The Masking Breakdown Point of Multivariate Outlier Identification Rules

Claudia Becker and Ursula Gather [1]

Abstract

In this paper, we consider one-step outlier identification rules for multivariate data, generalizing the concept of so-called α outlier identifiers, as presented in Davies and Gather (1993) for the case of univariate samples. We investigate, how the finite-sample breakdown points of estimators used in these identification rules influence the masking behaviour of the rules.

Keywords: Breakdown points; Outlier identification; Masking; Robust statistics.

1 Introduction

It is well known that outliers, i.e. observations lying "far away" from the main part of a data set and probably not following the assumed model, can strongly influence the statistical analysis of that data and even falsify the results. In particular, some classical parametric tests and estimators, e.g. the arithmetic mean as a location estimate, are prone to the influence of outlying observations. Therefore, one often finds the identification of outliers treated as a means to screen a data set for 'bad' observations firstly, thus avoiding a distortion of the statistical analysis. But outliers can be of fundamental interest in themselves and therefore their identification should also be considered as a goal in itself.

In multivariate data sets, it is almost impossible to detect outliers by pure vision, because they do not "stick out on the end" (Gnanadesikan and Kettenring, 1972, p. 109) as in univariate situations. Observations which are not conspicuous in any single variable may nevertheless differ clearly from the rest of the data if all variables are looked at simultaneously (cf. Rousseeuw and Leroy, 1987, p. 7 for an example). Some methods proposed for the identification of outliers in multivariate samples are of heuristic nature (e.g. Atkinson and

[1]Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany.

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475

Mulira, 1993, Bacon-Shone and Fung, 1987, Barnett and Lewis, 1994, p. 307 ff., Bhandary, 1992), others are of consecutive testing type (see Barnett and Lewis, 1994, chap. 7.3, Caroni and Prescott, 1992, Hara, 1988, Hawkins, 1980, chap. 8, Wilks, 1963). In this article, we discuss the approach of one-step outlier identification rules.

Our paper is organized as follows: In Section 2, we make precise in a formal way how we understand the task of outlier identification and we give a definition of a multivariate outlier identifier. For comparing the behaviour of such identifiers, performance criteria are needed. In Section 3, we deal with the masking breakdown point as a worst-case criterion and its relation to the finite-sample breakdown points of the estimators, the identification rule is based on. Finally, we give a small example.

2 Multivariate outlier identification

The identification of outliers heavily relies on the assumption of some underlying model for the data. An observation can finally only be considered as an outlier with respect to such a model in mind. Here, we look at the p-variate normal distribution $N(\ ,\Sigma)$ as a model distribution, where $\ \in I\!R^p$, $\Sigma \in I\!R^{p\times p}$, Σ positive definite (p.d.). In analogy to the definition of Davies and Gather (1993, p. 782) for the case of the univariate normal, Gather and Becker (1997, p. 129) give the general concept of an α outlier which can also be applied to the multivariate normal case. An α *outlier* with respect to $N(\ ,\Sigma)$ is then defined as an element of the set

$$\mathrm{out}(\alpha,\ ,\Sigma) := \{\mathbf{x} \in I\!R^p : (\mathbf{x} -\ )^T\Sigma^{-1}(\mathbf{x} -\ ) > \chi^2_{p;1-\alpha}\}$$

for $\alpha \in (0,1)$, $\chi^2_{p;1-\alpha}$ denoting the $(1-\alpha)$-quantile of the χ^2_p-distribution. The set $\mathrm{out}(\alpha,\ ,\Sigma)$ itself is called the α *outlier region* of $N(\ ,\Sigma)$. Thus, we have

$$\mathrm{P}(\mathbf{X} \in \mathrm{out}(\alpha,\ ,\Sigma)) = \alpha \quad \text{for } \mathbf{X} \sim N(\ ,\Sigma)$$

and for usual choices of α ($\alpha = 0.05, \alpha = 0.1$) this reflects the idea of an outlier being an observation which is rather unlikely under the assumed model and also situated 'outside the main mass of the distribution'.

The size of the outlier region may be adjusted to the sample size. For a sample of size N, one may consider out$(\alpha_N, \ \ , \Sigma)$, where, as in the univariate case, α_N can be chosen according to the condition

$$\mathrm{P}(\mathbf{X}_i \in \mathbb{R}^p \backslash \mathrm{out}(\alpha_N, \ \ , \Sigma), i = 1, \ldots, N) = 1 - \alpha \tag{1}$$

for $\mathbf{X}_i \sim N(\ \ , \Sigma)$, $i = 1, \ldots, N$, and some given $\alpha \in (0, 1)$. Thus, $\alpha_N = 1 - (1 - \alpha)^{1/N}$.

Our aim is now to detect all α_N outliers in a given sample $\underline{x}_N = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ of size N. As the parameters and Σ are unknown, the outlier region itself is unknown, and our task is equivalent to the task of estimating the α_N outlier region of the model distribution from data which themselves may not all be "clean".

This motivates the following definition (Gather and Becker, 1997, p. 132):

Let $\alpha_N \in (0, 1)$, $\underline{x}_N = (\underline{x}_n^r, \underline{x}_k^0)$ with $\underline{x}_n^r := (\mathbf{x}_1^r, \ldots, \mathbf{x}_n^r)$ be a sample of size n of i.i.d. $N(\ \ , \Sigma)$ distributed random vectors; let the remaining k observations $(\mathbf{x}_1^0, \ldots, \mathbf{x}_k^0) =: \underline{x}_k^0$ be δ_N outliers with respect to $N(\ \ , \Sigma)$ for some $\delta_N \in (0, 1)$, where $k = N - n, k < N/2$, k, δ_N unknown. An α_N *outlier identifier* is defined as a region

$$\mathbf{OR}(\underline{x}_N, \alpha_N) := \{\mathbf{x} \in \mathbb{R}^p : (\mathbf{x} - \mathbf{m})^T S^{-1} (\mathbf{x} - \mathbf{m}) \geq c\},$$

where $S = S(\underline{x}_N) \in \mathrm{PDS}(p)$, $\mathbf{m} = \mathbf{m}(\underline{x}_N) \in \mathbb{R}^p$, and $c = c(p, N, \alpha_N) \in \mathbb{R}, c \geq 0$, not depending on the arrangement and the existence of any δ_N outliers in $\underline{x}_N$ at all. All points $\mathbf{x} \in \mathbf{OR}(\underline{x}_N, \alpha_N)$ are identified as α_N outliers with respect to $N(\ \ , \Sigma)$.

Here, *PDS*$(p) = \{S \in \mathbb{R}^{p \times p} : S \text{ p.d. and symmetric}\}$, and $c(p, N, \alpha_N)$ is a normalizing constant. Several normalizing conditions are possible in order to fix c appropriately. We will restrict ourselves to the condition (following (1))

$$\mathrm{P}(\mathbf{X}_i \in \mathbb{R}^p \backslash \mathbf{OR}(\underline{X}_N, \alpha_N), i = 1, \ldots, N) = 1 - \alpha \tag{2}$$

for $\alpha \in (0, 1)$ and $\alpha_N = 1 - (1 - \alpha)^{1/N}$, where $\underline{X}_N = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$, $\mathbf{X}_1, \ldots, \mathbf{X}_N$ i.i.d. according to $N(\ \ , \Sigma)$. This means that with probability $1 - \alpha$ in a sample of size N from the p-variate normal, no observation will be identified as an outlier.

All further considerations will be restricted to affine equivariant identifiers $\mathbf{OR}$. Given an affine linear transformation $\underline{x}_N \mapsto A\underline{x}_N + \mathbf{b}$, $A \in \mathbb{R}^{p \times p}$, A nonsingular, $\mathbf{b} \in \mathbb{R}^p$, an affine equivariant outlier identifier fulfills

$$\mathbf{x} \in \mathbf{OR}(\underline{x}_N, \alpha_N) \Leftrightarrow A\mathbf{x} + \mathbf{b} \in \mathbf{OR}(A\underline{x}_N + \mathbf{b}, \alpha_N)$$

with $A\underline{x}_N + \mathbf{b} := (A\mathbf{x}_1 + \mathbf{b}, \ldots, A\mathbf{x}_N + \mathbf{b})$. This condition holds, if one chooses $\mathbf{m}$ and S as affine equivariant estimators of location and covariance.

3 Relations between finite-sample breakdown points and the masking effect

For the comparison of outlier identification rules one can think of various different criteria (cf. Jain and Pingel, 1981, Hampel, 1985, Simonoff, 1987, Barnett and Lewis, 1994, p. 121ff., Davies and Gather, 1993, or Gather and Becker, 1997, p. 133 f., for some possibilities). One of them is the masking breakdown point which is a worst-case criterion. The possible occurrence of the so-called masking effect is a well known problem when identifying outliers. It means that it can happen that some extremely outlying observations prevent the procedure from detecting even one outlier. In Davies and Gather (1993), the masking breakdown point of a univariate outlier identifier is defined, roughly spoken, as the smallest proportion of outliers in a sample needed to create a breakdown of the procedure by the masking effect. In the multivariate case, we give the following definition:

For a sequence $\quad = (\alpha_N)_{N \in \mathbb{N}}, 0 < \alpha_N < 1,\ \delta \in (0,1)$ and regular observations $\underline{x}_n^r$ let

$$\begin{aligned} \beta^M \quad := \quad & \beta^M(\mathbf{OR}, \alpha_N, \underline{x}_n^r, k, \delta) := \inf\{\, \beta > 0 : \text{there exist } \delta \text{ outliers } \underline{x}_k^0 \text{ such that} \\ & \qquad \text{based on } \underline{x}_N = (\underline{x}_n^r, \underline{x}_k^0) \text{ some } \beta \text{ outlier will not be identified as} \\ & \qquad \alpha_N \text{ outlier by } \mathbf{OR}\}, \qquad (3) \\ k^M \quad := \quad & k^M(\mathbf{OR}, \quad, \underline{x}_n^r, \delta) := \min\{k : \beta^M(\mathbf{OR}, \alpha_{n+k}, \underline{x}_n^r, k, \delta) = 0\}. \end{aligned}$$

Then β^M is called *masking point* and

$$\varepsilon^M(\mathbf{OR}) \ := \ \varepsilon^M(\mathbf{OR}, \quad, \underline{x}_n^r, \delta) \ := \ \frac{k^M}{n + k^M}$$

is called *masking breakdown point of* $\mathbf{OR}$.

The notion of breakdown is well known in the context of robust estimation. Donoho and Huber (1983, p. 160) developed the definition of the finite-sample breakdown point of an estimator. Lopuhaä and Rousseeuw (1991, p. 231) extended the formal definition to estimators of covariance. Tyler (1994) introduces the concept of a uniform breakdown point for pairs of location and scale estimators, which is also considered by Gather and Hilker (1997). The general idea is to determine the minimum number of arbitrarily badly placed observations in a sample needed to bring the estimator beyond all bounds. Formally, this reads as follows.

Let $\underline{x}_N = (\mathbf{x}_1^r, \ldots, \mathbf{x}_N^r)$ be a sample from i.i.d. $N(\ , \Sigma)$ distributed variables. Construct $\underline{y}_{N,k} = (\mathbf{x}_{i_1}^r, \ldots, \mathbf{x}_{i_n}^r, \mathbf{y}_1, \ldots, \mathbf{y}_k)$, $\mathbf{y}_j \in \mathbb{R}^p$, $j = 1, \ldots, k$, $N = n + k$, by replacing k observations from $\underline{x}_N$ by arbitrary vectors.

First, consider a sequence $T := \{T(\underline{x}_m)\}_{m \in \mathbb{N}}$ of location estimates for . The *finite-sample breakdown point of* T is defined as

$$\varepsilon^*(\underline{x}_N, T) := \min_{1 \le k \le N} \{ \frac{k}{N} : \sup_{\underline{y}_{N,k}} \| T(\underline{x}_N) - T(\underline{y}_{N,k}) \| = \infty \}.$$

Here, $\|.\|$ denotes the euclidean norm.

Consider a sequence $C := \{C(\underline{x}_m)\}_{m \in \mathbb{N}}$ of estimators for the covariance matrix Σ. For a symmetric matrix $A \in \mathbb{R}^{p \times p}$ let $\lambda_1(A) \ge \ldots \ge \lambda_p(A)$ denote the eigenvalues, and for $A, B \in \mathbb{R}^{p \times p}$, A, B p.d., let D be defined by $D(A, B) := \max\{|\lambda_1(A) - \lambda_1(B)|, |\frac{1}{\lambda_p(A)} - \frac{1}{\lambda_p(B)}|\}$. Then

$$\varepsilon^*(\underline{x}_N, C) := \min_{1 \le k \le N} \{ \frac{k}{N} : \sup_{\underline{y}_{N,k}} D(C(\underline{x}_N), C(\underline{y}_{N,k})) = \infty \}$$

is called *finite-sample breakdown point of* C.

We will only consider estimators for which the breakdown point does not depend on the special sample but only on the sample size N. As this condition is satisfied for most of the commonly used estimators, it does not seem to be too restrictive (cf. Donoho and Huber, 1983, p. 161, Gordaliza, 1991, p. 391).

Because an outlier identifier as defined above depends on estimators of location and covariance, we may expect strong relationships between the behaviour of the estimators and the behaviour of the identifier. In the following, bounds are given on the masking-breakdown point of an identifier **OR** depending on the finite-sample breakdown points of the estimators $\mathbf{m}$ and S used in **OR**.

Theorem 3.1 *Consider an identifier* **OR**, *based on estimators* $\mathbf{m}$ *and* S *for   and* Σ. *Let* $\varepsilon^*(\underline{x}_N, \mathbf{m}) =: k_1/N$ *and* $\varepsilon^*(\underline{x}_N, S) =: k_2/N$ *denote the finite-sample breakdown points of* $\mathbf{m}$ *and* S *with* $k_i < N/2$, $i = 1, 2$. *Let further denote* $k := \min\{k_1, k_2\}$, $\quad = (\alpha_N)_{N \in I\!N}$, $0 < \alpha_N < 1$, *and* $\delta \in (0, 1)$. *Suppose that* $\underline{x}_n^r = (\mathbf{x}_1^r, \ldots, \mathbf{x}_n^r)$ *is a sample of regular observations from* $N(\ \ , \Sigma)$. *Then*

$$\varepsilon^M(\mathbf{OR}, \quad , \underline{x}_n^r, \delta) \geq \frac{k}{N},$$

where $N = n + k$.

The proof of this theorem is given in the Appendix. Theorem 3.1 gives a lower bound for the masking breakdown point. In a similar way we can derive an upper bound.

Theorem 3.2 *Assume the conditions of Theorem 3.1 with* $K := \max\{k_1, k_2\}$. *Then*

$$\varepsilon^M(\mathbf{OR}, \quad , \underline{x}_n^r, \delta) \leq \frac{K}{N},$$

where $N = n + K$.

The proof is given in the Appendix. From Theorems 3.1 and 3.2 it can be seen that the finite-sample breakdown points of the estimators represent bounds for the masking breakdown point of the resulting outlier identifier. Therefore, the masking breakdown point equals the finite-sample breakdown points if $\varepsilon^*(\underline{x}_N, \mathbf{m})$ and $\varepsilon^*(\underline{x}_N, S)$ coincide. Together, the theorems can be used to derive the maximum possible masking breakdown point. This is done for samples where the regular observations are in general position. A p-variate sample is said to be *in general position* if no more than p points of the sample lie in any $(p-1)$-dimensional subspace of $I\!R^p$ (cf. Rousseeuw, 1985, S. 288).

Theorem 3.3 *Suppose that, under the conditions of Theorem 3.1, the sample* $\underline{x}_n^r$ *of regular observations is in general position, and that* $n \geq p + 1$. *Then*

$$\frac{[(N-p+1)/2]}{N} \leq \varepsilon^M_{\max} \leq \frac{1}{2},$$

where $\varepsilon^M_{\max} = \max_{\mathbf{OR}} \varepsilon^M(\mathbf{OR})$ *and* $[x]$ *denotes the integer part of* $x \in I\!R$.

The proof is given in the Appendix. From the results of the above theorems it becomes clear that the use of high-breakdown estimators such as certain S-estimators in outlier identifiers will give the best possible protection against the masking effect. Rousseeuw and Yohai (1984) introduce S-estimators in the context of robust regression. Davies (1987) extends the definition and derives a pair $(\mathbf{m}, S)$ of S-estimates for multivariate location and covariance, showing that the **m**aximum attainable **b**reakdown point for $(\mathbf{m}, S)$ is $[(N-p+1)/2]/N$, in which case we will denote them as $\mathrm{S_{MB}}$-estimators.

Corollary 3.1 *Under the conditions of Theorem 3.3 let* $\mathbf{OR}_{\mathrm{S_{MB}}}$ *be an outlier identifier based on* S_{MB}*-estimators for location and covariance. Then*

$$\varepsilon^M(\mathbf{OR}_{\mathrm{S_{MB}}}) \quad = \quad \frac{[(N-p+1)/2]}{N}.$$

The same high masking breakdown point is attained by using the minimum volume ellipsoid estimators introduced by Rousseeuw (1985) if one chooses the number h of data points on which the ellipsoid is based according to $h = [(N+p+1)/2]$, in which case the estimators have the best possible finite-sample breakdown points. This leads to a similar identification procedure as introduced by Rousseeuw and van Zomeren (1990). The difference lies in the normalizing condition: In this paper, we adjust the critical value to the sample size.

The use of $\mathrm{S_{MB}}$-estimators for the identification of outliers is illustrated by the following example.

Example 3.1 *The identifier* $\mathbf{OR}_{\mathrm{BW}}$ *based on Tukey's biweight (Beaton, Tukey, 1974; also cf. Rocke, 1996) is given by*

$$\mathbf{OR}_{\mathrm{BW}} := \{\mathbf{x} \in \mathbb{R}^p : (\mathbf{x} - \mathbf{m}_{\mathrm{BW}})^T S_{\mathrm{BW}}^{-1}(\mathbf{x} - \mathbf{m}_{\mathrm{BW}}) \geq c_{\mathrm{BW}}(p, N, \alpha_N)\},$$

where $\mathbf{m}_{\mathrm{BW}}$ *and* S_{BW} *are solutions of the minimization problem*

$$\min_{S \in PDS(p)} \det(S)$$

under the restriction

$$1/N \sum_{i=1}^{N} \rho\left(\sqrt{(\mathbf{X}_i - \mathbf{m})^T S^{-1}(\mathbf{X}_i - \mathbf{m})}\right) = b_0.$$

Table 1: Observation distances with respect to $\mathbf{m}_{\mathrm{BW}}$ and S_{BW}

Observation	Distance	Observation	Distance
1	**51.22**	11	1.44
2	22.30	12	1.78
3	**41.44**	13	3.51
4	**38.49**	14	2.36
5	0.78	15	1.41
6	1.16	16	1.55
7	1.52	17	2.79
8	1.98	18	1.03
9	1.13	19	1.30
10	2.03	20	2.35
		21	**32.19**

Here, $\rho = \rho_{\mathrm{BW}} : \mathbb{R}_+ \mapsto \mathbb{R}$*:*

$$\rho_{\mathrm{BW}}(d) = \begin{cases} d^2/2 - d^4/(2c_0^2) + d^6/(6c_0^4) & , \quad 0 \le d \le c_0 \\ c_0^2/6 & , \quad d > c_0 \end{cases},$$

$c_0 \in \mathbb{R}$ *such that the finite-sample breakdown point of* S_{BW} *is maximal. That means* c_0 *solves the equation* $E(\rho(D)) = r\rho(c_0)$*, where* $r = [(N-p+1)/2]/N$ *and* D *is a random variable with* $D^2 \sim \chi^2_p$*. The value* b_0 *is determined by* $E(\rho(D)) = b_0$ *(cf. Lopuhaä, 1989, Rocke, 1996).*

The constant $c_{\mathrm{BW}}(p, N, \alpha_N)$ *is calculated by simulation from the normalizing condition (2), where we choose* $\alpha = 0.1$*.*

We consider the data set known as "stackloss data" (Brownlee, 1965, p. 454) which is often investigated in the context of robust regression and outlier identification. The data come from an experiment for the oxidation of ammonia into nitric acid. Four variables are recorded: rate of incoming ammonia, cooling water temperature, acid concentration, and stackloss. In the regression approach, the stackloss is regarded as the dependent variable which has to be explained by means of the remaining variables. The observations can also be regarded as an unstructured multivariate data set. In this case, we have a sample of size $N = 21$ *with* $p = 4$ *variables in which we are searching for outliers. The value of the normalizing constant* c *for the identification procedure is then given as* $c_{\mathrm{BW}}(4, 21, 0.0050) = 31.57$*.*

Table 2: Observation distances with respect to $\overline{\mathbf{x}}_N$ and S_N

Observation	Distance	Observation	Distance
1	6.56	11	3.07
2	6.11	12	4.47
3	5.10	13	2.55
4	5.51	14	3.32
5	0.44	15	3.65
6	1.69	16	1.85
7	4.27	17	7.93
8	3.83	18	2.40
9	3.10	19	2.71
10	3.39	20	0.92
		21	11.13

Table 1 shows the (squared) distance $(\mathbf{x}_i - \mathbf{m}_{\mathrm{BW}})^T S_{\mathrm{BW}}^{-1}(\mathbf{x}_i - \mathbf{m}_{\mathrm{BW}})$ *for each observation* $\mathbf{x}_i$*,* $i = 1, \ldots, 21$*, of the data set.*

The identifier $\mathbf{OR}_{\mathrm{BW}}$ *declares four observations as* α_N *outliers, namely* x_1, x_3, x_4, x_{21}*, with* $\alpha_N = 0.0050$*. Most authors who investigated this often analysed data agree on that observations 3, 4 and 21 have to be regarded as outliers (cf. also Rousseeuw and Leroy, 1987, p. 76f.). Daniel, Wood (1971), Li (1985) and Andrews (1974) identify the same four observations like* $\mathbf{OR}_{\mathrm{BW}}$*, whereas Carroll and Ruppert (1985) declare observations 2, 3, 4, 21 as conspicuous. Dempster and Gasko-Green (1981) as well as Andrews and Pregibon (1978) even detect five outliers (1, 2, 3, 4, 21). Although observation 2 shows a relatively high distance, its identification is not justified by the robust procedure* $\mathbf{OR}_{\mathrm{BW}}$*.*

With the results of our outlier identifier, the interpretation of observation 21 becomes somewhat different to that of the previous investigations. All authors agree in regarding this observation as the clearest outlier. In Table 1 it can be seen that we cannot support this interpretation. But if we use the non-robust estimators $\overline{\mathbf{x}}_N = {}_{i=1}^{N} \mathbf{x}_i/N$ *and* $S_N = {}_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}}_N)^2/(N-1)$ *instead of* $\mathbf{m}_{\mathrm{BW}}$ *and* S_{BW}*, we find a similar behaviour of the resulting procedure as in the above mentioned investigations. Due to the masking effect, observation 21 has then the largest distance of all observations (see Table 2), even though it does not exceed the respective critical value* $\mathrm{c}(4, 21, 0.0050) = 14.86$*. At the same time, the distances of observations 1, 3, and 4 are not exceptionally large. Therefore, it can be* *concluded that the strong outliers (1, 3, 4) do not only mask themselves but also cause the impression that the less differing observation 21 is the most conspicuous one. The similarity of this situation to the behaviour of the above mentioned procedures leads to the conclusion that those procedures are still influenced by the strong outliers (1, 3, 4) when they make out observation 21 as the clearest outlier. In contrast to this,* $\mathbf{OR}_{\mathrm{BW}}$ *is less influenced by observations 1, 3, and 4 and therefore does not identify observation 21 as the most conspicuous one.*

Appendix: Proofs

Proof of Theorem 3.1: Consider a situation with $k-1$ outliers. Let $\underline{x}_{N-1} = (\underline{x}_n^r, \underline{x}_{k-1}^0)$ and $\underline{x}_{k-1}^0 = (\mathbf{x}_1^0, \ldots, \mathbf{x}_{k-1}^0)$ be an arbitrary constellation of δ outliers. With $k < N/2$ we have $(k-1)/(N-1) < k/N$. Therefore, neither $\mathbf{m}$ nor S can break down. From this it follows

$$\mathbf{OR}(\underline{x}_{N-1}, \alpha_{N-1}) \neq \emptyset, \qquad 0 < \mathrm{volume}(\mathbb{R}^p \backslash \mathbf{OR}) < \infty,$$

and there exists a sphere Sp with radius r, $0 < r < \infty$, such that $\mathbb{R}^p \backslash \mathbf{OR} \subseteq Sp$. For example choose $r = \|\mathbf{m}(\underline{x}_{N-1})\| + const \ \ \overline{\lambda_1(S)}$ where the constant is a factor of proportionality, because the squared volume of the ellipsoid $\mathbb{R}^p \backslash \mathbf{OR}$ is proportional to the product of the eigenvalues of S.

It now follows that

$$\mathbb{R}^p \backslash Sp \ \subseteq \mathbf{OR}(\underline{x}_{N-1}, \alpha_{N-1}),$$

thus all points outside the sphere are identified as α_{N-1} outliers.

Now there exists some $\beta \in \ (0,1)$, such that

$$Sp \ \subseteq \mathbb{R}^p \backslash \mathrm{out}(\beta, \quad, \Sigma) = \{\mathbf{x} \in \mathbb{R}^p : (\mathbf{x} - \quad)^T \Sigma^{-1} (\mathbf{x} - \quad) \leq \chi^2_{p;1-\beta}\}.$$

The maximal value β fulfilling this relation is denoted by β^*. Thus, we have

$$\mathrm{out}(\beta^*, \quad, \Sigma) \quad \subseteq \quad \mathbb{R}^p \backslash Sp \quad \subseteq \quad \mathbf{OR}(\underline{x}_{N-1}, \alpha_{N-1}),$$

that means, every β^* outlier is identified as an α_{N-1} outlier. The same statement holds for all $\beta < \beta^*$.

Together with (3) this yields $\beta^M(\alpha_{N-1}, \underline{x}_n^r, k-1, \delta) \geq \beta^* > 0$.

The same steps are possible for all j with $0 \leq j < k$ instead of $k-1$. Therefore,

$$\varepsilon^M(\mathbf{OR}(\underline{x}_{N-1}, \alpha_{N-1})) \; > \; \frac{k-1}{n+k-1},$$

and

$$\varepsilon^M(\mathbf{OR}(\underline{x}_N, \alpha_N)) \; \geq \; \frac{k}{n+k} = \frac{k}{N}.$$

Proof of Theorem 3.2: Assume $\varepsilon^M(\mathbf{OR}) > K/N$, that means $\varepsilon^M(\mathbf{OR}) \geq (K+1)/(N+1) = (K+1)/(n+K+1)$. Together with the definition of ε^M it follows that $k^M \geq K+1$. Then there must exist some $\beta^* > 0$ with

$$\beta^M(\alpha_{n+K}, \underline{x}_n^r, K, \delta) > \beta^*$$

for arbitrary constellations of K observations which are placed as δ outliers.

Now, because of (3), for any constellation $\underline{x}_K^0$ of δ outliers all points in out$(\beta^*, \quad, \Sigma)$ are identified as α_{n+K} outliers. This means that

$$\mathbb{R}^p \backslash \mathbf{OR}(\underline{x}_N, \alpha_N) \subseteq \mathbb{R}^p \backslash \text{out}(\beta^*, \quad, \Sigma)$$

for arbitrary $\underline{x}_K^0$. With this relation, the center of the ellipsoid $\mathbb{R}^p \backslash \mathbf{OR}(\underline{x}_N, \alpha_N)$ must lie within a closed subset of $\mathbb{R}^p$. On the other hand, the center is $\mathbf{m}(\underline{x}_N)$. From this it follows, that $\mathbf{m}$ will not break down for any constellation $\underline{x}_K^0$, thus $\varepsilon^*(\underline{x}_N, \mathbf{m}) > K/N$. But this contradicts the assumption on $\varepsilon^*(\underline{x}_N, \mathbf{m})$, finishing the proof.

Proof of Theorem 3.3: The proof follows immediately from Theorems 3.1 and 3.2, using the following results for the finite-sample breakdown points of affine equivariant estimators $\mathbf{m}$ and S:

$$\frac{[(N-p+1)/2]}{N} \quad \leq \quad \max_{\mathbf{m}} \varepsilon^*(\underline{x}_N, \mathbf{m}) \quad \leq \quad \frac{1}{2}$$

(Lopuhaä, Rousseeuw, 1991),

$$\max_{S} \varepsilon^*(\underline{x}_N, S) \quad = \quad \frac{[(N-p+1)/2]}{N}$$

(Davies, 1987).

References

Andrews, D.F. (1974), "A Robust Method for Multiple Linear Regression," *Technometrics*, 16, 523–531.

Andrews, D.F., Pregibon, D. (1978), "Finding the Outliers that Matter," *Journal of the Royal Statistical Society, Ser. B*, 44, 1–36.

Atkinson, A.C., Mulira, H.-M. (1993), "The Stalactite Plot for the Detection of Multivariate Outliers," *Statistics and Computing*, 3, 27–35.

Bacon–Shone, J., Fung, W.K. (1987), "A New Graphical Method for Detecting Single and Multiple Outliers in Univariate and Multivariate Data," *Applied Statistics*, 36, 153–162.

Barnett, V., Lewis, T. (1994), *Outliers in Statistical Data* (3rd ed.), New York: Wiley.

Beaton, A.E., Tukey, J.W. (1974), "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data," *Technometrics*, 16, 147–185.

Bhandary, M. (1992), "Detection of the Numbers of Outliers Present in a Data Set Using an Information Theoretic Criterion," *Communications in Statistics – Theory and Methods*, 21, 3263–3274.

Brownlee, K.A. (1965), *Statistical Theory and Methodology in Science and Engineering* (2nd ed.), New York: Wiley.

Caroni, C., Prescott, P. (1992), "Sequential Application of Wilks's Multivariate Outlier Test," *Applied Statistics*, 41, 355–364.

Carroll, R.J., Ruppert, D. (1985), "Transformations in Regression: A Robust Analysis," *Technometrics*, 27, 1–12.

Daniel, C., Wood, F.S. (1971), *Fitting Equations to Data*, New York: Wiley.

Davies, P.L. (1987), "Asymptotic Behaviour of S-Estimates of Multivariate Location Parameters and Dispersion Matrices," *The Annals of Statistics*, 15, 1269–1292.

Davies, P.L., Gather, U. (1993), "The Identification of Multiple Outliers," *Journal of the American Statistical Association*, 88, 782–792.

Dempster, A.P., Gasko-Green, M. (1981), "New Tools for Residual Analysis," *The Annals of Statistics*, 9, 945–959.

Donoho, D.L., Huber, P.J. (1983), "The Notion of Breakdown Point," in *A Festschrift for Erich L. Lehmann*, eds. P.J. Bickel, K.A. Doksum, and J.L. Hodges, Jr., Belmont, CA: Wadsworth, pp. 157–184.

Gather, U., Becker, C. (1997), "Outlier Identification and Robust Methods," in *Handbook of Statistics, Vol. 15: Robust Inference*, eds. G.S. Maddala, and C.R. Rao, Amsterdam: Elsevier, pp. 123–143.

Gather, U., Hilker, T. (1997), "A Note on Tyler's Modification of the MAD for the Stahel-Donoho Estimator," *The Annals of Statistics*, 25(5).

Gnanadesikan, R., Kettenring, J.R. (1972), "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data," *Biometrics*, 28, 81–124.

Gordaliza, A. (1991), "On the Breakdown Point of Multivariate Location Estimators Based on Trimming Procedures," *Statistics & Probability Letters*, 11, 387–394.

Hampel, F.R. (1985), "The Breakdown Points of the Mean Combined With Some Rejection Rules," *Technometrics*, 27, 95–107.

Hara, T. (1988), "Detection of Multivariate Outliers with Location Slippage or Scale Inflation in Left Orthogonally Invariant or Elliptically Contoured Distributions," *Annals of the Institute of Statistical Mathematics*, 40, 395–406.

Hawkins, D.M. (1980), *Identification of Outliers*, London: Chapman and Hall.

Jain, R.B., Pingel, L.A. (1981), "On the Robustness of Recursive Outlier Detection Procedures to Nonnormality," *Communications in Statistics – Theory and Methods*, 10, 1323–1334.

Li, G. (1985), "Robust Regression," in *Exploring Data Tables, Trends, and Shapes*, eds. D. Hoaglin, F. Mosteller, and J. Tukey, New York: Wiley, pp. 281–343.

Lopuhaä, H.P. (1989), "On the Relation Between S-Estimators and M-Estimators of Multivariate Location and Covariance," *The Annals of Statistics*, 17, 1662–1683.

Lopuhaä, H.P., Rousseeuw, P.J. (1991), "Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices," *The Annals of Statistics*, 19, 229–248.

Rocke, D.M. (1996), "Robustness Properties of S-Estimators of Multivariate Location and Shape in High Dimension," *The Annals of Statistics*, 24, 1327–1345.

Rousseeuw, P.J. (1985), "Multivariate Estimation with High Breakdown Point," in *Mathematical Statistics and Applications*, eds. W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Dordrecht: Reidel, pp. 283–297.

Rousseeuw, P.J., Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, New York: Wiley.

Rousseeuw, P.J., Yohai, V. (1984), "Robust Regression by Means of S-Estimators," in *Robust and Nonlinear Time Series Analysis, Lecture Notes in Statistics*, 26, New York: Springer, pp. 256–272.

Rousseeuw, P.J., van Zomeren, B.C. (1990), "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association*, 85, 633–639.

Simonoff, J.S. (1987), "The Breakdown and Influence Properties of Outlier Rejection-Plus-Mean Procedures," *Communications in Statistics A*, 16, 1749–1760.

Tyler, D.E. (1994), "Finite Sample Breakdown Points of Projection Based Multivariate Location and Scatter Statistics," *The Annals of Statistics*, 22, 1024–1044.

Wilks, S.S. (1963), "Multivariate Statistical Outliers," *Sankhyā A*, 25, 407–426.