,|\, \boldsymbol{x} \neq 0)$ lies between the $(1/2 - \epsilon)$ and $(1/2 + \epsilon)$ quantiles of $\mathcal{L}_P(\boldsymbol{x}^T \Sigma(P)^{-1}\boldsymbol{x} \,|\, \boldsymbol{x} \neq 0)$.

For the case of the restricted breakdown point $\epsilon(P \mid \mathcal{H}_r)$, replacing $\Sigma(\cdot)$ with $\bar{\Sigma}(\cdot)$ would result in replacing $\epsilon(P \mid \mathcal{H}_r)$ with $\min\{\epsilon(P \mid \mathcal{H}_r), 1/2\}$.

2.2   *Finite sample properties.*

Consider the special case $P = P_n$, with P_n being the empirical distribution of the q-dimensional data set $\boldsymbol{X} = \{x_1, \ldots, x_n\}$. Further suppose $n > q$ and that the data set $\boldsymbol{X}$ is in general position about the origin, which means that no more than $\dim(V)$ data points in $\boldsymbol{X}$ lie in $V \in \mathcal{V}$. This occurs with probability one when the data set represents a random sample from a smooth distribution in the sense of (5). For this case, we have $\epsilon(P_n) = \beta(P_n) = (n - q)/\{(n - 1)q\}$. The quantity $\epsilon(P_n)$ is related to the finite sample contamination breakdown point introduced by Donoho & Huber (1983). For the statistic $\Sigma(\boldsymbol{X}) := \Sigma(P_n)$, the finite sample contamination breakdown point is

defined to be $\epsilon_c(\boldsymbol{X}) = m_c/(n + m_c)$, where m_c is the smallest value of m such that the statistic $\Sigma(\boldsymbol{X} \cup \boldsymbol{Y})$ breaks down under arbitrary data sets $\boldsymbol{Y} = \{y_1, \ldots, y_m\}$ in $\mathbb{R}^q$ of size m. Since the finite sample size contamination breakdown point considers only contaminating distribution H which are themselves empirical distributions, it follows that $\epsilon_c(\boldsymbol{X}) \geq \epsilon(P_n)$, and so breakdown cannot occur whenever $m < (n - q)/(q - 1)$. Taking into account the integer nature of m, this implies $m_c \geq \lceil (n - q)/(q - 1) \rceil$. Furthermore, if $m > (n - q)/(q - 1)$, then by Proposition 1.c, $\Sigma(\boldsymbol{X} \cup \boldsymbol{Y})$ cannot exist when $\boldsymbol{Y}$ consists of the same value repeated m times and set equal to one of the values in $\boldsymbol{X}$. For the special case $m = (n - q)/(q - 1)$ and with $\boldsymbol{Y}$ choosen as above, the general position of $\boldsymbol{X}$ implies the consequences of Proposition 1.b cannot hold. Thus, $m_c = \lceil (n - q)/(q - 1) \rceil$ and

$$\epsilon_c(\boldsymbol{X}) = \frac{\lceil (n - q)/(q - 1) \rceil}{\lceil (n - 1)q/(q - 1) \rceil}. \tag{6}$$

Another finite sample version of the breakdown point also introduced by Donoho & Huber (1983) is the finite sample replacement breakdown point. For the statistic $\Sigma(\boldsymbol{X})$, the finite sample replacement breakdown point is defined to be $\epsilon_r(\boldsymbol{X}) = m_r/n$, where m_r is the smallest value of m such that the statistic $\Sigma(\mathbf{Z})$ breaks down if we replace $m \leq n$ of our data points $\boldsymbol{x}_i$ with arbitrary points $\boldsymbol{y}_i$, with $\mathbf{Z}$ being the resulting data set of size n. For this case, $\lambda_1(\Sigma(\mathbf{Z}))$ and $\lambda_q(\Sigma(\mathbf{Z}))^{-1}$ are uniformly bounded over all possible $\mathbf{Z}$, provided

$$m \leq \lceil n/q \rceil - 2. \tag{7}$$

The proof of this assertion is given in the appendix. Thus, $m_r \geq \lceil n/q \rceil - 1$. Again, by Proposition 1.b and c, if $m \geq n/q - 1$, then $\Sigma(\mathbf{Z})$ does not exist when the replacement values $\boldsymbol{y}_i$ consists of the same value repeated m times and set equal to one of the remaining values from $\boldsymbol{X}$. Thus, $m_r = \lceil n/q \rceil - 1$ and

$$\epsilon_r(\boldsymbol{X}) = (\lceil n/q \rceil - 1)/n. \tag{8}$$

Note that both (6) and (8) equals $(1/q + o(1))$ as $n \to \infty$.

3 The symmetrized scatter functional

A useful technique to circumvent defining scatter about a given "center" is to first symmetrize the distribution. Such a technique is used for example by Dietel (1993) and by Rousseeuw & Croux

(1993). Here, we consider Dümbgen's (1998) symmetrized version of Tyler's (1987a) M-estimate of scatter, which in its functional form, is defined to be

$$\Sigma_{\mathrm{s}}(P) := \Sigma(P \ominus P),$$

where $P \ominus Q := \mathcal{L}(\boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{x} \neq \boldsymbol{y})$ with independent random vectors $\boldsymbol{x} \sim P, \boldsymbol{y} \sim Q$. In general, $\Sigma_{\mathrm{s}}(P) \neq \Sigma(P)$ even if P itself has a symmetric distribution about the origin, that is even if $\mathcal{L}(\boldsymbol{x}) = \mathcal{L}(-\boldsymbol{x})$ for $\boldsymbol{x} \sim P$. However, if P has an elliptically symmetric distribution centered at the origin, see (2), then $\Sigma_{\mathrm{s}}(P) = \Sigma(P) \propto \Sigma$.

Unlike M-estimates of scatter in general, an important property of $\Sigma_{\mathrm{s}}(P)$ is that it is diagonal whenever the components of $\boldsymbol{x}$ are independent. This follows from the following proposition which provides also a property of the nonsymmetrized functional $\Sigma(\cdot)$ to be utilized later.

Proposition 5 *Let $\boldsymbol{x} \sim P$ be partitioned as $\boldsymbol{x} = (\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top)^\top$ with $\boldsymbol{x}_i \in \mathbb{R}^{q(i)}$, $q(1) + q(2) = q$.*

(a) Suppose that $\mathcal{L}(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) = \mathcal{L}(-\boldsymbol{x}_2 \mid \boldsymbol{x}_1)$. Then $\Sigma(P)$ is block diagonal with diagonal blocks of order $q(1)$ and $q(2)$ respectively.

(b) Suppose that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent. Then $\Sigma_{\mathrm{s}}(P)$ is block diagonal with diagonal blocks of order $q(1)$ and $q(2)$ respectively.

Denote the contamination breakdown point of $\Sigma_{\mathrm{s}}(P)$ by $\epsilon_{\mathrm{s}}(P)$. Also, observe that $P \ominus P$ is smooth in the sense of (5) if P is smooth in the stronger sense that

$$P(L) = 0 \text{ for any affine space } L \subset \mathbb{R}^q. \tag{9}$$

Two useful abbreviations are $\delta(P, Q) := \sum_{x \in \mathbb{R}^q} P\{x\} Q\{x\}$ (i.e. the probability that $\boldsymbol{x} = \boldsymbol{y}$), and $\delta(P) := \delta(P, P)$.

Theorem 6 *Suppose that $\Sigma_{\mathrm{s}}(P) \in \mathbb{M}^+$. Then*

(a)

$$\epsilon_{\mathrm{s}}(P) = \begin{cases} 1 - \sqrt{\dfrac{1 - \beta(P \ominus P)}{1 - \delta(P)\beta(P \ominus P)}} & \text{in general,} \\ 1 - \sqrt{1 - \beta(P \ominus P)} & \text{if } P \text{ has no atoms,} \\ 1 - \sqrt{1 - 1/q} & \text{if } P \text{ satisfies (9).} \end{cases}$$

10

(b) Suppose $\epsilon = \epsilon_s(P)$. Then $\Sigma_s(Q) \in \mathbb{M}^+$ for any Q in $\mathcal{U}(P, \epsilon)$. Moreover, for $k \geq 1$, let $Q_k = (1 - \epsilon)P + \epsilon H_k \in \mathcal{U}(P, \epsilon)$. Then $\lim_{k \to \infty} \lambda_1(\Sigma_s(Q_k)) = \infty$ or $\lim_{k \to \infty} \lambda_q(\Sigma_s(Q_k)) = 0$ if, and only if, the following three conditions are satisfied:

(i) $\lim_{k \to \infty} \max_{x \in \mathbb{R}^q} H_k\{x\} = 0$,

(ii) $|\boldsymbol{y}_k| \to_p \infty$ as $k \to \infty$, where $\boldsymbol{y}_k \sim H_k$, and

(iii) for any cluster point $(\widetilde{H}_1, \widetilde{H}_2)$ of $((\Pi H_k, \Pi(H_k \ominus H_k)))_k$ there exists a subspace $V \in \mathcal{V}(P \ominus P)$ such that $\widetilde{H}_1(V) = \widetilde{H}_2(V) = 1$.

In the theorem above, $\beta(\cdot)$ is defined as in the last section. The quantity $\beta(P \ominus P)$ though can be difficult to compute. However, for $V \in \mathcal{V}$,

$$
\begin{aligned}
(P \ominus P)(V) &= \left(\int P(x + V)\,P(dx) - \delta(P, P)\right) \Big/ (1 - \delta(P, P)) \\
&\leq \int P(x + V)\,P(dx) \leq \max_{x \in \mathbb{R}^q} P(x + V),
\end{aligned}
$$

and so $\beta(P \ominus P) \geq \beta_s(P) := \min_{x \in \mathbb{R}^q, V \in \mathcal{V}}[\dim(V)/q - P(x + V)]^+/(1 - P(x + V))$.

Theorem 6.a shows that symmetrization lowers the breakdown point of the M-functional. However, Theorem 6.b shows that the type of contamination required in order to cause breakdown for $\Sigma_s(P)$ is far more special then that needed to cause breakdown for $\Sigma(P)$. In particular, we have the additional necessary condition (ii). This leads to the question concerning breakdown caused by "tight" contamination or "inliers". By this we mean restricting attention to contaminating distributions H from an arbitrary tight family $\mathcal{H}$ of distributions. This restriction to "tight contamination" does not alter the breakdown point of $\Sigma(P)$, but it does affect the breakdown point of $\Sigma_s(P)$.

Theorem 7 *Suppose that $\Sigma(P \ominus P) \in \mathbb{M}^+$. Then*

(a)
$$
\epsilon_s(P \mid \mathcal{H}) \begin{cases} \geq & \sqrt{\beta_s(P)} \quad \text{in general,} \\ \geq & \sqrt{1/q} \quad \text{if } P \text{ satisfies (9).} \end{cases}
$$

(b) Suppose that P satisfies (9), and let $\epsilon = \sqrt{1/q}$. Then $\Sigma_s(Q) = 0$ for $Q = (1 - \epsilon)P + \epsilon H$ with $H \in \mathcal{H}$ if and only if H has no atoms and H is supported by some one-dimensional affine subspace of $\mathbb{R}^q$. Similarly, for $k \geq 1$, let $Q_k = (1 - \epsilon)P + \epsilon H_k$ with $H \in \mathcal{H}$ and such that

$\Sigma_s(Q_k) \in \mathbb{M}^+$. *Then* $\lim_{k \to \infty} \lambda_1(\Sigma_s(Q_k)) = \infty$ *or* $\lim_{k \to \infty} \lambda_q(\Sigma_s(Q_k)) = 0$ *if and only if the following two conditions are satisfied:*

    (i) $\lim_{k \to \infty} \max_{x \in \mathbb{R}^q} H_k\{x\} = 0$, *and*

    (ii) any cluster point of $(\Pi(H_k \ominus H_k))_k$ *is supported by some* $V \in \mathcal{V}$ *with* $\dim(V) = 1$.

One can easily show that Condition (ii) above implies that any cluster point of $(H_k)_k$ is supported by some one-dimensional affine subspace of $\mathbb{R}^q$.

4   The multivariate t M-functionals

For a distribution P on $\mathbb{R}^q$ and a given value of $\nu > 0$, a multivariate t M-functional of location and scatter is defined to be the solutions $\mu_\nu(P)$ and $\Sigma_\nu(P)$ for $m \in \mathbb{R}^q$ and $M \in \mathbb{M}^+$ respectively of the simultaneous M-functional equations

$$m = \mathbb{E}_P[u_\nu(s)x]/\mathbb{E}_P[u_\nu(s)] \quad \text{and} \quad M = \mathbb{E}_P[u_\nu(s)(x-m)(x-m)^\top], \quad (10)$$

where $s := (x-m)^\top M^{-1}(x-m)$ and $u_\nu(s) := (\nu+q)/(\nu+s)$.

If P_n is an empirical distribution, then $(\mu_\nu(P_n), \Sigma_\nu(P_n))$ represents the maximum likelihood estimate for the location-scatter family of elliptically symmetric t-distributions on ν degrees of freedom. The density of the t-distribution is given by

$$f_\nu(x; \mu, \Sigma) = c_{\nu,q}|\Sigma|^{-1/2}\{1 + (x-\mu)^\top \Sigma^{-1}(x-\mu)/\nu\}^{-(\nu+q)/2}, \quad x \in \mathbb{R}^q, \quad (11)$$

for some suitable normalizing constant $c_{\nu,q}$.

To make the development easier to follow, consider first the scatter-only t M-functional. That is, set $m = 0$ and consider only the solution $\Sigma_\nu^o(P)$ for $M \in \mathbb{M}^+$ to the M-functional equation

$$M = \mathbb{E}_P[u_\nu(x^\top M^{-1}x)xx^\top]. \quad (12)$$

For $\nu = 0$ this corresponds to Tyler's M-functional of scatter. For any integer value of $\nu > 0$, the t M-functional of scatter can be related to Tyler's M-functional in the following manner. First concatenate onto $x \sim P$ the random vector $u \sim U_\nu$ where U_ν represents the uniform distribution

on the $\nu - 1$ dimensional sphere in $\mathbb{R}^\nu$ with radius $\sqrt{\nu}$. Also, let $\boldsymbol{u}$ be independent from $\boldsymbol{x}$. This produces the concatenated random vector $\boldsymbol{y} = (\boldsymbol{x}^\top, \boldsymbol{u}^\top)^\top \sim P \otimes U_\nu$. Now if Tyler's M-functional of scatter exists for $P \otimes U_\nu$, then from Proposition 5.a and the properties of U_ν, it follows that $\Sigma_\nu^o(P)$ must exist and

$$\Sigma(P \otimes U_\nu) \;=\; \lambda \begin{pmatrix} \Sigma_\nu^o(P) & 0 \\ 0 & I_q \end{pmatrix}, \tag{13}$$

where λ is some positive constant. So, applying Proposition 1 on the existence of Tyler's M-functional of scatter at $P \otimes U_\nu$ gives the following existence results for the t M-functionals. This proposition is given in Kent & Tyler (1991) for empirical distributions, and with only strict inequality in part (b). We state the general result here for completeness. Strict inequality is not needed in part (b) since the implication of Proposition 1.b is not possible for $P \otimes U_\nu$.

Proposition 8 *Let $\mathcal{V}_0$ be the set of linear subspaces V of $\mathbb{R}^q$ with $0 \le \dim(V) < q$.*

(a) $\Sigma_\nu^o(P) \in \mathbb{M}^+$ exists and is unique if

$$P(V) \;<\; \frac{\dim(V) + \nu}{q + \nu} \quad \text{for all } V \in \mathcal{V}_0. \tag{14}$$

(b) If $P(V) \ge (\dim(V) + \nu)/(q + \nu)$ for some subspace $V \in \mathcal{V}_0$, then there is no matrix $M \in \mathbb{M}^+$ satisfying (12).

The continuity of $\Sigma(P)$ at any P satisfying (14) follows from Proposition 2.

We can also note from identity (13) that the breakdown point of $\Sigma(P \otimes U_\nu)$, which is $1/(q+\nu)$ when P is smooth in the sense of (5), is a lower bound for the breakdown point of $\Sigma_\nu^o(P)$. This is a lower bound since we are only interested in contaminating $P \otimes U_\nu$ by distributions of the form $H \otimes U_\nu$ and not by general distributions in $\mathbb{R}^{q+\nu}$. It turns out though that the breakdown point of $\Sigma_\nu^o(P)$, which we denote by $\epsilon_\nu^o(P)$, is equal to this lower bound. Analogous comments hold when we restrict attention to contaminations which are not concentrated near subspaces. We state this formally in next theorem. Within the proof of the theorem, which is given in the appendix, we show that $\beta(P \otimes U_\nu)$ equals

$$\beta(P; \nu) := \min\{\beta_0^*(P; \nu), \; 1/(q+\nu)\},$$

and $\beta_r(P \otimes U_\nu)$ equals

$$\beta_r(P; \nu) \;:=\; \begin{cases} \beta(P; \nu) & \text{if } 1 \le r \le \nu, \\ \min\{\beta_{r-\nu}^*(P; \nu), \; r/(q+\nu)\} & \text{if } \nu < r < q + \nu, \end{cases}$$

where $\beta_s^*(P; \nu) := \min_{V \in \mathcal{V}_s} [(\dim(V) + \nu)/(q + \nu) - P(V)]^+/(1 - P(V))$.

Theorem 9 *Let ν be a positive integer and suppose (14) holds. Then*

(a)
$$\epsilon_\nu^o(P) = \begin{cases} \beta(P; \nu) & \text{in general,} \\ \\ 1/(q + \nu) & \text{if } P \text{ satisfies (5).} \end{cases}$$

(b)
$$\epsilon_\nu^o(P \mid \mathcal{H}_r) \leq \begin{cases} \beta_r(P; \nu) & \text{in general,} \\ \\ r/(q + \nu) & \text{if } P \text{ satisfies (5).} \end{cases}$$

We now turn to the simultaneous location-scatter problem. Results for this case can be obtained from the scatter-only problem by using an identity introduced in Kent & Tyler (1991). This identity relates the t M-functionals of location and scatter in q dimensions with parameter ν to a scatter only t M-functional in $q+1$ dimension with parameter $\nu-1$. To obtain this identity, we concatenate to $x \sim P$ the fixed value 1. This produces the concatenated vector $y = (x^\top, 1)^\top \sim P \otimes \delta_1$. From equations (3.5) and (4.1) in Kent & Tyler (1991), we then get the identity

$$\Sigma_{\nu-1}^o(P \otimes \delta_1) = \begin{bmatrix} \Sigma_\nu(P) + \mu_\nu(P)\mu_\nu(P)^\top & \mu_\nu(P) \\ \\ \mu_\nu(P)^\top & 1 \end{bmatrix}, \tag{15}$$

This identity also holds for $\nu = 1$, in which case for the left-hand side we define $\Sigma_0^o(\cdot) := \gamma\Sigma(\cdot)$ for some constant $\gamma > 0$. That is, it is proportional to Tyler's M-functional at $P \otimes \delta_1$. The identity as presented in Kent & Tyler (1991) is for empirical distributions only. However, it is straightforward to note that the identity applies to any arbritrary P.

Using identity (15) together with Proposition 8, we obtain the following existence conditions for the multivariate t M-functionals of location and scatter. This proposition follows from noting that a set $\{y \in \mathbb{R}^{q+1} : y = s(x^\top, 1)^\top, x \in L, s \in \mathbb{R}\}$ with $L \subset \mathbb{R}^q$ is a vector space with dimension d if, and only if, L is an affine space with $\dim(L) = d - 1$.

Proposition 10 *Let $\mathcal{W}$ be the set of affine subspaces L of $\mathbb{R}^q$ with $0 \leq \dim(L) < q$.*

(a) $\mu_\nu(P) \in \mathbb{R}^q$ and $\Sigma_\nu(P) \in \mathbb{M}^+$ exist and are unique if

$$P(L) < \frac{\dim(L) + \nu}{q + \nu} \quad \text{for all } L \in \mathcal{W}. \tag{16}$$

14

(b) If $\nu > 1$ and $P(L) \geq (\dim(L) + \nu)/(q + \nu)$ for some affine space $L \in \mathcal{W}$, then there is no simultaneous solution $m \in \mathbb{R}^q$ and $M \in \mathbb{M}^+$ to (10).

The case that $\nu = 1$ but $P(L) \geq (\dim(L) + 1)/(q + 1)$ is covered by Proposition 1 applied to $P \otimes \delta_1$. The functional $(\mu_\nu(P), \Sigma_\nu(P))$ is continuous at those P for which (16) holds.

The breakdown point of $(\mu_\nu(P), \Sigma_\nu(P))$, which we denote by ϵ_ν, is defined as before with the modification that the maximum bias function is now taken to be

$$b_\nu(\epsilon; P) := \sup_{Q \in \mathcal{U}(P,\epsilon)} \max\{d_\nu^\top \Sigma_\nu(P)^{-1} d_\nu, \ \lambda_1(R_\nu), 1/\lambda_q(R_\nu)\}. \tag{17}$$

where $d_\nu := \mu_\nu(Q) - \mu_\nu(P)$ and $R_\nu := \Sigma_\nu(P)^{-1}\Sigma_\nu(Q)$. This maximum bias function is affine invariant. Breakdown of $(\mu_\nu(\cdot), \Sigma_\nu(\cdot))$ at P is now said to occur if either $\Sigma_\nu(Q)$ does not exist for some $Q \in \mathcal{U}(P, \epsilon)$, $\lambda_1(\Sigma_\nu(Q))$ can be made arbitrarily large, $\lambda_q(\Sigma_\nu(Q))$ can be made arbitrarily small, or $\|\mu_\nu(Q)\|$ can be made arbitrarily large over $Q \in \mathcal{U}(P, \epsilon)$. The breakdown point of $(\mu_\nu(P), \Sigma_\nu(P))$ follows from identity (15) and Theorem 9.

Some modification in our notation is needed for the location-scatter problem. First, replace $V \in \mathcal{V}_r$ with $L \in \mathcal{W}_r$ in the definition of $\beta_r^*(P, \nu)$, where $\mathcal{W}_r$ is the set of all affine subspaces L of $\mathbb{R}^q$ with $r \leq \dim(L) < q$. Also, we need to replace the notion of contamination being bounded away from subspaces to being bounded away from affine spaces, and so let $\mathcal{H}_r^*$ be any collection of distributions on $\mathbb{R}^q$ with the property that any cluster point $\widetilde{H}$ of the set $\{\Pi(H \otimes \delta_1) : H \in \mathcal{H}_r^*\}$ satisfies $\widetilde{H}(W) = 0$ for linear subspaces W of $\mathbb{R}^{q+1}$ with $\dim(W) < r$.

Theorem 11 *Let ν be a positive integer and suppose (16) holds. Then*

(a)
$$\epsilon_\nu(P) = \begin{cases} \beta(P; \nu) & \text{in general,} \\ 1/(q + \nu) & \text{if } P \text{ satisfies (9).} \end{cases}$$

(b)
$$\epsilon_\nu(P \mid \mathcal{H}_r^*) \leq \begin{cases} \beta_r(P; \nu) & \text{in general,} \\ r/(q + \nu) & \text{if } P \text{ satisfies (9).} \end{cases}$$

4.1 *Non-integer values of ν and more general M-functionals.*

In a techincal report, Tyler (1986) studied the finite sample breakdown properties for the "monotonic" M-estimates of scatter only, i.e. the solution over $M \in \mathbb{M}^+$ to the equation

$$M \; = \; \mathbb{E}_P[u(\boldsymbol{x}^\top M^{-1}\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^\top],$$

when the weight function $u(s)$ is such that $su(s)$ is monotonically increasing. This includes the case $u(s) = u_\nu(s)$ for any $\nu > 0$. The methods employed in this section for obtaining the breakdown points for the location-scatter t M-estimates based on $\nu + 1$ degrees of freedom from the breakdown points for the scatter only t M-estimates based on ν degrees of freedom can be applied to any $\nu > 0$ rather than to just integer values of ν.

The general approach used in Tyler (1986) for the scatter only problem is considerably less transparent than the approach used in this section for the scatter only problem for the special case of the t M-estimates based on an integer degrees of freedom. Outside of the t M-estimates, the relationship between a simultaneous M-estimate of location and scatter and a "monotonic" M-estimate of scatter only does not generally hold, nor does the uniqueness of the simultaneous M-estimates of location and scatter generally hold, see Kent & Tyler (1991). Studying the breakdown behavior of the simultaneous M-estimates location and scatter in general may require studying them as minimizers of objective functions, which is the approach used in Kent & Tyler (1991) in establishing stringent conditions on their existence.

References

ADROVER, J.G. (1998). Minimax bias-robust estimation of the dispersion matrix of a multivariate distribution. *Ann. Statist.* **26**, 2301-2320.

BLAKE, A. & MARINOS, C. (1990). Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence* **45**, 323-380.

DAVIES, P.L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15**, 1269-1292.

DIETEL, G. (1993). Global location and dispersion functionals. Ph.D. thesis, University of Essen, Germany.

DONOHO, D.L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Department of Statistics, Harvard University.

DONOHO, D.L. & HUBER P.J. (1983). The notion of breakdown point. In: *A Festschrift for Erich Lehmann (P. Bickel, K. Doksum, & J.L. Hodges, Jr., eds.)*, pp. 157-184. Wadsworth, Belmont, CA.

DÜMBGEN, L. (1998). On Tyler's M-Functional of scatter in high dimension. *Ann. Inst. Statist. Math.* **50**, 471-491.

HALLIN, M. & PAINDAVEINE, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30**, 1103-1133.

HALLIN, M. & PAINDAVEINE, D. (2002b). Multivariate signed ranks: Randles' interdirections or Tyler's angles? In: *Statistical data analysis based on the L1-norm and related methods (Y. Dodge, ed.)*, pp. 271-282. Birkhäuser, Basel.

HETTMANSPERGER, T.P. & RANDLES, R.H. (2003). A practical affine equivariant robust estimate of multivariate location. *Biometrika* **89**, 851-860.

HUBER, P.J. (1981). *Robust Statistics.* Wiley, New York.

KENT, J.T. & TYLER, D.E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Statist.* **15**, 247-254.

KENT, J.T. & TYLER, D.E. (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.* **19**, 2102-2119.

KENT, J.T. & TYLER, D.E. (1996). Constrained M-estimates of multivariate location and scatter. *Ann. Statist.* **24**, 1346-1370.

MARONNA, R.A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4**, 51-67.

MARONNA, R., STAHEL, W.A. & YOHAI, V.J. (1992). Bias-robust estimators of multivariate scatter based on projections. *J. Mult. Anal.* **42**, 141-161.

MARONNA, R. & YOHAI, V.J. (1990). The maximum bias of robust covariances. *Comm. Statist., Theory Meth.* **19**, 3925-3933.

MARONNA, R. & YOHAI, V.J. (1993). Bias-robust estimates of regression based on projections. *Ann. Statist.* **21**, 965-990.

MARTIN, R.D., YOHAI, V.J. & ZAMAR, R. (1989). Min-max robust regression. *Ann. Statist.* **17**, 1608-1630.

RANDLES, R.H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.* **95**, 1263-1268.

ROUSSEEUW, P.J. (1986). Multivariate estimation with high breakdown point. In: *Mathematical Statistics and Applications (W. Grossman, G. Pflug, I. Vincze & W. Wertz, eds.)*, pp. 283-297. Reidel, Dordrecht.

ROUSSEEUW, P.J. & CROUX, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88**, 1273-1283.

STAHEL, W.A. (1981). Breakdown of covariance estimators. Research report 31, Fachgruppe Statistik, ETH Zürich.

TATSUOKA, K. & TYLER, D.E. (2000). The uniqueness of S and M-functionals under non-elliptical distributions. *Ann. Statist.* **28**, 1219-1243.

TYLER, D.E. (1986). Breakdown properties of the M-estimators of multivariate scatter. Technical report, Rutgers University.

TYLER, D.E. (1987a). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15**, 234-251.

TYLER, D.E. (1987b). Statistical analysis for the angular central Gaussian distribution. *Biometrika* **74**, 579-590.

TYLER, D.E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.* **22**, 1024-1044.

TYLER, D.E. (2002). High breakdown point multivariate M-estimation. *Estadistica* **54**, 227-236.

David E. Tyler, Department of Statistics
Rutgers, The State University of New Jersey
Hill Center, Busch Campus, Piscataway, NJ 08854, U.S.A.
E-mail: dtyler@rci.rutgers.edu

Appendix

Lemma 12 *For $0 \leq d < q$, let $\mathcal{V}(d)$ be the set of all d-dimensional linear subspaces of $\mathbb{R}^q$. Then both $\max\{Q(V) \mid V \in \mathcal{V}(d)\}$ and $\max\{Q(x + V) \mid x \in \mathbb{R}^q, V \in \mathcal{V}(d)\}$ are well-defined and upper semicontinuous in Q.*

Proof of Lemma 12. Let $(Q_k)_k$ be any sequence of distributions converging weakly to some Q. Let $V_k \in \mathcal{V}(d)$ and $x_k \in \mathbb{R}^q$ such that either

(i) $x_k = 0$ and $Q_k(V_k) > \sup\{Q_k(V) - k^{-1} \mid V \in \mathcal{V}(d)\}$, or

(ii) $x_k \in V_k^{\perp}$ and $Q_k(x_k + V_k) > \sup\{Q_k(x + V) - k^{-1} \mid x \in \mathbb{R}^q, V \in \mathcal{V}(d)\}$.

Let $M_k \in \mathbb{M}$ describe the orthogonal projection from $\mathbb{R}^q$ onto V_k. After replacing $(Q_k)_k$ with a subsequence if necessary, one may assume that $(M_k)_k$ converges to some projection matrix M, and we define $V := M \mathbb{R}^q$. Further one may assume that $\lim_{k \to \infty} |x_k| = \infty$ or $\lim_{k \to \infty} x_k = x \in \mathbb{R}^q$. Since $x_k + V_k \subset \{y : |y| \geq |x_k|\}$ one easily deduces from $\lim_k Q_k = Q$ and $\lim_k |x_k| = \infty$ that $\lim_k Q_k(x_k + V_k) = 0$. If $\lim_k x_k = x$, then for any $R > 0$,

$$\limsup_{k \to \infty} Q_k(x_k + V_k) \leq \lim_{k \to \infty} \int \Big(1 - R|y - M_k y - x_k|\Big)^+ Q_k(dy) = \int \Big(1 - R|y - My - x|\Big)^+ Q(dy),$$

19

with the right-hand side going to $Q(x + V)$ as $R \to \infty$. This implies Lemma 12 holds. For the special case $(Q_k)_k \equiv Q$ one realizes that both suprema are attained.

Lemma 13 *Let $\lambda_1/\lambda_q(\cdot) := \lambda_1(\cdot)/\lambda_q(\cdot)$.*

(a) Let $\mathcal{Q}$ be a family of nondegenerate distributions on $\mathbb{R}^q$ such that $\Sigma(Q) \in \mathbb{M}^+$ for all $Q \in \mathcal{Q}$ and let $\{\Pi Q : Q \in \mathcal{Q}\}$ be closed. Then $\sup\{\lambda_1/\lambda_q(\Sigma(Q)) \mid Q \in \mathcal{Q}\} < \infty$.

(b) Let $(Q_k)_k$ be a sequence of nondegenerate distributions on $\mathbb{R}^q$ such that $\Sigma(Q_k) \in \mathbb{M}^+$ for all k with $\lim_{k \to \infty} \Pi Q_k = \widetilde{Q}$, and $\lim_{k \to \infty} \lambda_1/\lambda_q(\Sigma(Q_k)) = \rho \in [1, \infty]$. If $\rho = \infty$, then $\widetilde{Q}(V) \geq \dim(V)/q$ for some $V \in \mathcal{V}$. If $\rho < \infty$ but $\widetilde{Q}(V) \geq \dim(V)/q$ for some space $V \in \mathcal{V}$, then there is a second space $W \in \mathcal{V}$ such that $V \cap W = \{0\}$ and $\widetilde{Q}(V \cup W) = 1$.

Proof of Lemma 13. For part (a), Prohorov's Theorem implies that $\{\Pi Q : Q \in \mathcal{Q}\}$ is even compact. Since $\Sigma(Q) = \Sigma(\Pi Q) \in \mathbb{M}^+$ for all $Q \in \mathcal{Q}$, Proposition 2 yields
$$\sup\{\lambda_1/\lambda_q(\Sigma(Q)) \mid Q \in \mathcal{Q}\} = \max\{\lambda_1/\lambda_q(\Sigma(\Pi Q)) \mid Q \in \mathcal{Q}\} < \infty.$$

For part (b), suppose first that $\widetilde{Q}(V) < \dim(V)/q$ for all $V \in \mathcal{V}$. Then $\Sigma(\widetilde{Q}) \in \mathbb{M}^+$ by Proposition 1, and $\Sigma(\widetilde{Q}) = \lim_k \Sigma(Q_k)$ by Proposition 2, whence $\rho = (\lambda_1/\lambda_q)(\Sigma(\widetilde{Q})) < \infty$.

Now suppose that $\rho < \infty$. After replacing $(Q_k)_k$ with a subsequence if necessary, one may assume $\lim_k \Sigma(Q_k) = M \in \mathbb{M}^+$, and so $I = \lim_{k \to \infty} G(\Pi Q_k, \Sigma(Q_k)) = G(\widetilde{Q}, M)$, since $G(\cdot, \Sigma(Q_k))$ converges uniformly to $G(\cdot, M)$ as $k \to \infty$. Thus if $\widetilde{Q}(V) \geq \dim(V)/q$ for some $V \in \mathcal{V}$, then the second part of Proposition 1 says that $V \cap W = \{0\}$ and $\widetilde{Q}(V \cup W) = 1$ for some $W \in \mathcal{V}$.

Proof of Theorem 3. Note first that $\left\{\Pi Q : Q \in \mathcal{U}(P, \epsilon)\right\}$ is equal to the closed set
$$\left\{(1 - \epsilon_o)\Pi P + \epsilon_o \widetilde{H} \mid \widetilde{H} \text{ any symmetric distribution on } \mathbf{S}^{q-1}\right\},$$
where $\epsilon_o := \epsilon/\{1 - (1 - \epsilon)P(0)\}$. For if $Q = (1 - \epsilon)P + \epsilon H \in \mathcal{U}(P, \epsilon)$, then
$$\Pi Q = \frac{(1 - \epsilon)(1 - P\{0\})\Pi P + \epsilon(1 - H\{0\})\Pi H}{(1 - \epsilon)(1 - P\{0\}) + \epsilon(1 - H\{0\})} = (1 - \epsilon')\Pi P + \epsilon'\widetilde{H}$$
for some symmetric distribution $\widetilde{H}$ on $\mathbf{S}^{q-1}$ and
$$\epsilon' := \frac{\epsilon(1 - H\{0\})}{(1 - \epsilon)(1 - P\{0\}) + \epsilon(1 - H\{0\})} \leq \epsilon_o.$$

Further, $\Pi Q(V) \leq (1 - \epsilon_o)\Pi P(V) + \epsilon_o = (1 - \epsilon_o)P_o(V) + \epsilon_o$ with equality if, and only if, $H\{0\} = 0$ and $H(V) = 1$. Furthermore, this is strictly smaller than $\dim(V)/q$ if, and only if, $\epsilon_o < \{\dim(V)/q - P_o(V)\}/\{1 - P_o(V)\}$. Hence we can conclude the following: If $\epsilon_o < \beta(P_o)$ then $\Sigma(\cdot) \in \mathrm{I\!M}^+$ on $\mathcal{U}(P, \epsilon)$, and Lemma 13.a yields that $(\lambda_1/\lambda_q)(\Sigma(\cdot))$ is bounded on $\mathcal{U}(P, \epsilon)$. If $\epsilon_o = \beta(P_o)$, then $\Sigma(Q) = 0$ for $Q = (1 - \epsilon)P + \epsilon H \in \mathcal{U}(P, \epsilon)$ if, and only if, $H\{0\} = 0$ and $H(V) = 1$ for some $V \in \mathcal{V}(P_o)$. Since ϵ_o is strictly increasing in ϵ, inverting the equation $\epsilon_o = \beta(P_o)$ yields $\epsilon(P) = \{(1 - P(0))\beta(P_o)\}/\{1 - P(0)\beta(P_o)\}$.

Let $\epsilon = \epsilon(P)$ and $Q_k = (1 - \epsilon)P + \epsilon H_k \in \mathcal{U}(P, \epsilon)$ as stated in the theorem. After replacing $(Q_k)_k$ with a subsequence if necessary, one may assume that $\lim_k H_k\{0\} = a \in [0, 1]$, $\lim_k \Pi H_k = \widetilde{H}$ (where $\Pi \delta_0$ may be defined arbitrarily) and $\lim_k (\lambda_1/\lambda_q)(\Sigma(Q_k)) = \rho \in [1, \infty]$. This implies that

$$\lim_{k \to \infty} \Pi Q_k = \widetilde{Q} := \frac{(1 - \epsilon)(1 - P\{0\})\Pi P + \epsilon(1 - a)\widetilde{H}}{(1 - \epsilon)(1 - P\{0\}) + \epsilon(1 - a)}.$$

Since $\Sigma(P) \in \mathrm{I\!M}^+$, $\Pi P(V \cup W) < 1$ for arbitrary $V, W \in \mathcal{V}$ with $V \cap W = \{0\}$. The limit distribution $\widetilde{Q}$ inherits this property. Thus one can apply Lemma 13.b and conclude that $\rho = \infty$ if, and only if, $\widetilde{Q}(V) \geq \dim(V)/q$ for some $V \in \mathcal{V}$. But for any $V \in \mathcal{V}$,

$$\widetilde{Q}(V) = \frac{(1 - \epsilon)(1 - P(0))P_o(V) + \epsilon(1 - a)\widetilde{H}(V)}{(1 - \epsilon)(1 - P(0)) + \epsilon(1 - a)} \leq \frac{(1 - \epsilon)(1 - P(0))P_o(V) + \epsilon(1 - a)}{(1 - \epsilon)(1 - P(0)) + \epsilon(1 - a)}$$

$$\leq \frac{(1 - \epsilon)(1 - P(0))P_o(V) + \epsilon}{(1 - \epsilon)(1 - P(0)) + \epsilon} = (1 - \epsilon_o)P_o(V) + \epsilon_o \leq (1 - \epsilon_o)\frac{\dim(V)/q - \beta(P_o)}{1 - \beta(P_o)} + \epsilon_o.$$

The last expression equals $\dim(V)/q$. Equality holds, that is $\widetilde{Q}(V) = \dim(V)/q$, if and only if, $\widetilde{H}(V) = 1$, $a = 0$ and $V \in \mathcal{V}(P_o)$.

Proof of Statement (7). Let $\mathcal{S}_{n-m}$ be the set of all subsets of $\{1, \ldots, n\}$ with $n - m$ elements. For any $S \in \mathcal{S}_{n-m}$ let P_S be the empirical distrubution of the sample points $\boldsymbol{x}_i$, $i \in S$. Thus replacing up to m data points results in an empirical distribution Q_n belonging to $\bigcup\{\mathcal{U}(P_S, m/n) \mid S \in \mathcal{S}_{n-m}\}$. Statement (7) can then be restated as $\min\{\epsilon(P_S) \mid S \in \mathcal{S}_{n-m}\} > m/n$. Now, for any $S \in \mathcal{S}_{n-m}$,

$$\epsilon(P_S) = \min_{d=1,\ldots,q-1} \frac{d/q - d/(n - m)}{1 - d/(n - m)} = \frac{1/q - 1/(n - m)}{1 - 1/(n - m)}.$$

21

This quantity is strictly larger than m/n if, and only if, m is strictly less than $n/q - 1$, which is equivalent to $m \le \lceil n/q \rceil - 2$.

Proof of Proposition 5. As in Kent & Tyler (1988, 1991) one can show that the sequence $(M_k)_{k=0}^{\infty}$ with $M_0 := I$ and $M_{k+1} := q\, E_P[(\boldsymbol{x}^\top M_k^{-1} \boldsymbol{x})^{-1} \boldsymbol{x}\boldsymbol{x}^\top \,|\, \boldsymbol{x} \ne 0]$ converges to $\Sigma(P)$. The symmetry condition about $\mathcal{L}(\boldsymbol{x}_2 \,|\, \boldsymbol{x}_1)$ in part (a) is equivalent to saying that $\boldsymbol{x}$ has the same distribution as $(\boldsymbol{x}_1^\top, S\boldsymbol{x}_2^\top)^\top$, where S is independent from $\boldsymbol{x}$ and uniformly distributed on $\{-1, 1\}$. If M_k is a block diagonal with blocks A_k and B_k, which is true in case of $k = 0$, then

$$M_{k+1} = q\, E_P\left(\left(\boldsymbol{x}_1^\top A_k^{-1}\boldsymbol{x}_1 + \boldsymbol{x}_2^\top B_k^{-1}\boldsymbol{x}_2\right)^{-1} \begin{pmatrix} \boldsymbol{x}_1\boldsymbol{x}_1^\top & S\boldsymbol{x}_1\boldsymbol{x}_2^\top \\ S\boldsymbol{x}_2\boldsymbol{x}_1^\top & \boldsymbol{x}_2\boldsymbol{x}_2^\top \end{pmatrix} \,\Big|\, \boldsymbol{x} \ne 0 \right)$$

which is block diagonal with blocks, say A_{k+1} and B_{k+1}. This proves the assertion from part (a). Part (b) follows from part (a) applied to $\boldsymbol{x}_i - \boldsymbol{y}_i$ in place of $\boldsymbol{x}_i$.

The following preliminary result for the proof of Theorems 6 and 7 describes the possible limits of a sequence $(\Pi(P \ominus H_k))_k$.

Proposition 14 *Let $(H_k)_{k \ge 1}$ be a sequence of distributions on $\mathbb{R}^q$. A pair (a, B) is cluster point for the sequence of pairs $\left(\delta(P, H_k), \Pi(P \ominus H_k)\right)$ if, and only if, it can be represented as follows: $a = \sum_{x \in \mathbb{R}^q} P(x)a_x$ and*

$$B = \frac{\eta B_\infty + \displaystyle\sum_{x \in \mathbb{R}^q} P(x)\left((1-\eta)H\{x\} - a_x\right)B_x + (1-\eta)(1 - \delta(P, H))\Pi(P \ominus H)}{1 - \displaystyle\sum_{x \in \mathbb{R}^q} P\{x\}a_x},$$

for some distribution H on $\mathbb{R}^q$, some numbers $a_x \in [0, (1-\eta)H\{x\}]$ and some symmetric distributions B_∞ and B_x on $\mathbf{S}^{q-1}$, and where $\eta := \lim_{r \to \infty} \liminf_{k \to \infty} H_k\{x : |x| > r\}$.

Proof of Proposition 14. We compactify $\mathbb{R}^q$ via the mapping

$$x \mapsto \psi(x) := (1 + |x|)^{-1}x \in U(0, 1),$$

where $U(y, \delta)$ and $B(y, \delta)$ denote, respectively, the open and closed ball around $y \in \mathbb{R}^q$ with radius $\delta \ge 0$. Without loss of generality one may assume that the sequence of transformed distributions $H_k \circ \psi^{-1}$ converges weakly to some distribution D on $B(0, 1)$, and then $\eta = D(\mathbf{S}^{q-1})$. Even if D is concentrated on $U(0, 1)$ the Continuous Mapping Theorem is not applicable to $\Pi(P \ominus H_k)$,

because points within the countable set $\mathcal{X} := \left\{ x \in \mathbb{R}^q : D\{\psi(x)\} > 0 \right\}$ require special atten-tion. Since

$$D\{\psi(x)\} = \lim_{\delta \downarrow 0} \liminf_{k \to \infty} H_k U(x, \delta) = \lim_{\delta \downarrow 0} \limsup_{k \to \infty} H_k B(x, \delta)$$

for any $x \in \mathcal{X}$ and

$$\eta = \lim_{r \uparrow \infty} \liminf_{k \to \infty} H_k(\mathbb{R}^q \setminus B(0, r)) = \lim_{r \uparrow \infty} \limsup_{k \to \infty} H_k(\mathbb{R}^q \setminus U(0, r)),$$

one can find numbers $\delta_{x,k} \geq 0$ and $r_k > 0$ such that with $U_{x,k} := U(x, \delta_{x,k})$ and $U_{\infty,k} :=$ $\mathbb{R}^q \setminus B(0, r_k)$ the following requirements are met:

(i) $\lim_{k \to \infty} \delta_{x,k} = 0$ and $\lim_{k \to \infty} H_k U_{x,k} = D\{\psi(x)\}$ for $x \in \mathcal{X}$,

(ii) $\lim_{k \to \infty} r_k = \infty$ and $\lim_{k \to \infty} H_k U_{\infty,k} = \eta$, and

(iii) $U_{x,k} \cap U_{y,k} = \emptyset$ for different $x, y \in \mathcal{X} \cup \{\infty\}$.

After replacing $(H_k)_k$ with a suitable subsequence if necessary, one may assume further that for any $x \in \mathcal{X}$,

$$\lim_{k \to \infty} H_k\{x\} = a_x \in [0, D\{\psi(x)\}] \quad \text{and} \quad \lim_{k \to \infty} \Pi\mathcal{L}(x - \boldsymbol{y}_k \mid \boldsymbol{y}_k \in U_{x,k} \setminus \{x\}) = B_x,$$

where $\boldsymbol{y}_k \sim H_k$. Since $\lim_k H_k\{x\} = 0$ whenever $D\{\psi(x)\} = 0$, this implies that

$$\lim_{k \to \infty} \delta(P, H_k) = \sum_{x \in \mathcal{X}} P\{x\} a_x. \tag{18}$$

Now express $D = \eta B_* + (1 - \eta) H \circ \psi^{-1}$ with distributions B_* on $\mathbf{S}^{q-1}$ and H on $\mathbb{R}^q$, and let

$$f(x) := \begin{cases} g(|x|^{-1} x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

for some even, continuous function g on $\mathbf{S}^{q-1}$, and let $\boldsymbol{x} \sim P$, $\boldsymbol{y}_k \sim H_k$ and $\boldsymbol{y} \sim H$ be in-dependent. Then $\mathbb{E} f(\boldsymbol{x} - \boldsymbol{y}_k)$ may be split into $\mathbb{E} \mathbf{1}\{\boldsymbol{y}_k \in U_{\infty,k}\} f(\boldsymbol{x} - \boldsymbol{y}_k)$ and $\mathbb{E} \mathbf{1}\{\boldsymbol{y}_k \notin U_{\infty,k}\} f(\boldsymbol{x} - \boldsymbol{y}_k)$, and as $k \to \infty$,

$\mathbb{E} \mathbf{1}\{\boldsymbol{y}_k \in U_{\infty,k}\} f(\boldsymbol{x} - \boldsymbol{y}_k) = \eta \int g \, dB_* + o(1) = \eta \int g \, d\Pi B_* + o(1)$, and

$\mathbb{E} \mathbf{1}\{\boldsymbol{y}_k \notin U_{\infty,k}\} f(\boldsymbol{x} - \boldsymbol{y}_k) = \sum_{x \in \mathcal{X}} P\{x\} \mathbb{E} \mathbf{1}\{\boldsymbol{y}_k \in U_{x,k} \setminus \{x\}\} f(x - \boldsymbol{y}_k)$

$\qquad + \sum_{x \in \mathcal{X}} P\{x\} \mathbb{E} \mathbf{1}\{\boldsymbol{y}_k \notin U_{x,k} \cup U_{\infty,k}\} f(x - \boldsymbol{y}_k) + \mathbb{E} \mathbf{1}\{\boldsymbol{x} \notin \mathcal{X}, \boldsymbol{y}_k \notin U_{\infty,k}\} f(\boldsymbol{x} - \boldsymbol{y}_k)$

$= \sum_{x \in \mathcal{X}} P\{x\} \left(D\{\psi(x)\} - a_x \right) \int g \, dB_x + (1 - \eta) \sum_{x \in \mathcal{X}} P\{x\} \mathbb{E} \mathbf{1}\{\boldsymbol{y} \neq x\} f(x - \boldsymbol{y})$

$\qquad + (1 - \eta) \mathbb{E} \mathbf{1}\{\boldsymbol{x} \notin \mathcal{X}\} f(\boldsymbol{x} - \boldsymbol{y}) + o(1)$

$= \sum_{x \in \mathcal{X}} P\{x\} \left(D\{\psi(x)\} - a_x \right) \int g \, dB_x + (1 - \eta)(1 - \delta(P, H)) \int g \, d\Pi(P \ominus H) + o(1).$

23

Together with (18) this shows that $(\Pi(P \ominus H_k))_k$ converges weakly to a distribution B as stated in the proposition, where $B_\infty = \Pi B_*$.

Proof of Theorem 6. A detailed study of the closure of the set $\left\{ \Pi(Q \ominus Q) : Q \in \mathcal{U}(P, \epsilon) \right\}$ is the basis of this proof. For $k \geq 1$, let $Q_k = (1 - \epsilon)P + \epsilon H_k \in \mathcal{U}(P, \epsilon)$ be defined such that $\lim_{k \to \infty} \Pi(Q_k \ominus Q_k) = \widetilde{Q}$. By compactness arguments one may assume without loss of generality that $\lim_{k \to \infty} \delta(P, H_k) = a_{PH}$, $\lim_{k \to \infty} \Pi(P \ominus H_k) = B_{PH}$, $\lim_{k \to \infty} \delta(H_k) = a_{HH}$, and $\lim_{k \to \infty} \Pi(H_k \ominus H_k) = B_{HH}$. This yields the representation

$$\widetilde{Q} = \frac{(1 - \epsilon)^2 (1 - \delta(P))\Pi(P \ominus P) + 2\epsilon(1 - \epsilon)(1 - a_{PH})B_{PH} + \epsilon^2(1 - a_{HH})B_{HH}}{(1 - \epsilon)^2(1 - \delta(P)) + 2\epsilon(1 - \epsilon)(1 - a_{PH}) + \epsilon^2(1 - a_{HH})}.$$

Note first that $\widetilde{Q} = (1 - \epsilon')\Pi(P \ominus P) + \epsilon'\widetilde{H}$ for some symmetric distribution $\widetilde{H}$ on $\mathbf{S}^{q-1}$ and

$$\begin{aligned}
\epsilon' \; &:= \; \frac{2\epsilon(1 - \epsilon)(1 - a_{PH}) + \epsilon^2(1 - a_{HH})}{(1 - \epsilon)^2(1 - \delta(P)) + 2\epsilon(1 - \epsilon)(1 - a_{PH}) + \epsilon^2(1 - a_{HH})} \\
&\leq \; \frac{2\epsilon - \epsilon^2}{(1 - \epsilon)^2(1 - \delta(P)) + 2\epsilon - \epsilon^2} = \frac{1 - (1 - \epsilon)^2}{1 - (1 - \epsilon)^2\delta(P)} =: \epsilon_o.
\end{aligned}$$

Thus $\left\{ \Pi(Q \ominus Q) : Q \in \mathcal{U}(P, \epsilon) \right\}$ is *contained* in the closed set

$$\left\{ (1 - \epsilon_o)\Pi(P \ominus P) + \epsilon_o\widetilde{H} : \widetilde{H} \text{ any symmetric distribution on } \mathbf{S}^{q-1} \right\}.$$

Consequently, $\Sigma(Q \ominus Q) \in \mathbb{M}^+$ for all $Q \in \mathcal{U}(P, \epsilon)$ with $\sup_{Q \in \mathcal{U}(P, \epsilon)}(\lambda_1/\lambda_q)(\Sigma(Q \ominus Q))$ being finite, provided that $\epsilon_o < \beta(\Pi(P \ominus P)) = \beta(P \ominus P)$, which is equivalent to

$$\epsilon < \epsilon^*(P) := 1 - \sqrt{\frac{1 - \beta(P \ominus P)}{1 - \delta(P)\beta(P \ominus P)}}.$$

Now suppose that $\epsilon = \epsilon^*(P)$, that means, $\epsilon_o = \beta(P \ominus P)$. Then $\widetilde{Q}(V) \geq \dim(V)/q$ for some $V \in \mathcal{V}$ if, and only if, $a_{PH} = a_{HH} = 0$, $B_{PH}(V) = B_{HH}(V) = 1$ and $V \in \mathcal{V}(P \ominus P)$. These equations cannot hold if $\widetilde{Q} = \Pi(Q \ominus Q)$ for some distribution $Q = (1 - \epsilon)P + \epsilon H \in \mathcal{U}(P, \epsilon)$. For then $B_{PH}(V) = P \ominus H(V) \leq \max_{x \in \mathbb{R}^q} P(x + V) < 1$, because otherwise $P(x + V) = 1$ for some $x \in \mathbb{R}^q$, so that $P \ominus P(V) = 1$ and $\Sigma(P \ominus P) = 0$.

The equation $a_{HH} = 0$ is equivalent to Condition (i) in Theorem 6.b and entails that $a_{PH} = 0$ as well. Moreover, Proposition 14 implies that

$$B_{HP} = \eta B_\infty + (1 - \eta) \sum_{x \in \mathbb{R}^q} P\{x\}H\{x\}B_x + (1 - \eta)(1 - \delta(P, H))\Pi(P \ominus H)$$

for some distribution H on $\mathbb{R}^q$, some number $\eta \in [0,1]$ and symmetric distributions B_y, $y \in \mathbb{R}^q \cup \{\infty\}$, on $\mathbf{S}^{q-1}$. This representation shows that $B_{PH}(V) = 1$ for some $V \in \mathcal{V}(\Pi(P \ominus P))$ if, and only if, $\eta = 1$ and $B_\infty = B_{PH} = \lim_k \Pi H_k$ is concentrated on V. Together with the requirement $B_{HH}(V) = 1$ we end up with Conditions (ii) and (iii) of Theorem 6.b about the sequence $(H_k)_k$. All requirements (i), (ii), and (iii) of Theorem 6.b are satisfied, for instance, by $H_k := \mathcal{L}(k\boldsymbol{y})$, where $\boldsymbol{y}$ is some random vector whose distribution is concentrated on V but has no atoms. Thus $\epsilon_s(P) = \epsilon^*(P)$.

Proof of Theorem 7. Let $Q_k = (1-\epsilon)P + \epsilon H_k$ with $H_k \in \mathcal{H}$, and let $\widetilde{Q}$, a_{PH}, a_{HH}, B_{PH}, B_{HH} be as in the proof of Theorem 6. Since the sequence $(H_k)_k$ is tight, Proposition 14 yields that $a_{PH} = \sum_{x \in \mathbb{R}^q} P\{x\} a_x$ and

$$B_{PH} = \frac{\sum_{x \in \mathbb{R}^q} P\{x\}(H\{x\} - a_x)B_x + (1 - \delta(P,H))\Pi(P \ominus H)}{1 - \sum_{x \in \mathbb{R}^q} P\{x\} a_x}$$

for some distribution H on $\mathbb{R}^q$, numbers $a_x \in [0, H\{x\}]$ and symmetric distributions B_y, $y \in \mathbb{R}^q \cup \{\infty\}$, on $\mathbf{S}^{q-1}$. Thus for any $V \in \mathcal{V}$,

$$\begin{aligned}
(1 - a_{PH})B_{PH}(V) &\leq \sum_{x \in \mathbb{R}^q} P\{x\}(H\{x\} - a_x) + (1 - \delta(P,H))(P \ominus H)(V) \\
&= \int P(x + V)\, H(dx) - a_{PH} \leq \max_{x \in \mathbb{R}^q} P(x + V) - a_{PH},
\end{aligned}$$

and $(1 - \delta(P))\Pi(P \ominus P)(V) = \int P(x + V)P(dx) - \delta(P) \leq \max_{x \in \mathbb{R}^q} P(x + V) - \delta(P)$. Thus

$$\begin{aligned}
\widetilde{Q}(V) &\leq \frac{(1 - \epsilon^2)\max\limits_{x \in \mathbb{R}^q} P(x + V) + \epsilon^2 - (1 - \epsilon)^2\delta(P) - 2\epsilon(1 - \epsilon)a_{PH} - \epsilon^2 a_{HH}}{1 - (1 - \epsilon)^2 a_{PP} - 2\epsilon(1 - \epsilon)a_{PH} - \epsilon^2 a_{HH}} \\
&\leq (1 - \epsilon^2)\max_{x \in \mathbb{R}^q} P(x + V) + \epsilon^2.
\end{aligned}$$

This shows that $\epsilon_s(P \,|\, \mathcal{H}) \leq \sqrt{\beta_s(P)}$.

In case of P being smooth in the sense of (9),

$$\widetilde{Q}(V) = \frac{\epsilon^2(1 - a_{HH})B_{HH}(V)}{1 - \epsilon^2 + \epsilon^2(1 - a_{HH})} \leq \epsilon^2$$

with equality if, and only if, $a_{HH} = 0$ and $B_{HH}(V) = 1$.

Proof of Theorem 9. Let us first prove the representations of $\beta(P \otimes U_\nu)$ and $\beta_r(P \otimes U_\nu)$ for $1 \leq r < q + \nu$. Suppose that W is a linear subspace of $\mathbb{R}^{q+\nu}$ with dimension $d \in [1, q+\nu)$. Then

$$\frac{d/(q + \nu) - P \otimes U_\nu(W)}{1 - P \otimes U_\nu(W)} \leq \frac{d}{q + \nu}$$

25

with equality if, and only if, $P \otimes U_\nu (W) = 0$. In case of $P \otimes U_\nu (W) > 0$, let $w_1, w_2, \ldots, w_{q+\nu}$ be an orthonormal basis of $\mathbb{R}^{q+\nu}$ such that $W = \operatorname{span}(w_1, \ldots, w_d)$. Then for independent random vectors $\boldsymbol{x} \sim P$ and $\boldsymbol{u} \sim U_\nu$,

$$0 < P \otimes U_\nu (W) = P\left[(\boldsymbol{x}^\top, \boldsymbol{u}^\top)w_j = 0 \text{ for } j > d\right] \le \min_{j>d} P\left[(\boldsymbol{x}^\top, \boldsymbol{u}^\top)w_j = 0\right].$$

Since $\boldsymbol{u}^\top \widetilde{w}$ has a continuous distribution for any nonzero vector $\widetilde{w}$ in $\mathbb{R}^\nu$, all vectors w_j, $j > d$, have to belong to $\mathbb{R}^q \times \{0\}$. Hence W contains the space $\{0\} \times \mathbb{R}^\nu$ and thus may be written as $V \times \mathbb{R}^\nu$ for some linear subspace V of $\mathbb{R}^q$ with dimension $d - \nu$; in particular $d \ge \nu$. Then

$$\frac{d/(q+\nu) - P \otimes U_\nu (W)}{1 - P \otimes U_\nu (W)} = \frac{(\dim(V) + \nu)/(q+\nu) - P(V)}{1 - P(V)} \ge \beta^*_{d-\nu}(P; \nu)$$

with equality for a suitable $V \in \mathcal{V}_{d-\nu}$. These considerations entail the asserted formulae for $\beta(P \otimes U_\nu)$ and $\beta_r(P \otimes U_\nu)$.

Now it follows from Theorems 3 and 4 that $\epsilon^o_\nu(P) \ge \beta(P; \nu)$ and $\epsilon^o_\nu(P \,|\, \mathcal{H}_r) \ge \beta_r(P; \nu)$. Thus it suffices to show that $\epsilon^o_\nu(P) \le \beta(P; \nu)$.

At first let $\epsilon := \beta(P; \nu) = 1/(q+\nu)$. Then we define $Q_k = (1 - \epsilon)P + \epsilon\delta_{kz}$ for a fixed unit vector $z \in \mathbb{R}^q$. Here $\Pi(Q_k \otimes U_\nu) \to (1 - \epsilon)\Pi(P \otimes U_\nu) + \epsilon\delta_z$ as $k \to \infty$. Now it follows from part (b) of Theorem 3, applied to $(P \otimes U_\nu, \delta_{kz} \otimes U_\nu)$ in place of (P, H_k), that $(\lambda_1/\lambda_{q+\nu})(\Sigma_\nu(Q_k \otimes U_\nu)) \to \infty$. But this entails that $(\lambda_1(\Sigma^o_\nu(Q_k)))^\infty_{k=1}$ or $(\lambda_q(\Sigma^o_\nu(Q_k))^{-1})^\infty_{k=1}$ is unbounded.

Finally let $\epsilon := \beta(P; \nu) < 1/(q+\nu)$. That means, there exists a linear subspace V of $\mathbb{R}^q$ with dimension $d \in [0, q)$ such that $(d/(q+\nu) - P(V))/(1 - P(V)) = \epsilon$, which is equivalent to $P(V) = ((d+\nu)/(q+\nu) - \epsilon)/(1 - \epsilon)$. If we define $Q = (1 - \epsilon)P + \epsilon\delta_0$, then

$$Q \otimes U_\nu (V \times \mathbb{R}^q) = Q(V) = (d+\nu)/(q+\nu) = \dim(V \times \mathbb{R}^q)/(q+\nu).$$

Thus $\Sigma^o_\nu(Q)$ is not defined.