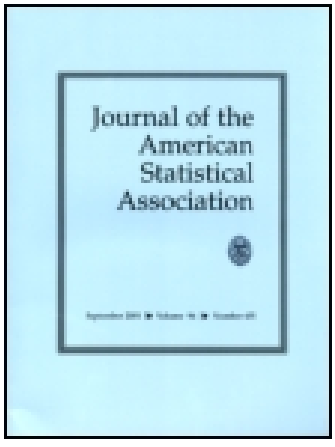

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

The Masking Breakdown Point of Multivariate Outlier Identification Rules

Claudia Becker [a] & Ursula Gather [a]
[a] Department of Statistics , University of Dortmund , D-44221 , Dortmund , Germany
Published online: 17 Feb 2012.

To cite this article: Claudia Becker & Ursula Gather (1999) The Masking Breakdown Point of Multivariate Outlier Identification Rules, Journal of the American Statistical Association, 94:447, 947-955

To link to this article: http://dx.doi.org/10.1080/01621459.1999.10474199

The Masking Breakdown Point of Multivariate Outlier Identification Rules

Claudia BECKER and Ursula GATHER

In this article, we consider simultaneous outlier identification rules for multivariate data, generalizing the concept of so-called α outlier identifiers, as presented by Davies and Gather for the case of univariate samples. Such multivariate outlier identifiers are based on estimators of location and covariance. Therefore, it seems reasonable that characteristics of the estimators influence the behavior of outlier identifiers. Several authors mentioned that using estimators with low finite-sample breakdown point is not recommended for identifying outliers. To give a formal explanation, we investigate how the finite-sample breakdown points of estimators used in these identification rules influence the masking behavior of the rules.

KEY WORDS: Breakdown point; Masking; Outlier identification; Robust statistics.

1. INTRODUCTION

It is well known that outliers (i.e., observations lying "far away" from the main part of a dataset 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 dataset for "bad" observations, thus avoiding distortion of the statistical analysis. But outliers can be of fundamental interest in themselves, and thus their identification should also be considered as a goal in itself. That this goal is not always easy to reach may be seen in the following examples.

Example 1. Consider the simple case of a univariate dataset of observations from the standard normal $N(0,1)$,

$$\begin{array}{cccccccccc} -2.21 & -1.84 & -.95 & -.91 & -.36 & -.19 & -.11 & -.10 & .18 & .30 \\ .31 & .43 & .51 & .64 & .67 & .72 & 1.22 & 1.35 & 8.10 & 17.60, \end{array}$$

where the original observations .81 and 1.76 have been erroneously misprinted as 8.10 and 17.60. From inspection, it is obvious that these two observations should be regarded as outliers. A possible (automatic) outlier identification procedure for this case would be to calculate $t(x_i) = |x_i - \overline{x}_N|/s_N$ for each observation x_i, $i = 1, \ldots, N$, where $\overline{x}_N = \sum_{i=1}^N x_i/N$ and $s_N = (\sum_{i=1}^N (x_i - \overline{x}_N)^2/(N-1))^{1/2}$. Any observation x_i with $t(x_i) > u_{1-\alpha}$, the $(1-\alpha)$ quantile of $N(0,1)$, would then be declared as outlying. For the foregoing data, we get $t(8.10) = 1.572$. With $\alpha = .05$, for example, we see that $u_{1-\alpha} = 1.6449$, such that the observation 8.10 will not be found to be an outlier. This gives an example of the well-known masking effect: The very large observation 17.60 influences both $\overline{x}_N$ and s_N, and thus it masks the smaller outlier and prevents its identification (also see Barnett and Lewis 1994, p. 90).

The situation becomes even worse in multivariate situations where the masking effect also occurs but the supplementary visual inspection of the dataset becomes much harder and detection of outliers by pure vision is almost impossible, because they do not "stick out on the end" (Gnanadesikan and Kettenring 1972, p. 109) as in univariate settings. Observations that are not conspicuous in any single variable nevertheless may differ clearly from the rest of the data if all variables are looked at simultaneously (see Rousseeuw and Leroy 1987, p. 7, for an example). Therefore, automatic rules for the identification of multivariate outliers that are resistant toward the masking effect are needed.

Various concepts for outlier identification in multivariate samples exist in the literature. Among them are methods of heuristic nature (e.g., Atkinson and Mulira 1993; Bacon-Shone and Fung 1987; Barnett and Lewis 1994, p. 307; Bhandary 1992) and those of the consecutive testing type (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 simultaneous outlier identification rules as considered by, for example, Barnett and Lewis (1994, p. 306), Gnanadesikan and Kettenring (1972), Healy (1968), Rousseeuw and Leroy (1987, p. 266), and Rousseeuw and van Zomeren (1990).

Example 2. We consider data from a study on the relation between air pollution and mortality. The original aim was to constitute a regression model. Here we regard the observations as a multivariate dataset. The data were collected from standard metropolitan statistical areas in the United States and consist of values of age-adjusted mortality and 15 variables on climate, air pollution, and some socioeconomic conditions. The complete dataset can be found in the Data and Story Library (*http://lib.stat.cmu.edu/DASL/*). Here we concentrate on mortality and 12 of the variables: mean January temperature, mean July temperature, relative humidity, annual rainfall, median education, population density, population, population per household, median income, HC pollution, NO_x pollution, and SO_2 pollution. From the 60 areas in the dataset, we exclude Fort Worth (TX) because of missing values. The remaining data then consist of $N = 59$ observations in $p = 13$ variables. We

Claudia Becker is Research Assistant and Ursula Gather is Professor, Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany (E-mail: *gather@amadeus.statistik.uni-dortmund.de*). This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

are now interested in whether there are "extreme" areas with respect to these variables; that is, we are looking for outliers in the data. The multivariate outlier identification procedure corresponding to the univariate method of Example 1 goes back to Healy (1968), who identified an observation $\mathbf{x}_i$ as an outlier if its (squared) Mahalanobis distance $(\mathbf{x}_i - \bar{\mathbf{x}}_N)^T \mathbf{S}_N^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_N)$ exceeds some appropriate critical value c. Here $\bar{\mathbf{x}}_N = \sum_{i=1}^N \mathbf{x}_i/N$ denotes the arithmetic mean and $\mathbf{S}_N = \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}}_N)(\mathbf{x}_i - \bar{\mathbf{x}}_N)^T/(N-1)$ denotes the sample covariance. Based on the asymptotic distribution of the Mahalanobis distance, the critical value c should be chosen as $\chi^2_{p;1-\alpha/N}$, the $(1 - \alpha/N)$ quantile of the χ^2_p distribution for some given $\alpha \in (0,1)$. For smaller samples, results of Barnett and Lewis (1994), Jennings and Young (1988), Penny (1996), and Wilks (1963), show that $c = (p(N-1)^2 F_{p,N-p-1;1-\alpha/N})/(N(N-p-1+pF_{p,N-p-1;1-\alpha/N}))$ is an appropriate choice. We adopt this choice of c here, as 59 observations may not be seen as a large sample in 13 dimensions. With $\alpha = .1$, we get $c = 27.532$. Table 1 shows the (squared) distance for each observed region. We see that observations 18, 28, 31, 37, 47, 48, and 58 are declared to be outliers by this procedure.

Example 1 shows that the estimators $\bar{x}_N$ and s_N are both prone to the influence of outlying observations. The same is true for the multivariate $\bar{\mathbf{x}}_N$ and $\mathbf{S}_N$. Therefore, we may ask whether the masking effect also occurs in Example 2. To work this out, we investigate in which sense the estimators affect the masking behavior of outlier identification procedures of the aforementioned type.

Our article is organized as follows: In Section 2 we make precise in a formal way how we understand the task of outlier identification and give a definition of a multivariate outlier identifier. For comparing the masking behavior of such identifiers, a suitable performance criterion is 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 on which the identification rule is based. Finally, we return to the question whether there are further outliers in the data of Example 2 and investigate the behavior of the outlier identifier proposed here in further examples.

2. MULTIVARIATE OUTLIER IDENTIFICATION

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

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

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

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

For usual choices of α ($\alpha = .05, \alpha = .1$), this reflects the idea of an outlier being an observation that is rather un-

Table 1. Observation Distances With Respect to $\bar{\mathbf{x}}_N$ and $\mathbf{S}_N$

Observation	Region	Distance
1	Akron, OH	3.50
2	Albany-Schenectady-Troy, NY	7.56
3	Allentown, Bethlehem, PA-NJ	5.09
4	Atlanta, GA	5.49
5	Baltimore, MD	10.16
6	Birmingham, AL	9.03
7	Boston, MA	16.98
8	Bridgeport-Milford, CT	24.40
9	Buffalo, NY	9.83
10	Canton, OH	3.01
11	Chattanooga, TN-GA	9.09
12	Chicago, IL	26.24
13	Cincinnatti, OH-KY-IN	4.47
14	Cleveland, OH	3.94
15	Columbus, OH	7.59
16	Dallas, TX	14.41
17	Dayton-Springfield, OH	2.76
18	**Denver, CO**	**36.88**
19	Detroit, MI	9.44
20	Flint, MI	6.02
21	Grand Rapids, MI	5.27
22	Greensboro-Winston-Salem-High Point, NC	6.03
23	Hartford, CT	4.79
24	Houston, TX	12.06
25	Indianapolis, IN	3.20
26	Kansas City, MO	11.29
27	Lancaster, PA	12.95
28	**Los Angeles, Long Beach, CA**	**52.16**
29	Louisville, KY-IN	12.16
30	Memphis, TN-AR-MS	9.46
31	**Miami-Hialeah, FL**	**29.24**
32	Milwaukee, WI	7.36
33	Minneapolis-St. Paul, MN-WI	13.22
34	Nashville, TN	5.03
35	New Haven-Meriden, CT	2.97
36	New Orleans, LA	19.64
37	**New York, NY**	**32.14**
38	Philadelphia, PA-NJ	9.30
39	Pittsburgh, PA	22.78
40	Portland, OR	22.22
41	Providence, RI	5.06
42	Reading, PA	9.19
43	Richmond-Petersburg, VA	6.06
44	Rochester, NY	3.54
45	St. Louis, MO-IL	10.80
46	San Diego, CA	18.28
47	**San Francisco, CA**	**35.19**
48	**San Jose, CA**	**29.33**
49	Seattle, WA	14.96
50	Springfield, MA	5.49
51	Syracuse, NY	5.26
52	Toledo, OH	5.44
53	Utica-Rome, NY	7.78
54	Washington, DC-MD-VA	19.46
55	Wichita, KS	14.36
56	Wilmington, DE-NJ-MD	14.09
57	Worcester, MA	17.74
58	**York, PA**	**32.11**
59	Youngstown-Warren, OH	3.70

NOTE: Observations identified as outliers are represented by boldface.

likely 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 $\text{out}(\alpha_N, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where, as in the univariate case, α_N can be chosen according to the condition

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

for $\mathbf{X}_i \sim \text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \ldots, N$, 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. Because the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ 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 that 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 iid $\text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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 $\text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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

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

where $\mathbf{S} = \mathbf{S}(\underline{x}_N) \in \text{PDS}(p)$, $\mathbf{m} = \mathbf{m}(\underline{x}_N) \in \mathbb{R}^p$, 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 $\underline{x}_N$. All points $\mathbf{x} \in \text{OR}(\underline{x}_N, \alpha_N)$ are identified as α_N outliers with respect to $\text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here $\text{PDS}(p) = \{\mathbf{S} \in \mathbb{R}^{p \times p}$: $\mathbf{S}$ pd and symmetric$\}$, and $c(p, N, \alpha_N)$ is a normalizing constant.

Several outlier identification procedures of this type exist in the statistical literature, using various estimators $\mathbf{m}$ and $\mathbf{S}$; often they are not presented in a strictly formalized way. The classical outlier identification rule in the multivariate normal setting as described in Example 2 yields the corresponding classical outlier identifier, which we denote by OR_{MD}, with $\mathbf{m} = \bar{\mathbf{x}}_N$ and $\mathbf{S} = \mathbf{S}_N$. Other outlier identification procedures of a similar kind have been given by, for example, Barnett and Lewis (1994, p. 306), Gnanadesikan and Kettenring (1972, sec. 4), Rousseeuw and Leroy (1987, pp. 266, 269–270), and Rousseeuw and van Zomeren (1990).

Several normalizing conditions are possible to fix c appropriately. As mentioned earlier, we are interested in estimating the α_N outlier region $\text{out}(\alpha_N, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $\alpha_N = 1 - (1 - \alpha)^{1/N}$. Therefore, we choose the same value of α_N to be used in the outlier identifier $\text{OR}(\underline{x}_N, \alpha_N)$, viewed as an estimator for $\text{out}(\alpha_N, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. The choice of α_N for the outlier region guarantees that under the model, with probability $1 - \alpha$, no observation lies in this region; see (1). Analogously to this and in accordance with the work of Davies and Gather (1993), we fix the normalizing constant c of the identifier such that for a sample of size N coming iid from the p-variate normal, with probability $1 - \alpha$, no observation will be identified as an outlier. We thus restrict ourselves to the condition

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

for $\alpha \in (0, 1)$ and $\alpha_N = 1 - (1 - \alpha)^{1/N}$, where $\underline{X}_N = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$, $\mathbf{X}_1, \ldots, \mathbf{X}_N$ iid according to $\text{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

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

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

with $\mathbf{A}\underline{x}_N + \mathbf{b} := (\mathbf{A}\mathbf{x}_1 + \mathbf{b}, \ldots, \mathbf{A}\mathbf{x}_N + \mathbf{b})$. This condition holds if one chooses $\mathbf{m}$ and $\mathbf{S}$ 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 criteria (for some possibilities see Barnett and Lewis 1994, p. 121; Davies and Gather 1993; Gather and Becker 1997, p. 133; Hampel 1985; Jain and Pingel 1981; Simonoff 1987). One of these is the masking breakdown point, which is a worst-case criterion. The possible occurrence of the so-called masking effect, as already shown in Example 1, is a well-known problem when identifying outliers. It means that some extremely outlying observations can prevent the procedure from detecting even one outlier. Davies and Gather (1993) defined the masking breakdown point of a univariate outlier identifier, roughly spoken, as the smallest proportion of outliers in a sample needed to create a breakdown of the procedure by the masking effect. For the multivariate case, we give the following definition.

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

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

and

$$\begin{aligned} k^M &:= k^M(\text{OR}, \boldsymbol{\alpha}, \underline{x}_n^r, \delta) \\ &:= \min\{k: \beta^M(\text{OR}, \alpha_{n+k}, \underline{x}_n^r, k, \delta) = 0\}. \end{aligned} \quad (4)$$

Then β^M is called the *masking point*, and

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

is called the *masking breakdown point of* 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) introduced the concept of a uniform breakdown point for pairs of location and scale estimators; this was 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 $\underline{x}_N = (\mathbf{x}_1^\tau, \ldots, \mathbf{x}_N^\tau)$ be a sample from iid $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distributed variables. Construct $\underline{y}_{N,k} = (\mathbf{x}_{i_1}^\tau, \ldots, \mathbf{x}_{i_n}^\tau, \mathbf{y}_1, \ldots, \mathbf{y}_k)$, $\mathbf{y}_j \in \mathbb{R}^p$, $j = 1, \ldots, k$, $N = n + k$, by replacing k observations from $\underline{x}_N$ by arbitrary vectors. First, consider a sequence $\mathbf{T} := \{\mathbf{T}(\underline{x}_m)\}_{m \in \mathbb{N}}$ of location estimates for $\boldsymbol{\mu}$. The *finite-sample breakdown point of* $\mathbf{T}$ is defined as

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

Here $\|\cdot\|$ denotes the Euclidean norm.

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

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

is called *finite-sample breakdown point of* $\mathbf{C}$.

We consider only 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 commonly used estimators, it does not seem too restrictive (Donoho and Huber 1983, p. 161; Gordaliza 1991, p. 391).

Because an outlier identifier as defined previously depends on estimators of location and covariance, we may expect strong relationships between the behavior of the estimators and the behavior of the identifier, as has been pointed out by several authors (e.g., Simonoff 1987). 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 $\mathbf{S}$ used in OR.

Theorem 1. Consider an affine equivariant identifier OR, based on estimators $\mathbf{m}$ and $\mathbf{S}$ for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Let $\varepsilon^*(\underline{x}_N, \mathbf{m}) =: k_1/N$ and $\varepsilon^*(\underline{x}_N, \mathbf{S}) =: k_2/N$ denote the finite-sample breakdown points of $\mathbf{m}$ and $\mathbf{S}$ with $k_i < N/2$, $i = 1, 2$. Further, let denote $k := \min\{k_1, k_2\}$, $K := \max\{k_1, k_2\}$, $\boldsymbol{\alpha} = (\alpha_N)_{N \in \mathbb{N}}$, $0 < \alpha_N < 1$, and $\delta \in (0, 1)$. Suppose that $\underline{x}_n^\tau = (\mathbf{x}_1^\tau, \ldots, \mathbf{x}_n^\tau)$ is a sample of regular observations from $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$\frac{k}{N} \le \varepsilon^M(\mathrm{OR}, \boldsymbol{\alpha}, \underline{x}_n^\tau, \delta) \le \frac{K}{N}.$$

The proof of this theorem is given in the Appendix.

In the general case, it is not possible to tighten the upper bound, due to the breakdown behavior of covariance estimators. On the one hand, an estimator $\mathbf{S}$ breaks down if its largest eigenvalue tends to infinity ("explosion" of the estimator). For OR, this corresponds to the ellipsoid $\mathbb{R}^p \backslash \mathrm{OR}$ growing along its largest axis. In this case there will be arbitrarily large outliers lying in the region $\mathbb{R}^p \backslash \mathrm{OR}$. These outliers cannot be identified, leading to the masking effect. But on the other hand, $\mathbf{S}$ could also break down because its smallest eigenvalue tends to 0. In this case ("implosion" of $\mathbf{S}$), the ellipsoid $\mathbb{R}^p \backslash \mathrm{OR}$ will shrink along its smallest axis and thus degenerate to a $(p-1)$-dimensional structure. This leads to swamping rather than masking, meaning that nonoutliers are falsely identified as outliers. This is the reason why tightening the upper bound generally is not possible. But if the finite-sample breakdown point of the location estimator is smaller than that of the covariance estimate, then we have the following theorem (the proof of which is given in the Appendix).

Theorem 2. Under the conditions of Theorem 1, let $k_1 < k_2$. Then

$$\varepsilon^M(\mathrm{OR}) \le \frac{k_1}{N},$$

and together with Theorem 1, it follows that

$$\varepsilon^M(\mathrm{OR}) = \varepsilon^*(\underline{x}_N, \mathbf{m}) = \frac{k_1}{N}.$$

From Theorem 1, 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 $\varepsilon^*(\underline{x}_N, \mathbf{S})$ coincide. From this, we can calculate the masking breakdown point of the classical outlier identifier $\mathrm{OR}_{\mathrm{MD}}$, based on $\bar{\mathbf{x}}_N$ and $\mathbf{S}_N$. It is well known that the finite-sample breakdown points of these estimators equal $1/N$ (Donoho and Huber 1983); thus we have $\varepsilon^M(\mathrm{OR}_{\mathrm{MD}}) = 1/N$, which is the lowest possible breakdown value. This gives a formal explanation why classical Mahalanobis distances are inadequate for the use in simultaneous outlier identification rules (also see Atkinson 1994, p. 1334).

Similarly, estimators based on iterative deletion are not suitable. As described by, for example, Rousseeuw and Leroy (1987, p. 254), such an estimator is calculated as follows: The Mahalanobis distance of each observation is calculated, the observation $\mathbf{x}_i$ yielding the largest distance is excluded, the distances are recalculated for the reduced dataset, and so on, up to a certain number of deleted observations. From the remaining data, the mean is calculated as a location estimate. Donoho and Gasko (1992, p. 1814) and Rousseeuw and Leroy (1987, p. 254) found that a location estimate based on iterative deletion has a finite-sample breakdown point $\varepsilon^* \le 1/(p+1)$. Therefore, with the result of Theorem 2, the masking breakdown point of any outlier identifier based on such an estimator will also be less than or equal to $1/(p+1)$. Even the use of "leave-one-out" versions of the classical estimates as described, for example, by Penny (1996), will not improve the result. Here sample mean and covariance are calculated without using the ith data point $\mathbf{x}_i$, and the Mahalanobis distance of $\mathbf{x}_i$ is then calculated with respect to these estimates (jackknifed Mahalanobis distance). But according to Atkinson and Mulira (1993), these jackknifed distances are monotone transformations of the original distances calculated from the complete data. Penny (1996) mentioned that using both distances will cause the same observations to be detected

as outliers of the same order of magnitude. Therefore, using "leave-one-out" versions will lead to identifiers with the same low masking breakdown points as those yielded by the classical estimators themselves.

The maximum possible masking breakdown point can be derived using Theorem 1, taking into account the known breakdown bounds for affine equivariant estimators. The following results hold for the finite-sample breakdown points of such estimators $\mathbf{m}$ and $\mathbf{S}$ of location and covariance. For the location case, Lopuhaä (1992) and Lopuhaä and Rousseeuw (1991) found that $[(N-p+1)/2]/N$ is surely a lower bound for the maximum possible finite-sample breakdown point. In addition, we have the findings of Donoho and Gasko (1992) that no affine equivariant estimator can exceed the breakdown bound $(n-p+1)/(2n-p+1)$, where n denotes the number of "good" observations. In our notation, this upper bound reads $((N-p+1)/2)/N$. Together, this leads to

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

where the bounds are equal for $N-p+1$ even and differ only slightly for $N-p+1$ odd.

For the covariance estimator, the sample $\underline{x}_N$ must be 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 $\mathbb{R}^p$ (cf. Rousseeuw 1985, p. 288). Then

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

(Davies 1987).

Using Theorem 1 leads immediately to the following corollary. The result is given for samples where the regular observations are in general position.

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

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

where $[x]$ denotes the integer part of $x \in \mathbb{R}$, $\varepsilon^M_{\max} = \max_{\mathrm{OR}} \varepsilon^M(\mathrm{OR})$, and the maximum is taken over all affine equivariant outlier identifiers as defined earlier.

From the results shown before, 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) introduced S estimators in the context of robust regression. Davies (1987) extended the definition to the multivariate location-scale setting. He derived the class of S estimates $(\mathbf{m}, \mathbf{S})$ for multivariate location and covariance, showing that the maximum attainable breakdown point for estimators of this class is $[(N-p+1)/2]/N$, in which case we will denote them as S_{MB} estimators.

Corollary 2. Under the conditions of Corollary 2, let $\mathrm{OR}_{S_{\mathrm{MB}}}$ be an outlier identifier based on S_{MB} estimators for location and covariance. Then

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

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 article, we adjust the critical value to the sample size.

In the same way, a wide variety of estimators can be used, leading to identification procedures with the same high masking breakdown point, for example minimum covariance determinant (MCD) estimators (Rousseeuw 1985), Stahel–Donoho estimators (Stahel 1981, Donoho 1982) and their modifications (e.g., Gather and Hilker 1997; Tyler 1994; Maronna and Yohai 1995), or constrained M estimators (Kent and Tyler 1996).

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

Example 3 (Continuing Example 2). The identifier $\mathrm{OR}_{\mathrm{BW}}$ based on Tukey's biweight (Beaton and Tukey 1974; also see Rocke 1996) is given by

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

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

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

under the restriction

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

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

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

$c_0 \in \mathbb{R}$, such that the finite-sample breakdown point of $\mathbf{S}_{\mathrm{BW}}$ is maximal. That means that 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 $D^2 \sim \chi_p^2$. 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 $\alpha = .1$. We get $c_{\mathrm{BW}}(13, 59, .0018) = 38.23$.

Table 2 shows the results of the identification procedure for the dataset from Example 2. We see that, except for observation 58, all outliers found by the classical identification procedure are also identified by $\mathrm{OR}_{\mathrm{BW}}$. Beyond that, three other observations are detected as outliers by the robust identifier: observations 40, 46, and 49. If we look at

Table 2. Observation Distances With Respect to $\mathbf{m}_{BW}$ and $\mathbf{S}_{BW}$

Observation	Distance	Observation	Distance
1	3.38	**31**	**118.33**
2	6.79	32	9.51
3	4.52	33	13.31
4	5.05	34	4.07
5	10.57	35	3.91
6	10.98	36	19.04
7	16.83	**37**	**38.28**
8	22.50	38	12.08
9	6.65	39	16.46
10	2.33	**40**	**148.54**
11	6.98	41	4.27
12	27.81	42	7.01
13	4.32	43	6.96
14	4.47	44	3.56
15	6.78	45	8.44
16	10.59	**46**	**412.00**
17	2.38	**47**	**1,521.27**
18	**94.79**	**48**	**225.31**
19	10.62	**49**	**91.28**
20	7.63	50	4.55
21	3.86	51	4.55
22	7.13	52	3.54
23	3.83	53	5.71
24	14.05	54	24.15
25	3.03	55	10.60
26	8.89	56	13.36
27	10.17	57	13.29
28	**6,267.88**	58	26.47
29	12.51	59	3.20
30	9.57		

NOTE: Observations identified as outliers are represented by boldface.

the areas that are now classified to be extreme, we find that most of them are located along the west coast of the United States: Seattle, WA (No. 49); Portland, OR (No. 40); San Francisco, CA (No. 47); San Jose, CA (No. 48); Los Angeles/ Long Beach, CA (No. 28); and San Diego, CA (No. 46). Table 2 shows that Los Angeles and San Francisco have the largest distances with respect to the robust estimators $\mathbf{m}_{BW}$ and $\mathbf{S}_{BW}$, where the distance for Los Angeles is striking. For the classical identifier OR_{MD}, Los Angeles also yields the largest distance (see Table 1), but it is not that remarkably large. Here we see the effect of one very extreme observation on the classical procedure: The estimators $\bar{\mathbf{x}}_N$ and $\mathbf{S}_N$ are both strongly influenced by this observation, leading to a smaller Mahalanobis distance and yielding the effect that—in the presence of further extreme outliers (Nos. 47 and 48)—some outlying observations (Nos. 40, 46 and 49) remain undetected. We get the same conclusion for the identification procedure based on the jackknifed Mahalanobis distances ("leave-one-out"), because this method leads to the same results as the usual Mahalanobis distances. (See the discussion of that point after Theorem 2.)

Example 4. As a second multivariate dataset, we investigate the bushfire data considered by Maronna and Yohai (1995). The data were collected by Campbell (1989) for the purpose of locating bushfire scars. They consist of $N = 38$ observations (each corresponding to one pixel) of satellite measurements on $p = 5$ frequency bands. Maronna and Yohai considered outlier identification rules of a similar type as presented here, using Mahalanobis-type distances with respect to various estimators but not giving a formal rule to determine a normalizing constant or rejection value c. They used four pairs of estimators $(\mathbf{m}, \mathbf{S})$: sample mean and covariance, a pair of M estimators (Cauchy maximum likelihood; Maronna 1976), the Stahel–Donoho estimators based on median and a modified median absolute deviation, and reweighted S estimators based on the biweight function. This last pair of estimators uses $\mathbf{m}_{BW}$ and $\mathbf{S}_{BW}$ as introduced in Example 3 as a starting point for a reweighting step. Maronna and Yohai found that using the Stahel–Donoho estimators gives the best results, followed by the reweighted S estimators. The outlier identification via M estimators suffers from the masking effect, as the amount of contamination in the bushfire data is larger than the breakdown point of the estimators. The classical identification rule does not find any observation to be outlying.

Maronna and Yohai concluded that observations 8, 9, and 32–38 are the most extreme outliers in this dataset, followed by observations 7, 10, and 11, which are still declared to be clearly outlying. Finally, observations 12 and 31 seem suspect.

The results for the reweighted S estimators are somewhat surprising with respect to the findings in our paper: Maronna and Yohai remark that "it appears that cases 7–11 have 'masked' the cluster 32–38" to the identification procedure based on these estimators (p. 337). We return to this later. First, we discuss the results of the outlier identifier OR_{BW} for the bushfire data. Table 3 shows the distances with respect to $\mathbf{m}_{BW}$ and $\mathbf{S}_{BW}$. The constant c of the identification procedure for this case is $c_{BW}(5, 38, .0028) = 27.86$. We see that observations 8–10 and 32–38 are clearly detected as outliers, whereas the distances of observations 7 and 11 do not exceed c. With our procedure, we find that the observations 32–38 are the most extreme ones. As discussed by Maronna and Yohai, they lie in a "corner" of the dataset.

Using $\mathbf{m}_{BW}$ and $\mathbf{S}_{BW}$ for calculating the reweighted S estimators, and applying the respective outlier identi-

Table 3. Observation Distances With Respect to $\mathbf{m}_{BW}$ and $\mathbf{S}_{BW}$

Observation	Distance	Observation	Distance
1	1.50	20	1.11
2	.83	21	.97
3	.91	22	2.30
4	.77	23	.84
5	1.36	24	.85
6	2.68	25	.64
7	6.82	26	.72
8	**98.61**	27	.70
9	**104.76**	28	1.34
10	**36.12**	29	2.26
11	26.51	30	4.15
12	2.67	31	26.71
13	2.54	**32**	**157.64**
14	1.50	**33**	**203.52**
15	1.77	**34**	**196.54**
16	1.46	**35**	**206.01**
17	.64	**36**	**194.03**
18	1.29	**37**	**200.80**
19	1.29	**38**	**203.54**

NOTE: Observations identified as outliers are represented by boldface.

Table 4. Observation Distances With Respect to $\mathbf{m}_{RWBW}$ and $\mathbf{S}_{RWBW}$

Observation	Distance	Observation	Distance
1	4.79	20	3.58
2	2.69	21	3.19
3	3.06	22	7.40
4	2.58	23	2.68
5	4.30	24	2.70
6	8.32	25	2.13
7	15.98	26	2.40
8	**310.86**	27	2.34
9	**330.88**	28	4.10
10	**109.61**	29	6.18
11	**78.62**	30	11.79
12	6.93	**31**	**80.41**
13	7.87	**32**	**489.83**
14	4.48	**33**	**632.76**
15	5.50	**34**	**610.22**
16	4.76	**35**	**643.00**
17	2.14	**36**	**602.34**
18	4.07	**37**	**625.85**
19	4.04	**38**	**634.98**

NOTE: Observations identified as outliers are represented by boldface.

fier $\mathrm{OR}_{\mathrm{RWBW}}$ (reweighted biweight), we get a result that differs from the findings of Maronna and Yohai, as can be seen in Table 4. Essentially the same observations as found by $\mathrm{OR}_{\mathrm{BW}}$ are identified as outliers, the only difference being observation 11, which is now also detected. The appropriate normalizing constant c for this identification procedure is also calculated by simulation, yielding $c_{\mathrm{RWBW}}(5, 38, .0028) = 29.11$. We thus see that this identifier does not suffer from masking for the bushfire data. The difference between our calculations and the results of Maronna and Yohai for identical estimators may be explained by the different algorithms used to obtain the estimators. S estimators are solutions to a minimization problem; their explicit calculation for a given dataset can yield only an approximative solution, commonly based on some sort of subsampling at one step of the algorithm. Because of this, it can happen that the true optimum will not be found but that the algorithm converges to some local optimum instead. We suppose that for the solution found by Maronna and Yohai, this may have happened, because from their construction, both estimators should lead to equally good procedures.

Finally, we regard a dataset that is rather often investigated in the context of outlier identification, the so-called "stackloss data." We therefore choose it as a reference dataset to compare the results of a wide variety of outlier identification rules with the results of our robust procedure, where we again concentrate on the identifier $\mathrm{OR}_{\mathrm{BW}}$ based on S estimators with Tukey's biweight function.

Example 5. The dataset well known as "stackloss data" (Brownlee 1965, p. 454) comes 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. As before, these observations can be regarded as an unstructured multivariate dataset. We have a sample of size $N = 21$ with $p = 4$ variables. The value of the normalizing constant c for the identification procedure is then given as $c_{\mathrm{BW}}(4, 21, .0050) = 31.57$. Table 5 shows the (squared) distance $(\mathbf{x}_i - \mathbf{m}_{\mathrm{BW}})^T \mathbf{S}_{\mathrm{BW}}^{-1} (\mathbf{x}_i - \mathbf{m}_{\mathrm{BW}})$ for each observation $\mathbf{x}_i$, $i = 1, \ldots, 21$, of the dataset.

The identifier $\mathrm{OR}_{\mathrm{BW}}$ declares four observations as α_N outliers: $\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_{21}$, with $\alpha_N = .0050$. Most authors who investigated these data agree on that observations 3, 4, and 21 have to be regarded as outliers (cf. also Rousseeuw and Leroy 1987, p. 76). Andrews (1974), Daniel and Wood (1971), and Li (1985) identified the same four observations like $\mathrm{OR}_{\mathrm{BW}}$, whereas Carroll and Ruppert (1985) declared observations 2, 3, 4, 21 as conspicuous. Dempster and Gasko-Green (1981), as well as Andrews and Pregibon (1978), even detected five outliers (1, 2, 3, 4, 21). Although observation 2 shows a relatively high distance, its identification is not justified by the robust procedure $\mathrm{OR}_{\mathrm{BW}}$.

With the results of $\mathrm{OR}_{\mathrm{BW}}$, 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. Table 5 does not support this interpretation. But if we use the nonrobust estimators $\bar{\mathbf{x}}_N$ and $\mathbf{S}_N$ instead of $\mathbf{m}_{\mathrm{BW}}$ and $\mathbf{S}_{\mathrm{BW}}$, we find a similar behavior of the resulting procedure as in the aforementioned investigations. Due to the masking effect, observation 21 has then the largest distance of all observations (see Table 6), even though it does just not exceed the respective critical value $c(4, 21, .0048) = 11.19$. At the same time, the distances of observations 1, 3, and 4 are not exceptionally large. Therefore, the strong outliers (1, 3, 4) not only mask themselves but also cause the impression that the less-differing obser-

Table 5. Observation Distances With Respect to $\mathbf{m}_{BW}$ and $\mathbf{S}_{BW}$

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	.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**

NOTE: Observations identified as outliers are represented by boldface.

Table 6. Observation distances with respect to $\bar{\mathbf{x}}_N$ and $\mathbf{S}_N$

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	.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	.92
		21	11.13

vation 21 is the most conspicuous one. The similarity of this situation to the behavior of the aforementioned procedures leads to the conclusion that those procedures are still influenced by the strong outliers (1, 3, 4) when they label observation 21 as the clearest outlier. In contrast to this, $\mathrm{OR}_{\mathrm{BW}}$ is less influenced by observations 1, 3, and 4 and thus does not identify observation 21 as the most conspicuous one.

Finally, we may summarize that robust outlier identifiers show good theoretical properties with respect to the resistance against the masking effect. They behave well in real data situations and may give some more information about the structure in the data.

APPENDIX: PROOFS

Proof of Theorem 1

We first show the left inequality. 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 $\mathbf{S}$ can break down. From this, it follows

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

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

It now follows that

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

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

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

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

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

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

which means that 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,

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

and

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

For the second inequality, assume that $\varepsilon^M(\mathrm{OR}) > K/N$, which means that $\varepsilon^M(\mathrm{OR}) \geq (K+1)/(N+1) = (K+1)/(n+K+1)$. 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 that are placed as δ outliers.

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

$$\mathbb{R}^p \backslash \mathrm{OR}(\underline{x}_N, \alpha_N) \subseteq \mathbb{R}^p \backslash \mathrm{out}(\beta^*, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

for arbitrary $\underline{x}_K^0$. With this relation, the center of the ellipsoid $\mathbb{R}^p \backslash \mathrm{OR}(\underline{x}_N, \alpha_N)$ 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 2

Without loss of generality, let $\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \mathcal{I}$. Consider a situation with $k = k_1$ outliers. Then the covariance estimator $\mathbf{S}$ does not break down. For the eigenvalues $\lambda_i(\mathbf{S})$, $i = 1, \ldots, p$, of $\mathbf{S}$, it follows that

$$\exists\, C \in \mathbb{N} : 0 < \lambda_p(\mathbf{S}) \leq \cdots \leq \lambda_1(\mathbf{S}) < C,$$

and $\mathbf{S}^{-1}$ exists. Therefore,

$$\mathbb{R}^p \backslash \mathrm{OR}(\underline{x}_N, \alpha_N) = \{\mathbf{x} \in \mathbb{R}^p : (\mathbf{x}-\mathbf{m})^T \mathbf{S}^{-1} (\mathbf{x}-\mathbf{m}) < c(p, N, \alpha_N)\} \neq \emptyset.$$

In particular, there exists some point $\mathbf{x} \in \mathbb{R}^p$ with $(\mathbf{x}-\mathbf{m})^T \mathbf{S}^{-1} (\mathbf{x}-\mathbf{m}) < c(p, N, \alpha_N)$; for example, $\mathbf{x} = \mathbf{m}$. Now $\varepsilon^*(\mathbf{m}, \underline{x}_N) = k_1/N = k/N$. This means that for every $D \in \mathbb{R}$ there exists a constellation $\underline{x}_k^0 = \underline{x}_k^0(D)$ of δ outliers such that for $\underline{x}_N = (\underline{x}_n^r, \underline{x}_k^0(D))$, we have

$$\mathbf{m}(\underline{x}_N)^T \mathbf{m}(\underline{x}_N) > D,$$

which just corresponds to a breakdown of the location estimator. We can find such a constellation especially for $D = \chi^2_{p;1-\beta}$ for arbitrary values of β.

Therefore, for each $\beta, 0 < \beta < 1$, we find a set $\underline{x}_k^0$ of δ outliers, such that for the location estimate $\mathbf{m}$ based on $\underline{x}_N = (\underline{x}_n^r, \underline{x}_k^0)$ the following hold:

$$\mathbf{m} \in \mathbb{R}^p \backslash \mathrm{OR}(\underline{x}_N, \alpha_N)$$

and

$$\mathbf{m}^T \mathbf{m} > \chi^2_{p;1-\beta}, \quad \text{i.e.,} \quad \mathbf{m} \in \mathrm{out}(\beta, \mathbf{0}, \mathcal{I}).$$

This describes a situation where OR suffers from the masking effect. Therefore,

$$\varepsilon^M(\mathrm{OR}) \leq \frac{k_1}{N}.$$

[Received December 1997. Revised February 1999.]

REFERENCES

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

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

Atkinson, A. C. (1994), "Fast Very Robust Methods for the Detection of Multiple Outliers," *Journal of the American Statistical Association*, 89, 1329–1339.

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

Bacon-Shone, J., and 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., and Lewis, T. (1994), *Outliers in Statistical Data* (3rd ed.), New York: Wiley.

Beaton, A. E., and 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 Dataset 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.

Campbell, N. A. (1989), "Bushfire Mapping Using NOAA A VHRR Data," technical report, CSIRO.

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

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

Daniel, C., and 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., and Gather, U. (1993), "The Identification of Multiple Outliers," *Journal of the American Statistical Association*, 88, 782–792.

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

Donoho, D. L. (1982), "Breakdown Properties of Multivariate Location Estimators," Ph.D. qualifying paper, Harvard University, Department of Statistics.

Donoho, D. L., and Gasko, M. (1992), "Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness," *The Annals of Statistics*, 20, 1803–1827.

Donoho, D. L., and 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., and 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., and Hilker, T. (1997), "A Note on Tyler's Modification of the MAD for the Stahel-Donoho Estimator," *The Annals of Statistics*, 25, 2024–2026.

Gnanadesikan, R., and 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 and 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.

Healy, M. J. R. (1968), "Multivariate Normal Plotting," *Applied Statistics*, 17, 157–161.

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

Jennings, L. W., and Young, D. M. (1988), "Extended Critical Values of the Multivariate Extreme Deviate Test for Detecting a Single Spurious Observation," *Communications in Statistics—Simulation and Computation*, 17, 1359–1373.

Kent, J. T., and Tyler, D. E. (1996), "Constrained M-Estimation for Multivariate Location and Scatter," *The Annals of Statistics*, 24, 1346–1370.

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.

——— (1992), "Highly Efficient Estimators of Multivariate Location With High Breakdown Point," *The Annals of Statistics*, 20, 398–413.

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

Maronna, R. A. (1976), "Robust M-Estimators of Multivariate Location and Scatter," *The Annals of Statistics*, 4, 51–67.

Maronna, R. A., and Yohai, V. J. (1995), "The Behavior of the Stahel–Donoho Robust Multivariate Estimator," *Journal of the American Statistical Association*, 90, 330–341.

Penny, K. I. (1996), "Appropriate Critical Values When Testing for a Single Multivariate Outlier by Using the Mahalanobis Distance," *Applied Statistics*, 45, 73–81.

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

Rousseeuw, P. J., and 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-Verlag, pp. 256–272.

Rousseeuw, P. J., and 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, Part A*, 16, 1749–1760.

Stahel, W. A. (1981), "Breakdown of Covariance Estimators," Research Report 31, ETH Zürich, Fachgruppe für Statistik.

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*, Ser. A, 25, 407–426.