CHAPTER 1

Robust Location and Scatter Estimators in Multivariate Analysis

Yijun Zuo

Department of Statistics and Probability, Michigan State University
East Lansing, MI 48824, USA
E-mail: zuo@msu.edu

The sample mean vector and the sample covariance matrix are the corner stone of the classical multivariate analysis. They are optimal when the underlying data are normal. They, however, are notorious for being extremely sensitive to outliers and heavy tailed noise data. This article surveys robust alternatives of these classical location and scatter estimators and discusses their applications to the multivariate data analysis.

1. Introduction

The sample mean and the sample covariance matrix are the building block of the classical multivariate analysis. They are essential to a number of multivariate data analysis techniques including multivariate analysis of variance, principal component analysis, factor analysis, canonical correlation analysis, discriminant analysis and classification, and clustering. They are optimal (most efficient) estimators of location and scatter parameters at any multivariate normal models. It is well-known, however, that these classical location and scatter estimators are extremely sensitive to unusual observations and susceptible to small perturbations in data. Classical illustrative examples showing their sensitivity are given in Devlin et al (1981), Huber (1981), Rousseeuw and Leroy (1987), and Maronna and Yohai (1998).

Bickel (1964) seems to be the first who considered the robust alternatives of the sample mean vector – the coordinate-wise median and the coordinate-wise Hodges-Lehmann estimator. Extending the univariate trimming and Winsorizing idea of Tukey (1949) and Tukey (1960) to higher dimensions, Bickel (1965) proposed the metrically trimmed and Winsorized means in the multivariate setting. All these estimators indeed are much more robust (than the sample mean) against outliers and contaminated data (and some are very efficient as well). They, however, lack the desired *affine equivariance* (see Section 2.2) property of the sample mean vector.

Huber (1972) discussed a "peeling" procedure for location parameters which was first proposed by Tukey. A similar procedure based on iterative trimming was presented by Gnanadesikan and Kettenring (1972). The resulting location estimators become affine equivariant but little seems to be known about their properties. Hampel (1973) was the first to suggest an affine equivariant iterative procedure for a scatter matrix, which turns out to be a special M-estimator (see Section 3.1) of the scatter matrix.

Inspired by Huber (1964)'s seminal paper, Maronna (1976) first introduced and treated systematically general M-estimators of multivariate location and scatter parameters. Huber (1977) considered the robustness of the covariance matrix estimator with respective to two measures: *influence function* and *breakdown point* (defined in Section 2).

Multivariate M-estimators are not greatly influenced by small perturbations in a data set and have reasonably good efficiencies over a broad range of population models. Ironically, they, introduced as robust alternatives to the sample mean vector and the sample covariance matrix, were frequently mentioned in the robust statistics literature in the last two decades, not because of their robustness but because of their not being robust enough globally (in terms of their breakdown point). Indeed, M-estimators have a relatively very low breakdown point in high dimensions and are not very popular choices of robust estimators of location and scatter parameters in the multivariate setting. Developing *affine equivariant* robust alternatives to the sample mean and the sample covariance matrix that also have *high breakdown points* consequently was one of the fundamental goals of research in robust statistics in the last two decades.

This paper surveys some influential robust location and scatter estimators developed in the last two decades. The list here is by no means exhaustive. Section 2 presents some popular robustness measures. Robust location and scatter estimators are reviewed in Section 3. Applications of robust estimators are discussed in Section 4. Concluding remarks and future research topics are presented in Section 5 at the end of the paper.

2. Robustness Criteria

Often a statistic T_n can be regarded as a functional $T(\cdot)$ evaluated at an empirical distribution F_n, where F_n is the empirical version of a distribution

F based on a random sample $X_1, \cdots, X_n$ from F, which assigns mass $1/n$ to each sample point X_i, $i = 1, \cdots, n$. In the following we describe three most popular robustness measures of functional $T(F)$ or statistic $T(F_n)$.

2.1. *Influence function*

One way to measure the robustness of the functional $T(F)$ at a given distribution F is to measure the effect on T when the true distribution slightly deviates from the assumed one F. In his Ph.D. thesis, Hampel (1968) explored this robustness and introduced the influence function concept. For a fix point $x \in \mathbb{R}^d$, let δ_x be the point-mass probability measure that assigns mass 1 to the point x. Hampel (1968, 1971) defined the *influence function* of the functional $T(\cdot)$ at a fixed point x and the given distribution F as

$$IF(x; T, F) = \lim_{0<\epsilon\to 0} \frac{T((1-\epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon}, \tag{1}$$

if the limit exists. That is, the influence function measures the relative effect (influence) on the functional T of an infinitesimal point mass contamination of the distribution F. Clearly, the relative effect (influence) on T is desired to be small or at least bounded. A functional $T(\cdot)$ with a bounded influence function is regarded as robust and desirable.

A straightforward calculation indicates that for the classical mean and covariance functionals $\mu(\cdot)$ and $\Sigma(\cdot)$ at a fixed point x and a given F in $\mathbb{R}^d$,

$$IF(x; \mu, F) = x - \mu(F), \qquad IF(x; \Sigma, F) = (x-\mu)(x-\mu)' - \Sigma(F).$$

Clearly, both influence functions are unbounded with respect to standard vector and matrix norms, respectively. That is, an infinitesimal point mass contamination can have an arbitrarily large influence (effect) on the classical mean and covariance functionals. Hence these functionals are not robust.

The model, $(1-\epsilon)F + \epsilon\delta_x$, a distribution with a slight departure from the F, is also called the *ϵ-contamination model*. Since only a point-mass contamination is considered in the definition, the influence function measures the *local* robustness of the functional $T(\cdot)$. General discussions and treatments of influence functions of statistical functionals could be found in Serfling (1980), Huber (1981), and Hampel et al. (1986).

In addition to being a measure of local robustness of a functional $T(F)$, the influence function can also be very useful for the calculation of the asymptotic variance of $T(F_n)$. Indeed, if $T(F_n)$ is asymptotically normal, then the asymptotic variance of $T(F_n)$ is just $E(IF(X; T, F))^2$ in general.

Furthermore, under some regularity conditions, the following asymptotic representation is obtained in terms of the influence function:

$$T(F_n) - T(F) = \int IF(x; T, F) d(F_n - F)(x) + o_p(n^{-1/2}),$$

which leads to the asymptotic normality of the statistic $T(F_n)$.

2.2. *Breakdown point*

The influence function captures the local robustness of a functional $T(\cdot)$. The breakdown point, on the other hand, depicts the *global* robustness of $T(F)$ or $T(F_n)$. Hampel (1968) and Hampel (1971) apparently are the first ones to consider the breakdown point of $T(F)$ in an asymptotic sense.

Donoho and Huber (1983) considered a finite sample version of the notion, which since then has become the most popular quantitative measure of global robustness of an estimator $T_n = T(F_n)$, largely due to its intuitive appeal, non-probabilistic nature of the definition, and easy calculation in many cases. Roughly speaking, the finite sample breakdown point of an estimator T_n is the minimum fraction of "bad" (or contaminated) data points in a data set $X^n = \{X_1, \cdots, X_n\}$ that can render the estimator useless. More precisely, Donoho and Huber (1983) defined the finite sample *breakdown point* of a *location* estimator $T(X^n) := T(F_n)$ as

$$BP(T; X^n) = \min\{\frac{m}{n} : \sup_{X_m^n} |T(X_m^n) - T(X^n)| = \infty\}, \tag{2}$$

where X_m^n is a contaminated data set resulting from replacing (contaminating) m points of X^n with arbitrary m points in $\mathbb{R}^d$. The above notion sometimes is called *replacement* breakdown point. Donoho (1982) and Donoho and Huber (1983) also considered *addition* breakdown point. The two versions, however, are actually interconnected quantitatively; see Zuo (2001). Thus we focus on the replacement version throughout in this paper.

For a *scatter* (or covariance) estimator S of the matrix Σ in the probability density function $f((x-\mu)'\Sigma^{-1}(x-\mu))$, to define its breakdown point one can still use (2) but with T on the left side replaced by S and $T(\cdot)$ on the right side by the vector of the logarithms of the eigenvalues of $S(\cdot)$. Note that for a location estimator, it becomes useless if it approaches ∞. On the other hand, for a scatter estimator, it becomes useless if one of its eigenvalues approaches 0 or ∞ (this is why we use the logarithm).

Clearly, the higher the breakdown point of an estimator, the more robust the estimator against outliers (or contaminated data points). It is not difficult to see that one bad (or contaminating one) point of a data set of size n is enough to ruin the sample mean or the sample covariance matrix. Thus, their breakdown point is $1/n$, the lowest possible value. That is, the sample mean vector and the sample covariance matrix are not robust globally (and locally as well due to the unbounded influence functions).

On the other hand, to have the sample median (Med) breakdown (unbounded), one has to move 50% of data points to the infinity. Precisely, the univariate median has a breakdown point $\lfloor (n+1)/2 \rfloor / n$ for any data set of size n, where $\lfloor x \rfloor$ is the largest integer no larger that x. Likewise, it can be seen that the median of the absolute deviations (from the median) (MAD) has a breakdown point $\lfloor n/2 \rfloor / n$ for a data set with no overlapping data points. These breakdown point results turn out to be the best for any *reasonable* location and covariance (or scale) estimators, respectively. Note that the breakdown point of a constant estimator is 1 but the estimator is not reasonable since it lacks some equivariance property.

Location and scatter estimators T and S are called *affine equivariant* if

$$T(AX^n + b) = A \cdot T(X^n) + b, \quad S(AX^n + b) = A \cdot S(X^n) \cdot A', \tag{3}$$

respectively, for any $d \times d$ non-singular matrix A and any vector $b \in \mathbb{R}^d$, where $AX^n + b = \{AX_1 + b, \cdots, AX_n + b\}$. They are called *rigid-body* or *translation* equivariant if (3) holds for any orthogonal A or any identity A ($A = I_d$), respectively. When $b = 0$ and $A = sI_d$ for a scalar $s \neq 0$, T and S are called *scale* equivariant. The following breakdown point upper bound results are due to Donoho (1982). We provide here a much simpler proof.

Lemma 1: *For any translation* (*scale*) *equivariant location* (*scatter*) *estimator* T (S) *at any sample* X^n *in* $\mathbb{R}^d$, $BP(T(S), X^n) \leq \lfloor (n+1)/2 \rfloor / n$.

Proof: It suffices to consider the location case. For $m = \lfloor (n+1)/2 \rfloor$ and $b \in \mathbb{R}^d$, let $Y_m^n = \{X_1 + b, \cdots, X_m + b, X_{m+1}, \cdots, X_n\}$. Both Y_m^n and $Y_m^n - b$ are data sets resulting from contaminating at most m points of X^n. Observe

$$\|b\| = \|T(Y_m^n) - T(Y_m^n - b)\| \leq \sup_{X_m^n} 2 \cdot \|T(X_m^n) - T(X^n)\| \to \infty \text{ as } \|b\| \to \infty.$$

Here (and hereafter) $\|\cdot\|$ is the Euclidean norm for a vector and $\|A\| = \sup_{\|u\|=1} \|Au\|$ for a matrix A. □

The coordinate-wise and the L_1 (also called *spatial*) medians are two known location estimators that can attain the breakdown point upper bound in the lemma; see Lopuhaä and Rousseeuw (1991), for example.

Both estimators, however, are not affine equivariant (the first is only translation and the second is just rigid-body equivariant). On the other hand, no scatter matrices constructed can reach the upper bound in the lemma. In fact, for affine equivariant scatter estimators and for data set X^n in a *general position* (that is, no more than d data points lie in the same $d-1$ dimensional hyperplane), Davies (1987) provided a negative answer and proved the following breakdown point upper bound result.

Lemma 2: *For any affine equivariant scatter estimator S and data set X^n in general position in $\mathbb{R}^d$, $BP(S, X^n) \le \lfloor (n-d+1)/2 \rfloor / n$.*

MAD is a univariate affine equivariant scale estimator that attains the upper bound in this lemma. Higher dimensional affine equivariant *scatter* estimators that reach this upper bound have been proposed in the literature. The following questions about *location* estimators, however, remain open:

(1) Is there any affine equivariant location estimator in high dimensions that can attain the breakdown point upper bound in Lemma 1? If not,
(2) What is the breakdown point upper bound of an affine equivalent location estimator?

A partial answer to the first question is given in Zuo (2004a) where under a slightly narrow definition of the finite sample breakdown point a location estimator attaining the upper bound in Lemma 1 is introduced.

2.3. *Maximum bias*

The point-mass contamination in the definition of influence function is very special. In practice, a deviation from the assumed distribution can be due to the contamination of any distribution. The influence function consequently measures a special *local* robustness of a functional $T(\cdot)$ at F. A very broad measure of *global* robustness of $T(\cdot)$ at F is the so-called maximum bias; see Huber (1964) and Huber (1981). Here any possible contaminating distribution G and the contaminated model $(1-\epsilon)F + \epsilon G$ are considered for a fixed $\epsilon > 0$ and the *maximum bias* of $T(\cdot)$ at F is defined as

$$B(\epsilon; T, F) = \sup_G \|T((1-\epsilon)F + \epsilon G) - T(F)\|. \tag{4}$$

$B(\epsilon; T, F)$ measures the worst case bias due to an ϵ amount contamination of the assumed distribution. $T(\cdot)$ is regarded as robust if it has a moderate maximum bias curve for small ϵ. It is seen that the standard mean and covariance functionals have an unbounded maximum bias for any $\epsilon > 0$ and hence are not robust in terms of this maximum bias measure.

The minimum contamination amount ϵ^* that can lead to an unbounded maximum bias is called the *asymptotic breakdown point* of T at F. Its finite sample version is exactly the one given by (2). On the other hand, if G is restricted to a point-mass contamination, then the rate of the change of $B(\epsilon; T, F)$ relative to ϵ, for ϵ arbitrarily small, is closely related to $IF(x; T, F)$. Indeed the "slope" of the maximum bias curve $B(\epsilon; T, F)$ at $\epsilon = 0$ is often the same as the supremum (over x) of $\|IF(x; T, F)\|$. Thus, the maximum bias really depicts the entire picture of the robustness of the functional T whereas the influence function and the breakdown point serve for two extreme cases. Though a very important robustness measure, the challenging derivation of $B(\epsilon; T, F)$ for a location or scatter functional T in high dimensions makes the maximum bias a less popular one than the influence function and the finite sample breakdown point in the literature.

To end this section, we remark that robustness is one of the most important performance criteria of a statistical procedure. There are, however, other important performance criteria. For example, efficiency is always a very important performance measure for any statistical procedure. In his seminal paper, Huber (1964) took into account both the robustness and the efficiency (in terms of the asymptotic variance) issues in the famous "minimax" (minimizing worst case asymptotic variance) approach. Robust estimators are commonly not very efficient. The univariate median serves as a perfect example. It is the most robust affine equivariant location estimator with the best breakdown point and the lowest maximum bias at symmetric distributions (see Huber 1964). Yet for its best robustness, it has to pay the price of low efficiencies relative to the mean at normal and other light-tailed models. In our following discussion about the robustness of location and scatter estimators, we will also address the efficiency issue.

3. Robust multivariate location and scatter estimators

This section surveys important affine equivariant robust location and scatter estimators in high dimensions. The efficiency issue will be addressed.

3.1. *M-estimators and variants*

As pointed out in Section 1, affine equivariant M-estimators of location and scatter parameters were the early robust alternatives to the classical sample mean vector and sample covariance matrix. Extending Huber (1964)'s idea of the univariate M-estimators as minimizers of objective functions, Maronna (1976) defined multivariate M-estimators as the solutions T (in $\mathbb{R}^d$) and V (a positive definite symmetric matrix) of

$$\frac{1}{n}\sum_{i=1}^{n} u_1(((X_i - T)'V^{-1}(X_i - T))^{1/2})(X_i - T) = 0, \tag{5}$$

$$\frac{1}{n}\sum_{i=1}^{n} u_2((X_i - T)'V^{-1}(X_i - T))(X_i - T)(X_i - T)' = V, \tag{6}$$

where u_i, $i = 1, 2$, are weight functions satisfying some conditions. They are a generalization of the *maximum likelihood estimators* and can be regarded as weighted mean and covariance matrix as well. Maronna (1976) discussed the existence, uniqueness, consistency, asymptotic normality, influence function and breakdown point of estimators. Though possessing bounded influence functions for suitable u_i's, $i = 1, 2$, T and V have relatively low breakdown points ($\leq 1/(d+1)$) (see, Maronna (1976) and p. 226 of Huber (1981)) and hence are not robust globally in high dimensions. The latter makes the M-estimators less appealing choices in robust statistics, though they can be quite efficient at normal and other models.

Tyler (1991) considered some sufficient conditions for the existence and uniqueness of M-estimators with special redescending weight functions. Constrained M-estimators, which combine both good local and good global robustness properties, are considered in Kent and Tyler (1996).

3.2. *Stahel-Donoho estimators and variants*

Stahel (1981) and Donoho (1982) "outlyingness" weighted mean and covariance matrix appear to be the first location and scatter estimators in high dimensions that can integrate affine equivariance with high breakdown points. In $\mathbb{R}^1$, the outlyingness of a point x with respect to (w.r.t.) a data set $X^n = \{X_1, \cdots, X_n\}$ is simply $|x - \mu(X^n)|/\sigma(X^n)$, the absolute deviation of x to the center of X^n standardized by the scale of X^n. Here μ and σ are univariate location and scale estimators with typical choices including (mean, standard deviation), (median, median absolute deviation), and more generally, univariate M-estimators of location and scale (see Huber (1964, 1981)). Mosteller and Tukey (1977) (p. 205) introduced an outlyingness weighted mean in $\mathbb{R}^1$. Stahel and Donoho (SD) considered a multivariate analog and defined the outlyingness of a point x w.r.t. X^n in $\mathbb{R}^d$ $(d \geq 1)$ as

$$O(x, X^n) = \sup_{\{u:\ u \in R^d,\ \|u\|=1\}} |u'x - \mu(u \cdot X^n)|/\sigma(u \cdot X^n) \tag{7}$$

where $u'x = \sum_{i=1}^d u_i x_i$ and $u \cdot X^n = \{u'X_1, \cdots, u'X_n\}$. If $u'x - \mu(u \cdot X^n) = \sigma(u \cdot X^n) = 0$, we define $|u'x - \mu(u \cdot X^n)|/\sigma(u \cdot X^n) = 0$. Then

$$T_{SD}(X^n) = \sum_{i=1}^n w_i X_i \Big/ \sum_{i=1}^n w_i, \tag{8}$$

$$S_{SD}(X^n) = \sum_{i=1}^n w_i (X_i - T_{SD}(X^n))(X_i - T_{SD}(X^n))' \Big/ \sum_{i=1}^n w_i \tag{9}$$

are the SD outlyingness weighted mean and covariance matrix, where $w_i = w(O(X_i, X^n))$ and w is a weight function down-weighting outlying points.

Since μ and σ^2 are usually affine equivariant, $O(x, X^n)$ is then *affine invariant*: $O(x, X^n) = O(Ax+b, AX^n+b)$ for any non-singular $d \times d$ matrix A and vector $b \in \mathbb{R}^d$. It follows that T_{SD} and S_{SD} are affine equivariant.

Stahel (1981) considered the asymptotic breakdown point of the estimators. Donoho (1982) derived the finite sample breakdown point for (μ, σ) being median (Med) and median absolute deviation (MAD), for X in a general position, and for suitable weight function w. His result, expressed in terms of *addition* breakdown point, amounts to (see, e.g., Zuo (2001))

$$BP(T_{SD}, X^n) = \frac{\lfloor (n - 2d + 2)/2 \rfloor}{n}, \quad BP(S_{SD}, X^n) = \frac{\lfloor (n - 2d + 2)/2 \rfloor}{n}.$$

Clearly, BPs of the SD estimators depend essentially on the BP of MAD (since Med already provides the best possible BP). As a scale estimator, MAD breaks down (*explosively* or *implosively*) as it tends to ∞ or 0. Realizing that it is easier to implode MAD with a projected data set $u \cdot X^n$ for X^n in high dimension (since there will be d overlapping projected points along some projection directions), Tyler (1994), Gather and Hilker (1997), and Zuo (2000) all modified MAD to get a higher BP of the SD estimators:

$$BP(T^*_{SD}, X^n) = \lfloor (n - d + 1)/2 \rfloor / n, \quad BP(S^*_{SD}, X^n) = \lfloor (n - d + 1)/2 \rfloor / n.$$

Note that the latter is the best possible BP result for S_{SD} by Lemma 2.

The SD estimators stimulated tremendous researches in robust statistics. Seeking *affine equivariant* estimators with *high BPs* indeed was one primary goal in the field in the last two decades. The asymptotic behavior of the SD estimators, however, was a long-standing problem. This hindered the estimators from becoming more popular in practice. Maronna and Yohai

(1995) first proved the $\sqrt{n}$-consistency. Establishing the limiting distributions, however, turned out to be extremely challenging. Indeed, there once were doubts in the literature about the existence or the normality of their limit distributions; see, e.g., Lopuhaä (1999) and Gervini (2003).

Zuo, Cui and He (2004) and Zuo and Cui (2005) studied general data depth weighted estimators, which include the SD estimators as special cases, and established a general asymptotic theory. The asymptotic normality of the SD estimators thus follows as a special case from the general results there. The robustness studies of the general data depth induced estimators carried out in Zuo, Cui and Young (2004) and Zuo and Cui (2005) also show that the SD estimators have bounded influence functions and moderate maximum bias curves for suitable weight functions. Furthermore, with suitable weight functions, the SD estimators can outperform most leading competitors in the literature in terms of robustness and efficiency.

3.3. *MVE and MCD estimators and variants*

Rousseeuw (1985) introduced affine equivariant *minimum volume ellipsoid* (MVE) and *minimum covariance determinant* (MCD) estimators as follows. The MVE estimators of location and scatter are respectively the center and the ellipsoid of the minimum volume ellipsoid containing (at least) h data points of X^n. It turns out that the MVE estimators can possess a very high breakdown point with a suitable h $(= \lfloor (n+d+1)/2 \rfloor)$ (Davies (1987)). They, however, are neither asymptotically normal nor $\sqrt{n}$ consistent (Davis (1992a)) and hence are not very appealing in practice. The MCD estimators are the mean and the covariance matrix of h data points of X^n for which the determinant of the covariance matrix is minimum. Again with $h = \lfloor (n+d+1)/2 \rfloor$, the breakdown point of the estimators can be as high as $\lfloor (n-d+1)/2 \rfloor / n$, the best possible BP result for any affine equivariant *scatter* estimator by Lemma 2; see Davies (1987) and Lopuhaä and Rousseeuw (1991). The MCD estimators have bounded influence functions that have jumps (Croux and Haesbroeck (1999)). The estimators are $\sqrt{n}$-consistent (Butler, Davies and Jhun (1993)) and the asymptotical normality is also established for the *location* part but not for the scatter part (Butler, Davies and Jhun (1993)). The estimators are not very efficient at normal models and this is especially true at the h selected in order for the estimators to have a high breakdown point; see Croux and Haesbroeck (1999). In spite of their low efficiency, the MCD estimators are quite popular in the literature, partly due to the availability of fast computing algorithms of the estimators (see, e.g., Hawkins (1994) and Rousseeuw and Van Driessen (1999)).

To overcome the low efficiency drawback of the MCD estimators, reweighted MCD estimators were introduced and studied; see Lopuhaä and Rousseeuw (1991), Lopuhaä (1999), and Croux and Haesbroeck (1999).

3.4. *S-estimators and variants*

Davis (1987) introduced and studied S-estimators for multivariate location and scatter parameters, extending an earlier idea of Rousseeuw and Yohai (1984) in regression context to the location and scatter setting. Employing a *smooth* ρ function, the S-estimators extend Rousseeuw's MVE estimators which are special S-estimators with a non-smooth ρ function. The estimators become $\sqrt{n}$-consistent and asymptotically normal. Furthermore they can have a very high breakdown point $\lfloor (n-d+1)/2 \rfloor /n$, again the upper bound for any affine equivariant *scatter* estimator; see Davies (1987). The S-estimators of location and scatter are defined as the vector T_n and the positive definite symmetric (PDS) matrix C_n which minimize the determinant of C_n, $\det(C_n)$, subject to

$$\frac{1}{n}\sum_{i=1}^{n} \rho\Big(((X_i - T_n) C_n^{-1} (X_i - T_n))^{1/2} \Big) \le b_0, \tag{10}$$

where the non-negative function ρ is symmetric and continuously differentiable and strictly increasing on $[0, c_0]$ with $\rho(0) = 0$ and constant on $[c_0, \infty)$ for some $c_0 > 0$ and $b_0 < a_0 := \sup \rho$. As shown in Lopuhaä (1989), S-estimators have a close connection with M-estimators and have bounded influence functions. They can be highly efficient at normal models; see Lopuhaä (1989) and Rocke (1996). The latter author, however, pointed out that there can be problems with the breakdown point of the S-estimators in high dimensions and provided remedial measures. Another drawback is that the S-estimators can not simultaneously attain a high breakdown point and a given efficiency at the normal models. Modified estimators that can overcome the drawback were given in Lopuhaä (1991, 1992) and Davies (1992b). The S-estimators can be computed with a fast algorithm such as the one given in Ruppert (1992).

3.5. *Depth weighted and maximum depth estimators*

Data depth has recently been increasingly pursued as a promising tool in multi-dimensional exploratory data analysis and inference. The key idea of data depth in the location setting is to provide a center-outward ordering of multi-dimensional observations. Points deep inside a data cloud receive high depth and those on the outskirts get lower depth. Multi-dimensional points then can be ordered based on their depth. Prevailing notions of data depth include Tukey (1975) halfspace depth, Liu (1990) simplicial depth and projection depth (Liu (1992), Zuo and Serfling (2000a) and Zuo (2003)). All these depth functions satisfy desirable properties for a general depth functions; see, e.g., Zuo and Serfling (2000b). Data depth has found applications to nonparametric and robust multivariate analysis. In the following we focus on the application to multivariate location and scatter estimators.

For a give sample X^n from a distribution F, let F_n be the empirical version of F based on X^n. For a general depth function $D(\cdot,\cdot)$ in $\mathbb{R}^d$, depth-weighted location and scatter estimators can be defined as

$$L(F_n) = \frac{\int x w_1(D(x,F_n))dF_n(x)}{\int w_1(D(x,F_n))dF_n(x)}, \tag{11}$$

$$S(F_n) = \frac{\int (x - L(F_n))(x - L(F_n))' w_2(D(x,F_n))dF_n(x)}{\int w_2(D(x,F_n))dF_n(x)}, \tag{12}$$

where w_1 and w_2 are suitable weight functions and can be different; see Zuo, Cui and He (2004) and Zuo and Cui (2005). These depth-weighted estimators can be regarded as generalizations of the univariate L-statistics. A similar idea is first discussed in Liu (1990) and Liu, Parelius and Singh (1999), where the depth-induced location estimators are called DL-statistics. Note that equations (11) and (12) include as special cases depth trimmed and Winsorized multivariate means and covariance matrices; see Zuo (2004b) for related discussions. With the projection depth (PD) as the underlying depth function, these equations lead to as special cases the Stahel-Donoho location and scatter estimators, where the projection depth is defined as

$$PD(x,F_n) = 1/(1 + O(x,F_n)), \tag{13}$$

where $O(x,F_n)$ is defined in (7). Replacing F_n with its population version F in (11), (12) and (13), we obtain population versions of above definitions.

Common depth functions are affine invariant. Hence $L(F_n)$ and $S(F_n)$ are affine equivariant. They are unbiased estimators of the center θ of symmetry of a symmetric F of X (i.e., $\pm(X-\theta)$ have the same distribution) and of the covariance matrix of an elliptically symmetric F, respectively; see Zuo, Cui and He (2004) and Zuo and Cui (2005). Under mild assumptions on w_1 and w_2 and for common depth functions, $L(F_n)$ and $S(F_n)$ are strongly consistent and asymptotically normal. They are locally robust with bounded influence functions and globally robust with moderate maximum biases and very high breakdown points. Furthermore, they can be extremely efficient at normal and other models. For details, see Zuo, Cui and He (2004) and Zuo and Cui (2005).

General depth weighted location and scatter estimators include as special cases the re-weighted estimators of Lopuhaä (1999) and Gervini (2003), where Mahalanobis type depth (see Liu (1992)) is utilized in the weight calculation of sample points. With appropriate choices of weight functions, the re-weighted estimators can possess desirable efficiency and robustness properties. Since Mahalanobis depth entails some initial location and scatter estimators, the performance of the re-weighted estimators depends crucially on the initial choices in both finite and large sample sense, though.

Another type of depth induced estimators is the maximum depth estimators, which could be regarded as an extension of the univariate median type estimators to the multivariate setting. For a given location depth function $D_L(\cdot,\cdot)$ and scatter depth function $D_S(\cdot,\cdot)$ and a sample X^n (or equivalently F_n), maximum depth estimators can be defined as

$$MDL(F_n) = \arg\sup_{x\in\mathbb{R}^d} D_L(x, F_n) \tag{14}$$

$$MDS(F_n) = \arg\sup_{\Sigma\in\mathcal{M}} D_S(\Sigma, F_n), \tag{15}$$

where $\mathcal{M}$ is the set of all positive definite $d\times d$ symmetric matrices. Aforementioned depth notions are all location depth functions. An example of the scatter depth function, given in Zuo (2004b), is defined as follows. For a given univariate scale measure σ, define the outlyingness of a matrix $\Sigma\in\mathcal{M}$ with respect to F_n (or sample X^n) as

$$O(\Sigma, F_n) = \sup_{u\in S^{d-1}} g\big(\sigma^2(u\cdot X^n)/u'\Sigma u\big), \tag{16}$$

where g is a nonnegative function on $[0,\infty)$ with $g(0)=\infty$ and $g(\infty)=\infty$; see, e.g., Maronna et al. (1992) and Tyler (1994). The (projection) depth of a scatter matrix $\Sigma\in\mathcal{M}$ then can be defined as (Zuo (2004b))

$$D_S(\Sigma, F_n) = 1/(1+O(\Sigma, F_n)). \tag{17}$$

A scatter depth defined in the same spirit was first given in Zhang (2002).

The literature is replete with discussions on location depth D_L and its induced deepest estimator $MDL(F_n)$; see, e.g., Liu (1990), Liu et al. (1999), Zuo and Serfling (2000a), Arcones et al. (1994), Bai and He (1999), Zuo (2003) and Zuo, Cui and He (2004). There are, however, very few discussions on scatter depth D_S and its induced deepest estimator $MDS(F_n)$ (exceptions are made in Maronna et al. (1992), Tyler (1994), Zhang (2002), and Zuo (2004b) though). Further studies on D_S and MDS such as robustness, asymptotics, efficiency, and inference procedures are called for.

Maximum depth estimators tend to be highly robust locally and globally as well. Indeed, the maximum projection depth estimators of location have bounded influence functions and moderately maximum biases; see Zuo, Cui and Young (2004). Figure 1 clearly reveals the boundedness of the influence functions of the maximum projection depth estimator (PM) (and the projection depth weighted mean (PWM)) with Med and MAD for μ and σ.

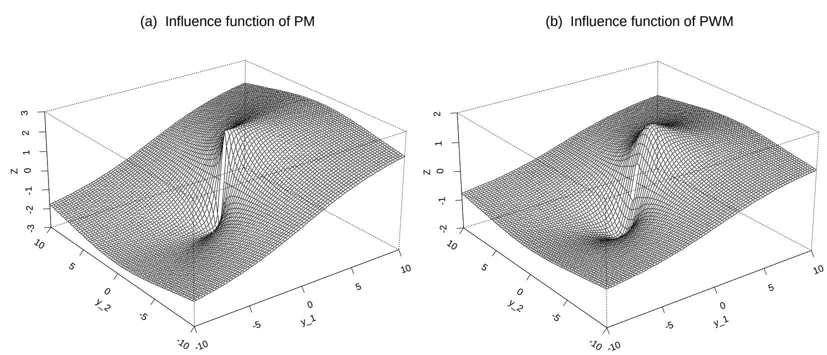

Fig. 1. (a) The first coordinate of the influence function of maximum projection depth estimator of location (projection median (PM)). (b) The first coordinate of the influence function of the projection depth weighted mean (PWM).

Maximum depth estimators can also possess high breakdown points. For example, both the maximum projection depth estimators of location and scatter can possess the highest breakdown points among their competitors,

$$BP(MDL,\, X^n) = \frac{\lfloor (n-d+2)/2 \rfloor}{n}, \quad BP(MDS,\, X^n) = \frac{\lfloor (n-d+1)/2 \rfloor}{n},$$

where PD is the depth function with Med and a modified version of MAD as μ and σ in its definition; see Zuo (2003) and Tyler (1994). Maximum depth estimators can also be highly efficient. For example, with appropriate choices of μ and σ, the maximum projection depth estimator of location can be highly efficient; see Zuo (2003) for details.

4. Applications

Robust location and scatter estimators find numerous applications to multivariate data analysis and inference. In the following we survey some major applications including robust Hotelling's T^2, robust multivariate control charts, robust principal component analysis, robust factor analysis, robust canonical correlation analysis and robust discrimination and clustering. We skip the application to the multivariate regression (see, e.g., Croux et al. (2001), Croux et al. (2003) and Rousseeuw et al. (2004) for related studies).

4.1. *Robust T^2 and control charts*

Hotelling's T^2: $n(\bar{X}-E(X))S^{-1}(\bar{X}-E(X))$ is the single most fundamental statistic in the classical inference about the multivariate mean vectors of populations as well as in the classical multivariate analysis of variance. It is also the statistic for the classical multivariate quality control charts. Built on the sample mean $\bar{X}$ and the sample covariance matrix S, T^2, unfortunately, is not robust. The T^2 based procedures also depend heavily on the normality assumption.

A simple and intuitive way to robustify the Hotelling's T^2 is to replace $\bar{X}$ and S with robust location and scatter estimators, respectively. An example was given in Willems et al. (2002), where re-weighted MCD estimators were used instead of the mean and the covariance matrix. A major issue here is the (asymptotic) distribution of the robust version of T^2 statistic. Based on the multivariate sign and sign-rank tests of Randles (1989), Peters and Randles (1991) and Hettmansperger et al. (1994), robust control charts are constructed by Ajmani and Vining (1998) and Ajmani et al. (1998).

Another approach to construct robust multivariate quality charts is via data depth. Here a quality index is introduced based on the depth of points and the multivariate processes are monitored based on the index. Representative studies include Liu (1995) and Liu and Singh (1993). Others include Ajmani et al. (1997) and Stoumbos and Allison (2000).

Finally, the projection (depth) pursuit idea has also been employed to construct multivariate control charts; see, e.g., Ngai and Zhang (2001).

4.2. *Robust principal component analysis*

Classical principal component analysis (PCA) is carried out based on the eigenvectors (eigenvalues) of the sample covariance (or correlation) matrix. Such analysis is extremely sensitive to outlying observations and the conclusions drawn based on the principal components may be adversely affected by the outliers and misleading. A most simple and appealing way to robustify the classical PCA is to replace the matrix with a robust scatter estimator. Robust PCA studies started in as early as 1970's and include Maronna (1976), Campbell (1980) and Devlin et al. (1981), where M-estimators of location and scatter were utilized instead of the sample mean and covariance matrix. Some recent robust PCA studies focus on the investigation of the influence function of the eigenvectors and eigenvalues; see, e.g., Jaupi and Saporta (1993), Shi (1997) and Croux and Haesbroeck (2000).

A different approach to robust PCA uses projection pursuit (PP) techniques; see Li and Chen (1985), Croux and Ruiz-Gazen (1996) and Hubert et al. (2002). It seeks to maximize a robust measure of spread to obtain consecutive directions along which the data points are projected. This idea has been generalized to common principal components in Boente et al. (2002).

Recently, Hubert et al. (2005) combined the advantages of the above two approaches and proposed a new method to robust PCA where the PP part is used for the initial dimension reduction and then the ideas of robust scatter estimators are applied to this lower-dimensional data space.

4.3. *Robust factor analysis*

The classical factor analysis (FA) starts with the usual sample covariance (or correlation) matrix and then the eigenvectors and eigenvalues of the matrix are employed for estimating the loading matrix (or the matrix is used in the likelihood equation to obtain the maximum likelihood estimates of the loading matrix and specific variances). The analysis, however, is not robust since outliers can have a large effect on the covariance (or correlation matrix) and the results obtained may be misleading or unreliable.

A straightforward approach to robustify the classical FA is to replace the sample covariance (or correlation) matrix with a robust one. One such example was given in Pison et al. (2003) where MCD estimators were employed. Further systematic studies on robust FA such as robustness, efficiency and performance, and inference procedures are yet to be conducted.

4.4. *Robust canonical correlation analysis*

The classical canonical correlation analysis (CCA) seeks to identify and quantify the associations between two sets of variables. It focuses on the correlation between a linear combination of the variables in one set and a linear combination of the variables in another set. The idea is to determine first the pair of linear combinations having the largest correlation, then the next pair of linear combinations having the largest correlation among all pairs uncorrelated with the previous selected pair, and so on. In practice, sample covariance (or correlation) matrix is utilized to achieve the goal. The result obtained, however, is not robust to outliers in the data since the sample covariance (or correlation) matrix is extremely sensitive to unusual observations. To robustify the classical approach, Karnel (1991) proposed to use M-estimators and Croux and Dehon (2002) the MCD estimators. The latter paper also studied the influence functions of canonical correlations and vectors. Robustness and asymptotics of robust CCA were discussed in Taskinen et al. (2005). More studies on robust CCA are yet to be seen.

4.5. *Robust discrimination, classification and clustering*

In the classical discriminant analysis and classification, the sample mean and the sample covariance matrix are often used to build discriminant rules which however are very sensitive to outliers in data. Robust rules can be obtained by inserting robust estimators of location and scatter into the classical procedures. Croux and Dehon (2001) employed S-estimators to carry out a robust linear discriminant analysis. A robustness issue related to the quadratic discriminant analysis is addressed by Croux and Joossens recently. He and Fung (2000) discussed the high breakdown estimation and applications in discriminant analysis. Hubert and Van Driessen (2004) discussed fast and robust discriminant analysis based on MCD estimators.

In the classical clustering methods, the sample mean and the sample covariance matrix likewise are often employed to build clustering rules. Robust estimators of location and scatter could be used to replace the mean vector and the covariance matrix to obtain robust clustering rules. References on robust clustering methods include Kaufman and Rousseeuw (1990). Robust clustering analysis is a very active research area of computer scientists; see, e.g., Davé and R Krishnapuram (1997) and Fred and Jain (2003) and references therein. More studied on clustering analysis from statistical perspective with robust location and scatter estimators are needed.

5. Conclusions and future works

Simulation studied by Maronna and Yohai (1995), Gervini (2002), Zuo, Cui and He (2004), Zuo, Cui and Young (2004) and Zuo and Cui (2005) indicate that the projection depth weighted mean and covariance matrix (the Stahel-Donoho estimators) with suitable weight functions can outperform most of its competitors in terms of local and global robustness as well as efficiency at a number of distribution models. We thus recommend the Stahel-Donoho estimators and more generally projection depth weighted mean and covariance matrix as favorite choices of robust location and scatter estimators. Maximum depth estimators of location and scatter are strong competitors, especially from robustness view point. They (especially maximum depth scatter estimators) deserve further attention and investigations.

Computing high breakdown point robust affine equivariant location and scatter estimators is always a challenging task and there is no exception for the projection depth related estimators. Recent studies of this author, however, indicate that some of these estimators can be computed exactly in two and higher dimensions for robust μ and σ such as Med and MAD. Though fast approximate algorithms for computing these estimators already exist for moderately high dimensional data, issues involving the computing of these depth estimators such as how accurate and how robust are the approximate algorithms are yet to be addressed.

At this point, all applications of robust location and scatter estimators to multivariate data analysis are centered around the MCD based procedures. Since MCD estimators are not very efficient and can sometime have unstable behavior, we thus recommend replacing MCD estimators with the projection depth weighted estimators and expect that more reliable and efficient procedures are to be obtained. Asymptotic theory involving the robust multivariate analysis procedures is yet to be established.

Finally we comment that data depth is a natural tool for robust multivariate data analysis and more researches along this direction which can lead to very fast, robust, and efficient procedures are needed.

6. Acknowledgments

The author thanks Professors Hira Koul and Catherine Dehon for their insightful and constructive comments and suggestions. The work is partly supported by NSF grant DMS-0234078.

References

1. V. Ajmani and G. Vining (1998). A ronust multivariate exponentially weighted moving average chart. Preprint.
2. V. Ajmani, R. Randles and G. Vining (1998). Robust multivariate control charts. Preprint.
3. V. Ajmani, R. Randles, G. Vining and W. Woodall (1997). Robustness of multivariate control charts. In *Proceedings of the Section on Quality and Productivity, 1997 Quality and Productivity Reserach Conference*, American Statistical Association, Alexandria, VA. 152-157.
4. M. A. Arcones, Z. Chen and E. Giné (1994). Estimators related to U-

processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.* **22** 1460-1477.
5. Z. D. Bai and X. He (1999). Asymptotic distributions of the maximum depth estimators for regression and multivariate location. *Ann. Statist.* **27** 1616-1637.
6. P. J. Bickel (1964). On some alternative estimates for shift in the p-variate one sample problem. *Ann. Math. Statist.* **35** 1079-1090.
7. P. J. Bickel (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847-858.
8. G. Boente, A. M. Pires and I. Rodrigues (2002). Influence functions and outlier detection under the common principal components model: a robust approach. *Biometrika* **89** 861-875.
9. R. W. Butler, P. L. Davies and M. Jhun (1993). Asymptotics for the minimum covariance determinant estimator. *Ann. Statist.* **21** 1385-1400.
10. N. A. Campbell (1980). Robust procedures in multivariate analysis: robust covariance estimation. *Appl. Statist.* **29** 231-237.
11. C. Croux and C. Dehon (2001). Robust linear discriminant analysis using S-estimators. *Canad. J. Statist.* **29** 473-492.
12. C. Croux and C. Dehon (2002). Analyse canonique Basée sur des estimateurs robustes de la matrice de covariance. *La Revue de Statistique Appliquee* **2** 5-26
13. C. Croux, C. Dehon, S. Van Aelst, and P. J. Rousseeuw (2001). Robust Estimation of the Conditional Median Function at Elliptical Models. *Statist. Probab. Letters* **51** 361-368.
14. C. Croux and G. Haesbroeck (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *J. Multivariate Anal.* **71** 161-190.
15. C. Croux and G. Haesbroeck (2000). Principal compoment analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika* **87** 603-618.
16. C. Croux and K. Joossens. Influence of Observations on the Misclassification Probability in Quadratic Discriminant Analysis *J. Multivariate Anal.* (to appear).
17. C. Croux and A. Ruiz-Gazen (1996). A fast algorithm for robust principal components based on projection pursuit. In *COMPSTAT 1996, Proceedings in Computational Statistics* (Ed A. Prat). Heidelberg, Physica-Verlag. 211-217.
18. C. Croux, S. Van Aelst, and C. Dehon (2003). Bounded Influence Regression using High Breakdown Scatter Matrices. *Ann. Instit. Statist. Math.* **55** 265-285.
19. R. N. Dav and R. Krishnapuram (1997). Robust clustering methods: A unified view. *IEEE Trans. Fuzzy Syst.* **5** 270-293.
20. P. L. Davies (1987). Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269-1292.
21. P. L. Davies (1992a). The asymptotics of rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist.* **20** 1828-1843.

22. P. L. Davies (1992b). An efficient Fréchet-differentiable high breakdown multivariate location and dispersion estimator. *J. Multivariate Anal.* **40** 311-327.
23. S. J. Devlin, R. Gnanadesikan andJ. R. Kettenring (1981). Robust estimation of dispersion matrics and principal components. *J. Amer. Statist. Assoc.* **76** 354-762.
24. D. L. Donoho (1982). Breakdown properties of multiavraite location estiamtors. Ph.D. thesis, Harvard University.
25. D. L. Donoho and P. J. Huber (1983). The notion of breakdown point. In *A Festschrift for Erich Lehmann* (Eds. P. J. Bickel, K. Doksum, and J. L. Hodges, Jr.). Wadsworth, Belmont, CA. 157-184.
26. A. Fred and A. K. Jain (2003). Robust data clustering. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition II.* 128-133.
27. U. Gather and T. Hilker (1997). A note on Tyler's modification of the MAD for the Stahel-Donoho estimator. *Ann. Statist.* **25** 2024-2026.
28. D. Gervini (2002). The influence function of the Stahel-Donoho estimator of multivariate location and scatter. *Statist. Probab. Lett.* **60** 425-435.
29. D. Gervini (2003). A robust and efficient adaptive reweighted estimator of multivariate location and scatter. *J. Multivariate Anal.* **84** 116-144.
30. R. Gnanadesikan and J. R. Kettenring (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28** 81-124.
31. F. R. Hampel (1968). Contributions to the theory of robust estimation. Ph.D. dissertion, University of California, Berkeley, CA.
32. F. R. Hampel (1971). A general qualitative definition of robustness. *Ann. Math. Statis.* **42** 1887-1896.
33. F. R. Hampel (1973). Robust estimation: A condensed partial survey. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **27** 87-104.
34. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The approach based on influence functions.* John Wiley & Sons, New York.
35. D. M. Hawkins (1994). The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data. *Comput. Statist. Data Anal.* **17** 197-210.
36. X. He and W. K. Fung (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *J. Multivariate Anal.* **72** 151-162.
37. P. J. Huber (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.
38. P. J. Huber (1972). Robust Statistics: A review. *Ann. Math. Statist.* **35** 1041-1067.
39. P. J. Huber (1977). Robust covariances. In *Statistical Decision Theory and Related Topics* (Eds. S. S. Gupta and D. S. Moore). Academic Press, New York. 165-191.
40. P. J. Huber (1981). *Robust Statistics.* (John Wiley & Sons, New York.
41. M. Hubert, P. J. Rousseeuw and S. Verboven (2002). A fast method for

robust principal components with applications to chemometrics. *Chemom. Intell. Lab. Sys.* **60** 101-111.
42. M. Hubert, P. J. Rousseeuw and K. Vanden Branden (2005). ROBPCA: a new approach to robust principal component analysis. *Technometrics* **47** 64-79.
43. M. Hubert and K. Van Driessen (2004). Fast and robust discrimant analysis. *Comput. Statist. Data Anal.* **45** 301-320.
44. L. Jaupi and G. Saporta (1993). Using the influence function in robust principal components analysis. In *New Directions in Statistical Data Analysis and Robustness* (Eds, S. Morgenthaler, E. Ronchetti and W. A. Stahel). Basel, Birkhäuser. 147-156.
45. G. Karnel (1991). Robust canonical correlation and correspondance analysis. In *The Frontiers of Statistical Scientific and Industrial Applications*, Volume II of the Proceedings of ICOSCO-I. 335-354.
46. L. Kaufman and P. J. Rousseeuw (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley-Interscience, New York.
47. G. Li and Z. Chen (1985). Projection-Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo. *J. Amer. Statist. Assoc.* **80** 759-766.
48. R. Y. Liu (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405-414.
49. R. Y. Liu (1992). Data depth and multivariate rank tests. In L_1-*Statistical Analysis and Related Mathods* (Ed. Y. Dodge). North-Holland, Amsterdam. 279-294.
50. R. Y. Liu (1995). Control Charts for Multivariate Processes. *J. Amer. Statist. Assoc.* **90** 1380-1387.
51. R. Y. Liu, J. M. Parelius, and K. Singh (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. Statist.* **17** 1662-1683.
52. R. Y. Liu and K. Singh (1993). A quality index based on data depth and multivariate rand tests. *J. Amer. Statist. Assoc.* **88** 252-260.
53. H. P. Lopuhaä (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17** 1662-1683.
54. H. P. Lopuhaä (1991). Multivarite τ-estimators for location ans scatter. *Can. J. Statist.* **19** 307-321.
55. H. P. Lopuhaä (1992). Highly efficient estimators of multivariate location with high breakdown point. *Ann. Statist.* **20** 398-413.
56. H. P. Lopuhaá (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *Ann. Statist.* **27** 1638-1665.
57. H. P. Lopuhaä and P. J. Rousseeuw (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229-248.
58. R. A. Maronna (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **1** 51-67.
59. R. A. Maronna and V. J. Yohai (1995). The behavior of the Stahel-Donoho

robust multivariate estimator. *J. Amer. Statist. Assoc.* **90** 330-341.
60. R. A. Maronna and V. J. Yohai (1998). Robust estimation of multivariate location and scatter. In *Encyclopedia of Statistical Sciences, Updated volume 2* (Eds. S. Kotz, C. Read and D. Banks). Wiley, New York. 589-596.
61. R. A. Maronna, W. A. Stahel and V. J. Yohai (1992). Bias-robust estimates of multivariate scatter based on projections. *J. Multivariate Anal.* **42** 141-161.
62. F. Mosteller and J. W. Tukey (1997). *Data Analysis and Regression.* Addison Wesley, Reading, MA.
63. H. Ngai and J. Zhang (2001). Multivariate cumulative sum control charts based on projection pursuit. *Statist. Sinica* **11** 747-766.
64. G. Pison, P. J. Rousseeuw, P. Filzmoser and C. Croux (2003). Robust Factor Analysis. *J. Multivariate Anal.* **84** 145-172.
65. D. M. Rocke (1996). Robustness Properties of S-estimators of multivariate location and shape in high dimension. *Ann. Statist.* **24** 1327-1345.
66. P. J. Rousseeuw (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications* (Eds. W. Grossmann, G. Pflug, I. Vincze and W. Wertz). Reidel. 283-297.
67. P. J. Rousseeuw and A. M. Leroy (1987). *Robust Regression and Outlier Detection.* Wiley, New York.
68. P. J. Rousseeuw, S. Van Aelst, K. Van Driessen, and J. Agullió (2004). Robust multivariate regression. *Technometrics* **46** 293-305.
69. P. J. Rousseeuw and K. Van Driessen (1999). A fast Algorithm for the minimum covariance determinant estimator. *Technometrics* **41** 212-223.
70. P. J. Rousseeuw and V. J. Yohai (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* Springer, New York. **26** 256-272.
71. D. Ruppert (1992). Computing S-Estimators for Regression and Multivariate Location/Dispersion. *J. Comput. Graph. Statist.* **1** 253-270
72. R. J. Serfling (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.
73. L. Shi (1997). Local influence in principal components analysis. *Biometrika* **84** 175-186.
74. W. A. Stahel (1981). Robust estimation: Infinitesimal optimality and covariance matrix estimators. Ph.D. thesis, ETH, Zurich.
75. S. Taskinen, C. Croux, A. Kankainen, E. Ollila and H. Oja. Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *J. Multivariate Anal.* (to appear).
76. J. W. Tukey (1949). Memorandum Reports 31-34, *Statist. Research Group.* Princeton Univ.(unpublished).
77. J. W. Tukey (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* (Ed. I. Olkin). Standford Univ. Press. 448-485.
78. J. W. Tukey (1975). Mathematics and the picturing of data. In *Proc. International Congress of Mathematicians Vancouver 1974.* Canadian Mathematics Congress, Montreal. **2** 523-531.

79. D. E. Tyler (1994). Finite sample breakdown points of projection based multivarite location and scatter statistics. *Ann. Statist.* **22** 1024-1044.
80. G. Willems, G. Pison, P. J. Rousseeuw and S. Van Aelst (2002). A Robust Hotelling Test. *Metrika* **55** 125-138.
81. J. Zhang (2002). Some extensions of Tukey's depth function. *J. Multivariate Anal.* **82** 134-165.
82. Y. Zuo (2000). A note of finite sample breakdown points of projection based multivariate location and scatter statistics. *Metrika* **51** 259-265.
83. Y. Zuo (2001). Some quantitative relationships between two types of finite sample breakdown point. *Statit. Probab. Lett.* **51** 369-375.
84. Y. Zuo (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460-1490.
85. Y. Zuo (2004a). *Statistica Sinica* **14** 1199-1208.
86. Y. Zuo (2004b). Statistical depth functions and some applications. *Adv. Math.* (China) **33** 1-25.
87. Y. Zuo and H. Cui (2005). Depth weighted scatter estiamtors. *Ann. Statist.* **33** 381-413.
88. Y. Zuo, H. Cui and X. He (2004). On the Stahel-Donoho estimator and depth weighted means for multiavraite data. *Ann. Statist.* **32** 167-188.
89. Y. Zuo, H. Cui and X. He (2004). Influence function and maximum bias of projection depth based estimators. *Ann. Statist.* **32** 189-218.
90. Y. Zuo and R. Serfling (2000a). General notions of statistical depth function. *Ann. Statist.* **28** 461-482.
91. Y. Zuo and R. Serfling (2000b). Structural properties and convergence results for contours of sample statistical depth functons. *Ann. Statist.* **28** 483-499.