Multivariate Estimation With High Breakdown Point

Article · January 1985

DOI: 10.1007/978-94-009-5438-0_20

Peter Rousseeuw
University of Leuven

Appeared in: (1985) Mathematical Statistics and Applications, edited by W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Reidel Publishing Company, Dordrecht, pp. 283-297.

MULTIVARIATE ESTIMATION WITH HIGH BREAKDOWN POINT

PETER ROUSSEEUW

1. INTRODUCTION

Suppose we have a sample $X = \{x_1, \ldots, x_n\}$ of n points in p dimensions, and we want to estimate its location by means of an estimator T which is translation equivariant:

$$T(x_1+b, \ldots, x_n+b) = T(x_1, \ldots, x_n) + b \text{ for all } b \tag{1.1}$$

and permutation invariant:

$$T(x_{\pi(1)}, \ldots, x_{\pi(n)}) = T(x_1, \ldots, x_n) \text{ for any permutation } \pi. \tag{1.2}$$

The best known estimator is the arithmetic mean $\bar{x} = \frac{1}{n}\Sigma_{i=1}^{n} x_i$, which is the least squares estimator because it minimizes $\Sigma_{i=1}^{n} ||x_i - T||^2$ where $||..||$ is the ordinary Euclidean norm. However, it is well-known that $\bar{x}$ is not robust, because even a single (very bad) outlier in the sample can move $\bar{x}$ arbitrarily far away. To formalize this, we shall make use of the breakdown point ε^* defined by Hampel (1971). In this paper, we use a version introduced by Donoho and Huber (1983) which is intended for finite samples, like the precursor idea of Hodges (1967). Denote by $\beta(m;T,X)$ the supremum of $||T(X')-T(X)||$ for all corrupted samples X' where m of the original data points are replaced by arbitrary values. Then the breakdown point of the estimator T at the sample X is defined as

$$\varepsilon_n^*(T,X) = \min\{\frac{m}{n};\ \beta(m;T,X) \text{ is infinite}\}. \qquad (1.3)$$

In words, it is the smallest fraction of contamination that can cause the estimator to take on values arbitrarily far away from T(X). For the arithmetic mean we find $\varepsilon_n^*(T,X) = \frac{1}{n}$ because one bad observation can already cause breakdown. This ε_n^* depends only slightly on n; in order to have only a single value, one often considers the limit for $n \to \infty$ (with p fixed), so we can say that $\bar{x}$ has a breakdown point $\varepsilon^* = 0$ %.

The maximal breakdown point is 50 % for translation equivariant and permutation invariant estimators (consider a configuration of outliers which is just a translation image of the "good" data points). In one dimension, the median reaches $\varepsilon^* = 50$ %. There are two straightforward generalizations to the case $p > 1$:

(i) the coordinatewise median $(\operatorname{med}_i(x_i^1), \ldots, \operatorname{med}_i(x_i^p))$ where x_i^j is the j-th coordinate of x_i

(ii) the L_1 estimator, which minimizes $\Sigma_{i=1}^n ||x_i - T||$.

Both estimators attain $\varepsilon^* = 50$ %.

Remark 1 For $p \geq 3$, the coordinatewise median does not have to lie in the convex hull of the sample. (As an example, consider the p unit vectors X:= {(1,0,..,0), (0,1,..,0),.., (0,0..,1)}, the convex hull of which is a simplex which does not contain the coordinatewise median (0,0,..,0) .) However, it does lie in the convex hull for $p \leq 2$, as can be shown by a geometrical argument.

2. AFFINE EQUIVARIANCE

In many situations, one wants the estimation to commute with coordinate transformation. We say that T is affine equivariant if and only if

$$T(Ax_1+b,..,Ax_n+b) = AT\,(x_1,..,x_n)+b \qquad (2.1)$$

for any vector b and any nonsingular matrix A. The arithmetic mean $\bar{x}$ is affine equivariant, unlike the coordinatewise median and the L_1 estimator. However, the L_1 still satisfies the weaker condition of orthogonal equivariance (this is (2.1) restricted to orthogonal matrices A) because it only depends on distances.

On the other hand, multivariate M-estimators are affine equivariant, but their breakdown point is at most $1/(p+1)$ (see Maronna 1976). Donoho (1982) lists some other well-known affine equivariant estimators, and shows that they all satisfy $\varepsilon^* \leq 1/(p+1)$. This raises the question whether it is at all possible to combine a high breakdown point with affine equivariance. In the next section we shall see that this is indeed the case, by discussing the Stahel-Donoho estimator and a new proposal. However, it seems that the affine equivariance condition is very demanding, because the latter estimators are quite complicated. The remainder of this section will be devoted to illustrating the severeness of this condition.

Remark 2 If T is affine equivariant, permutation invariant and continuous in the observations, then T is the arithmetic mean. This follows from taking A in (2.1) equal to diagonal matrices with (p-1) entries λ and one entry 1 . If one lets λ

tend to zero, it follows by continuity that T has to commute with all projections on coordinates. But then it follows from Donoho (1982, proposition 4.6) that T is the arithmetic mean. (This result can also be proven from section 3 of Obenchain 1971).

Remark 3 Some people have expressed the opinion that it should be very easy to obtain high breakdown estimators which are affine equivariant, by rescaling the observations. They proposed the following procedure: calculate the covariance matrix V, and take a root S (that is, $SS^t = V$). Transform the data as $\tilde{x}_i := S^{-1}x_i$, apply to these $\tilde{x}_i$ an easily computable estimator $\tilde{T}$ with $\varepsilon^* = 50\%$ which is not affine equivariant, and then transform back by putting $T := S\tilde{T}$. First, we observe that this construction gives a unique result if and only if $\tilde{T}$ is equivariant for orthogonal transformations, because for any orthogonal matrix 0 the product SO is also a root of V. Also $\tilde{T}$ has to be translation equivariant to insure affine equivariance of T. Therefore, it seems that taking the L_1 estimator for $\tilde{T}$ will do the job. However, we shall see that its good breakdown behaviour does not carry over.

To show this, we consider a two-dimensional example where a fraction $(1-\varepsilon)$ of the data is spherically bivariate normal (we assume that both coordinates follow a univariate normal distribution N(0,1) and are independent of each other), and there is a fraction ε of outliers which are concentrated at the point with coordinates (0,u). Here, $u > 0$ and $0 < \varepsilon < \frac{1}{2}$. Calculation shows that the covariance matrix V is diagonal, with the diagonal elements $(1-\varepsilon)$ and $(1-\varepsilon)(1+\varepsilon u^2)$. Therefore, we can easily construct a root S of V by taking the square root of these diagonal elements, yielding $S_{11} = (1-\varepsilon)^{1/2}$, and $S_{22} = ((1-\varepsilon)\ (1+\varepsilon u^2))^{1/2}$, and of course $S_{12} = S_{21} = 0$. For u tending to ∞, the transformation $S^{-1} : (x,y) \to (\tilde{x},\tilde{y})$ yields the following situation: the fraction $(1-\varepsilon)$ gets concentrated on the x-axis with univariate normal distribution $N(0,(1-\varepsilon)^{1/2})$, and the fraction ε lands at the point $(0,(\varepsilon(1-\varepsilon))^{-1/2})$. For any finite value of u, the L_1 estimate of the transformed data lies on the vertical axis by symmetry, so we may denote it by $\tilde{T} = (0,\Delta(u))$. Let us now look at the limit $\Delta := \lim_{u\to\infty} \Delta(u)$. By definition, Δ minimizes the expected value of $\|(\tilde{x},\tilde{y}) - (0,\Delta)\|$ which is denoted by $D := \varepsilon((\varepsilon(1-\varepsilon))^{-1/2}-\Delta) + (1-\varepsilon)\int[\Delta^2 + \tilde{x}^2]^{1/2}\ dF(\tilde{x})$, where F is the normal distribution $N(0,(1-\varepsilon)^{-1/2})$. By the substitution $v = (1-\varepsilon)^{1/2}\tilde{x}$, the average distance D equals $\varepsilon((\varepsilon(1-\varepsilon))^{-1/2}-\Delta) + (1-\varepsilon)^{1/2}\int[(1-\varepsilon)\Delta^2 + v^2]^{1/2}\, d\Phi(v)$ where Φ is the standard normal distribution. Because Δ minimizes D, it follows that $0 = \partial D/\partial\Delta = -\varepsilon + (1-\varepsilon)\int\{(1-\varepsilon)\ \Delta^2 / [(1-\varepsilon)\Delta^2 + v^2]\}^{1/2}$ $d\Phi(v)$, hence $\int\{(1-\varepsilon)\Delta^2/[(1-\varepsilon)\Delta^2+v^2]\}^{1/2} d\Phi(v)$ equals the positive constant $\varepsilon/(1-\varepsilon)$. This cannot happen for $\Delta = 0$, hence $\Delta > 0$. As the final estimate T equals $S\tilde{T}$, we conclude that $\|T\| = \|S\tilde{T}\| = \|(0,S_{22}(u)\Delta(u))\| = \{(1-\varepsilon)\ (1+\varepsilon u^2)\}^{1/2}\Delta(u)$ tends to infinity for $u \to \infty$. (Note that this effect is due to the explosion of the first factor, caused by the breakdown of S !) This means that any fraction $\varepsilon > 0$ can make T break down, so $\varepsilon^*(T) = 0$ %. Therefore, rescaling does not solve our problem unless we could start with a high-breakdown covariance estimator, but that is precisely what we are looking for in the first place.

3. HIGH-BREAKDOWN ESTIMATORS

The first affine equivariant multivariate location estimator with a breakdown point of 50 % was obtained independently by W. Stahel (1981) and D. Donoho (1982). This estimator, called "outlyingness-weighted mean", is defined as follows. For each observation x_i in X, one looks for a one-dimensional projection in which x_i is most outlying:

$$r_i := \sup_{\|u\|=1} \frac{u^t x_i - \operatorname{med}_j(u^t x_j)}{\operatorname{med}_k |u^t x_k - \operatorname{med}_j(u^t x_j)|} \tag{3.1}$$

where $\operatorname{med}_j(u^t x_j)$ is the median of all projections of data points x_j on the direction of the vector u. Then one estimates location by the weighted mean

$$T(x_1,\ldots,x_n) := \frac{\Sigma_{i=1}^n w(r_i) x_i}{\Sigma_{i=1}^n w(r_i)} \tag{3.2}$$

where w(r) is a strictly positive and decreasing function of $r \geq 0$, such that $r \cdot w(r)$ is bounded. (The bound on $r \cdot w(r)$ is motivated by the one-dimensional case, where (3.2) is a one-step W-estimator starting from the median (Hoaglin et al 1983, page 291), the influence function of which is proportional to $r \cdot w(r)$.) Formula (3.1) is related to projection pursuit (Friedman and Tukey 1974), because one really has to search over all possible directions to find the "best" projection. It is shown that (3.2) is affine equivariant, and has a breakdown point converging to 50 % (Donoho 1982, proposition 3.1) at samples X in general position, which means that no more than p points of X lie in any (p-1) dimensional affine subspace. (3.3)

For two-dimensional data (visualized as points in a plane) this means that there are no more than two points of X on any line, so any three points of X determine a triangle with nonzero area.

Let us now introduce a second affine equivariant estimator with high breakdown, by putting

$$T(X) := \text{center of the minimal volume ellipsoid covering (at least) } h \text{ points of } X \tag{3.4}$$

where $h = \left[\frac{n}{2}\right]+1$ (here $\left[\frac{n}{2}\right]$ is the largest integer $\leq \frac{n}{2}$). We call this the Minimal Volume Ellipsoid estimator (MVE). The affine equivariance of the MVE follows from the fact that for each ellipsoid E the image f(E) through the nonsingular affine transformation $f(x) = Ax+b$ is again an ellipsoid, and

$$\text{Volume } (f(E)) = |\det (A)| \text{ Volume } (E).$$

Because det(A) is a constant, the relative sizes of ellipsoids do not change under affine transformations.

Proposition 3.1 At any p-dimensional sample X in general position, the breakdown point of the MVE estimator equals

$$\varepsilon_n^*(T,X) = \left(\left[\frac{n}{2}\right]-p+1\right) /n$$

which converges to 50 % as $n \to \infty$.

Proof Without loss of generality, let $T(X) = 0$. We put $M :=$ volume of smallest ellipsoid with center zero containing all points of X. Because X is in general position, each of its $\binom{n}{p+1}$ subsets S_j of $p+1$ points determines a simplex with nonzero volume. Therefore, for each S_j there exists a bound d_j such that any ellipsoid with center $\|c\| > d_j$ and containing S_j has a volume strictly larger than M. Put $d := \max_j d_j < \infty$.

Let us first show that $\varepsilon_n^*(T,X) \geq ([\frac{n}{2}]-p+1)/n$. Take any sample X' obtained by replacing at most $[\frac{n}{2}]-p$ points of X. Suppose $\|T(X')\| > d$, and let E be the corresponding smallest ellipsoid containing (at least) $[\frac{n}{2}]+1$ points of X'. But then E contains at least $([\frac{n}{2}]+1) - ([\frac{n}{2}]-p) = p+1$ points of X, so Volume $(E) > M$. This is a contradiction, because the smallest ellipsoid with center zero around the $n-([\frac{n}{2}]-p) \geq [\frac{n}{2}]+1$ "good" points of X' has a volume of at most M. Therefore, $\|T(X')\| \leq d$. (Even if $T(X')$ is not unique, then $\|T(X')\| \leq d$ still holds for all solutions.)

On the other hand, $\varepsilon_n^*(T,X) \leq ([\frac{n}{2}]-p+1)/n$. Indeed, take any p points of X, and consider the $(p-1)$-dimensional affine subspace H they determine. Now replace $[\frac{n}{2}]-p+1$ other points of X by points on H. Then H contains $[\frac{n}{2}]+1$ points of the new sample X', so the minimal volume ellipsoid covering these $[\frac{n}{2}]+1$ points degenerates to zero volume. Because X is in general position, no other ellipsoid covering $[\frac{n}{2}]+1$ points of X' can have zero volume, so $T(X')$ lies on H. Finally, we note that $T(X')$ is not bounded because the $[\frac{n}{2}]-p+1$ contaminated data points on H may have arbitrarily large norm. This ends the proof. □

A variant of the MVE is the estimator

$T(X)$:= mean of the h points of X for which the determinant of the covariance matrix is minimal. (3.5)

We call this the Minimal Covariance Determinant estimator (MCD), which corresponds to finding the h points for which the confidence ellipsoid (for any given level) has minimal volume, and then taking its center. This estimator is also affine equivariant because the determinant of the covariance matrix of the transformed data equals

$$\det(AVA^t) = \det^2(A) \det(V).$$

The MCD also has the same breakdown point as the MVE, using the same reasoning as in Proposition 3.1.

Both the MVE and the MCD are very drastic, because they are intended to safeguard against up to 50 % of outliers. If one is certain that the fraction of outliers is at most α (where $0 < \alpha \leq \frac{1}{2}$), one can work with the estimators MVE(α) and MCD(α) obtained by replacing h by $k(\alpha) := [n(1-\alpha)] + 1$ in (3.4) and (3.5). The breakdown point of these estimators is equal to α (for $n \to \infty$). For $\alpha \to 0$, the MVE yields the center of the smallest ellipsoid covering all the data, Whereas the MCD tends to the arithmetic mean.

Also note that both the MVE(α) and the MCD(α) yield very robust covariance estimators at the same time, if one makes use of the selected fraction $k(\alpha)$ of observations and multiplies the resulting matrix with a constant to obtain consistency in the case of multivariate normality.

4. A PARTICULAR CASE

In the general multivariate setting, the asymptotic properties of the MVE and MCD estimators are hard to determine. However, a beginning can be made by considering the special case where p=1. In this situation, it is easy to give an explicit algorithm for calculating the MVE and the MCD estimates. First one orders the data, so we may assume that

$$x_1 \leq x_2 \leq \ldots \leq x_n.$$

Only n-h+1 "halves" of this sample need to be considered, namely

$$H_1 = \{x_1, \ldots, x_h\},\ H_2 = \{x_2, \ldots, x_{h+1}\} \ \ldots$$

$$\ldots,\ H_{n-h+1} = \{x_{n-h+1}, \ldots, x_n\}.$$

For each of these H_i, one defines

$$\text{midpoint}(i) = \frac{1}{2}(x_i + x_{i+h-1}) \quad \text{length}(i) = x_{i+h-1} - x_i$$

$$\text{mean}(i) = \frac{1}{h} \Sigma_{j=i}^{i+h-1} x_j \qquad \text{var}(i) = \frac{1}{h-1} \Sigma_{j=i}^{i+h-1} (x_j - \text{mean}(i))^2.$$

For the MVE one selects the H_i for which length(i) is minimal, and one puts T(X):=midpoint(i). (In case more than one length(i) is minimal, T(X) can be taken as the average of the corresponding numbers midpoint(i).) For the MCD, one selects the H_i for which var(i) is minimal and puts T(X):=mean(i); if there are several solutions, one again can take their average. Also the MVE(α) and MCD(α) estimates can be calculated this way, by replacing h by $k(\alpha) = [n(1-\alpha)]+1$.

Remark 4 When p=1, the MVE(α) estimator can be rewritten as

$$\underset{t}{\text{minimize}}\ (r^2)_{k(\alpha):n} \tag{4.1}$$

where the $(r^2)_{i:n}$ are the squared residuals $(x_j-t)^2$, ordered from smallest to largest. The MCD(α) estimator is the solution of

$$\underset{t}{\text{minimize}}\ \Sigma_{i=1}^{k(\alpha)}\ (r^2)_{i:n}.$$

Both criteria (4.1) and (4.2) can also be applied to the estimation of regression coefficients, and yield a breakdown point equal to α in that case too (Rousseeuw 1982, 1983).

Let us now consider a symmetric and strongly unimodal distribution F with density f and finite Fisher information.

Proposition 4.1 If the observations are iid according to $F(x-\theta)$, then the distribution of the MVE(α) estimator T_n converges as:

$$\mathcal{L}(n^{1/3}(T_n-\theta)) \xrightarrow{\text{weak}} \mathcal{L}(c\tau/f(F^{-1}(1-\tfrac{\alpha}{2})).$$

Here $c=(\frac{1}{2}\Lambda^2(F^{-1}(1-\frac{\alpha}{2})))^{-1/3}$ where $\Lambda=-f'/f$ corresponds to the maximum likelihood scores, and τ is the random time s for which $s^2+Z(s)$ attains its minimum, where Z(s) is a standard Brownian motion.

Proof The proof is a simple adaptation of the result for a very similar estimator, namely the "shorth" (=mean of the shortest half of the sample). Parts 1, 2 and 3 of the reasoning in (Ahdrews et al. 1972, page 51) still apply and yield the constant c. We slightly adapt the remaining part of the calculation, using the same notations. If $\theta=0$ and $\hat{t}$ is the minimizing value of t, then main asymptotic variability of T_n is given by

$\frac{1}{2}\left[F_n^{-1}\left(\left(\frac{1}{2}+\hat{t}\right)+\left(\frac{1-\alpha}{2}\right)\right)+F_n^{-1}\left(\left(\frac{1}{2}+\hat{t}\right)-\left(\frac{1-\alpha}{2}\right)\right)\right] \cong \hat{t}(F^{-1})'\left(\frac{1}{2}+\left(\frac{1-\alpha}{2}\right)\right) =$
$\hat{t}/f(F^{-1}(1-\frac{\alpha}{2}))$, where $n^{1/3}\hat{t}$ behaves asymptotically like $c\tau$. □

This slow rate of convergence is a disadvantage of the MVE(α) method. However, note that number of arithmetic operations needed to compute MVE(α) is of the order $0(n\log(n))$, which is much faster than the $0(n^2)$ needed for the MCD(α) estimator.

Proposition 4.2 If the observations are iid according to $F(x-\theta)$, then the MCD(α) estimator T_n is asymptotically normal:

$$\mathcal{L}\left(n^{1/2}(T_n-\theta)\right) \xrightarrow{\text{weak}} N(0,V(\alpha))$$

where the asymptotic variance aquals

$$V(\alpha) = \frac{2\int_0^q x^2\, dF(x)}{(1-\alpha-2qf(q))^2}$$

(here, q is the quantile $F^{-1}(1-\frac{\alpha}{2})$).

Proof For each t, denote by $R_i(t)$ the rank of $|x_i-t|$. Moreover, the function a on $[0,1]$ is defined by $a(u)=(1-\alpha)^{-1}$ if $u \leq 1-\alpha$, and $a(u)=0$ otherwise. Then the MCD(α) estimator corresponds to $\underset{t}{\text{minimize}}\ \Sigma_{i=1}^{n}\ a(R_i(T)/n)(x_i-t)^2$. As there exists only a finite number of points t where some $R_i(t)$ changes, we may assume w.l.o.g. that all $R_i(t)$ are locally constant around the solution $t = T_n$. On this neighbourhood $(T_n-\delta, T_n+\delta)$ we put $a_i:=(R_i(T_n)/n)$ and

$S_n := \frac{1}{n} \Sigma_{i=1}^n a(R_i(T_n)/n)x_i$. Now T_n minimizes

$$\Sigma_{i=1}^n a_i((x_i-S_n)+(S_n-t))^2 = \Sigma_{i=1}^n a_i(x_i-S_n)^2 + \Sigma_{i=1}^n a_i(S_n-t)^2$$

for all t in

$(T_n-\delta, T_n+\delta)$. Differentiating this expression with respect to t at T_n, we find that $T_n=S_n$. Therefore, T_n is a solution of the equation

$$\frac{1}{n} \Sigma_{i=1}^n a(R_i(t)/n)x_i = t. \qquad (4.3)$$

Estimators satisfying (4.3) with this function a go back to Gnanadesikan and Kettenring (1972), and their asymptotic normality was shown by Yohai and Maronna (1976), giving the desired result. $\square$

Remark 5 This proposition implies that MCD(α) has the same asymtotic behaviour at F as the Huber-type "skipped" mean (Hampel 1974, page 392), which is the M-estimator of location (Huber 1981) corresponding to the ψ function

$$\begin{aligned} \psi(x) &= x \quad |x| \leq F^{-1}(1-\tfrac{\alpha}{2}) \\ &= 0 \quad \text{elsewhere} \end{aligned} \qquad (4.4)$$

Remark 6 The results of this section imply that our 50 % breakdown estimators MVE and MCD possess low asymptotic efficiencies. However, this can be repaired by applying a one-step M-estimator afterwards (Huber 1981, page 146), preferably using a smooth redescending ψ-function such as the one proposed by Hampel, Rousseeuw and Ronchetti (1981). In this way, the asymptotic efficiency can be improved without losing $\varepsilon^* = 50$ %.

Acknowledgments This research was supported by the Belgian National Science Foundation, and most of it took place while the author was visiting the Mathematical Sciences Research Institute at Berkeley. Thanks are due to D. Donoho for valuable suggestions and stimulating discussions. Helpful comments were made by F. Hampel, A. Leroy, F. Plastria, W. Stahel and V. Yohai.

REFERENCES

ANDREWS, D., BICKEL, P., HAMPEL, F., HUBER, P., ROGERS, W., and TUKEY, J. (1972).: Robust Estimates of Location: Survey and Advances. Princeton University Press.

DONOHO, D. (1982)."Breakdown Properties of Multivariate Location Estimators", Ph.D. qualifying paper, Harvard University.

DONOHO, D. and HUBER, P. (1983). "The Notion of Breakdown Point", in: A Festschrift for Erich L. Lehmann, Wadsworth.

FRIEDMAN, J. and TUKEY, J. (1974). "A Projection Pursuit Algorithm for Exploratory Data Analysis", IEEE Trans. Comp. C-23, 881-889.

GNANADESIKAN, R. and KETTENRING, J. (1972). "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data" Biometrics 28, 81-124.

HAMPEL, F. (1971). "A General Qualitative Definition of Robustness", Ann. Math. Statist. 42, 1887-1896.

HAMPEL, F. (1974). "The Influence Curve and its Role in Robust Estimation", JASA 69, 383-393.

HAMPEL, F. ROUSSEEUW, P., and RONCHETTI, E. (1981). "The Change-of-Variance Curve and Optimal Redescending M-estimators," JASA 76, 643-648.

HOAGLIN, D., MOSTELLER, F., TUKEY, J. (1983). Understanding Robust and Exploratory Data Analysis. Wiley, New York.

HODGES, J. (1967). "Efficiency in Normal Samples and Tolerance of Extreme Values for some Estimates of Location", Proc. 5th Berkeley Symp. on Math. Stat. and Prob., Vol. 1, 163-168.

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

MARONNA, R. (1967). "Robust M-estimators of Multivariate Location and Scatter", Ann. Statist. 4. 51-67.

OBENCHAIN, R. (1971). "Multivariate Procedures Invariant under Linear Transformations", Ann. Math. Statist. 42. 1569-1578.

ROUSSEEUW, P. (1982). "Least Median of Squares Regression", Technical Report, University of Brussels (to appear in JASA).

ROUSSEEUW, P. (1983). "Regression Techniques with High Breakdown Point", IMS Bulletin 12, page 155.

STAHEL, W. (1981). "Breakdown of Covariance Estimators" Research Report No. 31, Fachgruppe für Statistik, ETH Zürich.

YOHAI, V. and Maronna, R. (1976). "Location Estimators based on Linear Combinations of Modified Order Statistics", Commun. Statist. A5, 481-486.