COMMENTS ON BREAKDOWN

David J. Olive *

Southern Illinois University

June 10, 2008

Abstract

It is shown that the property of being a high breakdown estimator is weaker than the property of being an asymptotically unbiased estimator. Hence the breakdown and maximal bias properties should only make up a small part of a research paper. In particular, papers solely on breakdown or maximal bias should no longer be published.

KEY WORDS: LMS; LTA; LTS; MCD; MVE; Outliers; Robust Regression.

*David J. Olive is Associate Professor, Department of Mathematics, Southern Illinois University, Mailcode 4408, Carbondale, IL 62901-4408, USA. This work was supported by the National Science Foundation under grant DMS 0600933.

Introduction

The *multiple linear regression (MLR) model* is $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients and $\boldsymbol{e}$ is an $n \times 1$ vector of errors. The ith case $(\boldsymbol{x}_i^T, Y_i)$ corresponds to the ith row $\boldsymbol{x}_i^T$ of $\boldsymbol{X}$ and the ith element of $\boldsymbol{Y}$.

A *multivariate location and dispersion (MLD) model* is a joint distribution for a $p \times 1$ random vector $\boldsymbol{x}$ that is completely specified by a $p \times 1$ population *location* vector $\boldsymbol{\mu}$ and a $p \times p$ symmetric positive definite population *dispersion* matrix $\boldsymbol{\Sigma}$. Elliptically contoured distributions are important MLD models. The data is collected in an $n \times p$ matrix $\boldsymbol{X}$ with n rows $\boldsymbol{x}_i^T$.

Let the $p \times 1$ column vector $T(\boldsymbol{X})$ be a multivariate location estimator, and let the $p \times p$ symmetric positive definite matrix $\boldsymbol{C}(\boldsymbol{W})$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

$$D_i^2 = D_i^2(T(\boldsymbol{X}), \boldsymbol{C}(\boldsymbol{X})) = (\boldsymbol{x}_i - T(\boldsymbol{X}))^T \boldsymbol{C}^{-1}(\boldsymbol{X})(\boldsymbol{x}_i - T(\boldsymbol{X})) \tag{1.1}$$

for each observation $\boldsymbol{x}_i$. Notice that the Euclidean distance of $\boldsymbol{x}_i$ from the estimate of center $T(\boldsymbol{X})$ is $D_i(T(\boldsymbol{X}), \boldsymbol{I}_p)$ where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. The classical Mahalanobis distance corresponds to the sample mean and sample covariance matrix

$$T(\boldsymbol{X}) = \overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \quad \text{and} \quad \boldsymbol{C}(\boldsymbol{X}) = \boldsymbol{S} = \frac{1}{\mathrm{n}-1}\sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{x}_\mathrm{i} - \overline{\boldsymbol{x}})(\boldsymbol{x}_\mathrm{i} - \overline{\boldsymbol{x}})^\mathrm{T}.$$

Many of the most used practical "robust estimators" generate a sequence of K trial fits called *attractors*: $\boldsymbol{b}_1, ..., \boldsymbol{b}_K$ for MLR and $(T_1, \boldsymbol{C}_1), ..., (T_K, \boldsymbol{C}_K)$ for MLD. Then some criterion is evaluated and the attractor $\boldsymbol{b}_A$ or $(T_A, \boldsymbol{C}_A)$ that minimizes the criterion is used as the final estimator. One way to obtain attractors is to generate trial fits called *starts*, and then use the *concentration* technique. For multivariate data, let $(T_{0,j}, \boldsymbol{C}_{0,j})$ be the jth start and compute all n Mahalanobis distances $D_i(T_{0,j}, \boldsymbol{C}_{0,j})$. At the next iteration, the classical estimator $(T_{1,j}, \boldsymbol{C}_{1,j})$ is computed from the $c_n \approx n/2$ cases corresponding to the smallest distances. This iteration can be continued for k steps resulting in the sequence of estimators $(T_{0,j}, \boldsymbol{C}_{0,j}), (T_{1,j}, \boldsymbol{C}_{1,j}), ..., (T_{k,j}, \boldsymbol{C}_{k,j})$. Then $(T_{k,j}, \boldsymbol{C}_{k,j}) = (\overline{\boldsymbol{x}}_{k,j}, \boldsymbol{S}_{k,j})$ is the jth attractor. For MLR, let $\boldsymbol{b}_{0,j}$ be the jth start and compute all n residuals $r_i(\boldsymbol{b}_{0,j}) = Y_i - \boldsymbol{b}_{0,j}^T \boldsymbol{x}_i$. At the next iteration, the OLS estimator $\boldsymbol{b}_{1,j}$ is computed from the $c_n \approx n/2$ cases corresponding to the smallest squared residuals. This iteration can be continued for k steps resulting in the sequence of estimators $\boldsymbol{b}_{0,j}, \boldsymbol{b}_{1,j}, ..., \boldsymbol{b}_{k,j}$. Then $\boldsymbol{b}_{k,j}$ is the jth attractor for $j = 1, ..., K$. Using $k = 10$ concentration steps often works well, and the basic resampling algorithm is a special case with $k = 0$, i.e., the attractors are the starts.

A common method for generating starts is to use randomly selected "elemental sets" of p cases for MLR and $p+1$ cases for MLD. The jth elemental fit is a classical estimator $\boldsymbol{b}_j$ or $(T_j, \boldsymbol{C}_j)$ computed from the jth elemental set.

Hampel, Ronchetti, Rousseeuw and Stahel (1986, p. 96-98) and Donoho and Huber (1983) provide some history for breakdown. Maguluri and Singh (1997) have interesting examples on breakdown. If d of the cases have been replaced by arbitrarily bad contaminated cases, then the contamination fraction is $\gamma = d/n$. Then the breakdown value of $\hat{\boldsymbol{\beta}}$ or of a multivariate location estimator is the smallest value of γ needed to make $\|\hat{\boldsymbol{\beta}}\|$ or $\|T\|$ arbitrarily large. Let $0 \leq \lambda_p(\boldsymbol{C}(\boldsymbol{X})) \leq \cdots \leq \lambda_1(\boldsymbol{C}(\boldsymbol{X}))$ denote the eigenvalues of the dispersion estimator applied to data $\boldsymbol{X}$. The breakdown value of a dispersion estimator $\boldsymbol{C}$ is the smallest value of γ needed to drive the smallest eigenvalue to zero or the largest eigenvalue to ∞.

Breakdown for MLR

The following result greatly simplifies some breakdown proofs and shows that an MLR estimator basically breaks down if the median absolute residual MED($|r_i|$) can be made arbitrarily large. The result implies that if the breakdown value ≤ 0.5, breakdown can be computed using the median absolute residual MED($|r_i|(\boldsymbol{W}_d^n)$) instead of $\|T(\boldsymbol{W}_d^n)\|$.

Suppose that the proportion of outliers is less that 0.5. If the $\boldsymbol{x}_i$ are fixed, and the outliers are moved up and down the Y axis, then for high breakdown (HB) estimators, $\hat{\boldsymbol{\beta}}$ and MED($|r_i|$) will eventually be bounded. Thus if the $|Y_i|$ values of the outliers are large enough, the $|r_i|$ values of the outliers will be large.

If the Y_i's are fixed, arbitrarily large $\boldsymbol{x}$-outliers tend to drive the slope estimates to 0, not ∞. If both $\boldsymbol{x}$ and Y can be varied, then a cluster of outliers can be moved arbitrarily far from the bulk of the data but still have small residuals. For example, move the outliers along the regression plane formed by the clean cases.

Let $\boldsymbol{W}$ be the $n \times (p+1)$ matrix with ith row $(\boldsymbol{x}_i^T, Y_i)$. Let $\boldsymbol{W}_d^n$ denote the data matrix where any d of the cases have been replaced by arbitrarily bad contaminated cases.

Proposition 1 (Olive 2005): If the breakdown value ≤ 0.5, computing the breakdown value using the median absolute residual MED($|r_i|$) is asymptotically equivalent to using $\|\hat{\boldsymbol{\beta}}\|$.

Proof: Consider a fixed data set $\boldsymbol{W}_d^n$ with ith row $(\boldsymbol{w}_i^T, Z_i)^T$. If the regression estimator $T(\boldsymbol{W}_d^n) = \hat{\boldsymbol{\beta}}$ satisfies $\|\hat{\boldsymbol{\beta}}\| = M$ for some constant M, then the median absolute residual $\text{MED}(|Z_i - \hat{\boldsymbol{\beta}}^T \boldsymbol{w}_i|)$ is bounded by $\max_{i=1,\ldots,n} |Y_i - \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i| \leq \max_{i=1,\ldots,n}[|Y_i| + \sum_{j=1}^p M|x_{i,j}|]$ if $d < n/2$.

Now suppose that $\|\hat{\boldsymbol{\beta}}\| = \infty$. Since the absolute residual is the vertical distance of the observation from the hyperplane, the absolute residual $|r_i| = 0$ if the ith case lies on the regression hyperplane, but $|r_i| = \infty$ otherwise. Hence $\text{MED}(|r_i|) = \infty$ if fewer than half of the cases lie on the regression hyperplane. This will occur unless the proportion of outliers $d/n > (n/2 - q)/n \to 0.5$ as $n \to \infty$ where q is the number of "good" cases that lie on a hyperplane of lower dimension than p. In the literature it is usually assumed that the original data is in general position: $q = p - 1$. QED

If the contamination is high but less than 0.5, then $||\hat{\boldsymbol{\beta}}||$ and $\text{MED}(|r_i|)$ stay bounded. If the $\boldsymbol{x}$ values of the outliers are fixed, add a constant c to Y values of the outliers. As $|c|$ goes to infinity, the outliers will eventually have the largest squared residuals. Notice that if the HB estimator is $\hat{Y} = x$, then the point $(10^9, 10^9)$ will have zero residual. For estimators like LMS, LTS and LTA that have a narrowest band interpretation, let the outliers tilt the narrowest band away from the true regression plane $\boldsymbol{x}^T\boldsymbol{\beta}$, then slide the outliers up and down the narrowest band. They will have small squared residuals although both the $\boldsymbol{x}$ and Y values can be arbitrarily large. The correlation of the clean cases is important in that if $\boldsymbol{x}^T\boldsymbol{\beta}$ and Y are highly correlated, then the narrowest band is more narrow and $|c|$ is smaller. The limiting cases are exact fit with band width of 0 and correlation 1, and 0 correlation where $|c|$ is huge.

Rousseeuw and Leroy (1987, p. 29, 206) conjectured that high breakdown (HB) regression estimators can not be computed cheaply, and that if the algorithm is also affine equivariant, then the complexity of the algorithm must be at least $O(n^p)$. The following counterexample shows that these two conjectures are false.

Proposition 2 (Olive 2005): Suppose that the MLR model has an intercept β_1. Let $\hat{\boldsymbol{\beta}}_J$ be the OLS estimator applied to the set J of approximately $n/2$ cases that have $Y_i \in [\text{MED}(Y_i) \pm \text{MAD}(Y_i)]$ where $\text{MED}(Y_i)$ is the median and $\text{MAD}(Y_i) = \text{MED}(|Y_i - \text{MED}(Y_i)|)$ is the median absolute deviation of the response variable. Then $\hat{\boldsymbol{\beta}}_J$ is an affine equivariant HB regression estimator.

Proof. Consider the estimator

$$\hat{\boldsymbol{\beta}}_M = (\text{MED}(Y_i), 0, ..., 0)^T$$

which yields the predicted values $\hat{Y}_i \equiv \text{MED}(Y_i)$. The squared residual $r_i^2(\hat{\boldsymbol{\beta}}_M) \leq (\text{MAD}(Y_i))^2$ if the ith case is in J. Hence the OLS fit $\hat{\boldsymbol{\beta}}_J$ to the cases in J has

$$\sum_{i \in J} r_i^2(\hat{\boldsymbol{\beta}}_J) \leq n(\text{MAD}(Y_i))^2,$$

and

$$\text{MED}(|r_i(\hat{\boldsymbol{\beta}}_J)|) \leq \sqrt{n}\text{MAD}(Y_i) < \infty$$

if $\text{MAD}(Y_i) < \infty$. Hence the estimator $\hat{\boldsymbol{\beta}}_J$ is HB, but it only resists large Y–outliers. $\hat{\boldsymbol{\beta}}_J$ is affine equivariant because the cases that determine the OLS fit do not depend on $\boldsymbol{X}$. (Note that $\hat{\boldsymbol{\beta}}_J$ is scale but not regression equivariant.) QED

Proposition 3 (Hawkins and Olive 2002). The breakdown value of MLR concentration algorithms that use K elemental starts is bounded above by K/n.

Proof: To cause an MLR algorithm to break down, simply contaminate one observation in each starting elemental set so as to displace the fitted coefficient vector by a large amount. Since K elemental starts are used, at most K points need to be contaminated. QED

Proposition 4. If a high breakdown start is added to an MLR concentration algorithm, then the resulting estimator is HB.

Proof: The MLR algorithm uses the LTS(c_n) criterion, and concentration reduces the HB LTS criterion that is based on $c_n \geq n/2$ absolute residuals, so the median absolute residual of the resulting estimator is bounded as long as the criterion applied to the HB estimator is bounded. QED

For example, suppose the ordered squared residuals from the mth start $\boldsymbol{b}_{0m}$ are obtained. Then $\boldsymbol{b}_{1m}$ is simply the OLS fit to the cases corresponding to the c_n smallest squared residuals. Denote these cases by $i_1, ..., i_{c_n}$. Then

$$\sum_{i=1}^{c_n} r^2_{(i)}(\boldsymbol{b}_{1m}) \leq \sum_{j=1}^{c_n} r^2_{i_j}(\boldsymbol{b}_{1m}) \leq \sum_{j=1}^{c_n} r^2_{i_j}(\boldsymbol{b}_{0m}) = \sum_{j=1}^{c_n} r^2_{(i)}(\boldsymbol{b}_{0m})$$

where the second inequality follows from the definition of the OLS estimator. Hence concentration steps reduce the LTS criterion. If $c_n = (n+1)/2$ for n odd and $c_n = 1+n/2$ for n even, then the criterion is bounded iff the median squared residual is bounded.

Notice that if $\boldsymbol{b}_{0m} = \hat{\boldsymbol{\beta}}_J$ is the $m = (K+1)$th start, then the attractor $\boldsymbol{b}_{km}$ found after k concentration steps is also a HB regression estimator. Let $\hat{\boldsymbol{\beta}}_{k,B} = 0.99\boldsymbol{b}_{km}$. Then $\hat{\boldsymbol{\beta}}_{k,B}$ is a HB biased estimator of $\boldsymbol{\beta}$ (biased if $\boldsymbol{\beta} \neq \boldsymbol{0}$). The following result shows that it is easy to make a HB estimator that is asymptotically equivalent to any consistent estimator, although the outlier resistance of the HB estimator is poor.

Proposition 5: Let $\hat{\boldsymbol{\beta}}$ be any consistent estimator of $\boldsymbol{\beta}$ and let $\hat{\boldsymbol{\beta}}_H = \hat{\boldsymbol{\beta}}$ if $\text{MED}(r_i^2(\hat{\boldsymbol{\beta}}_H))$ $\leq \text{MED}(r_i^2(\hat{\boldsymbol{\beta}}_{k,B}))$. Let $\hat{\boldsymbol{\beta}}_H = \hat{\boldsymbol{\beta}}_{k,B}$, otherwise. Then $\hat{\boldsymbol{\beta}}_H$ is a HB estimator that is asymptotically equivalent to $\hat{\boldsymbol{\beta}}$.

Proof: Since $\hat{\boldsymbol{\beta}}$ is consistent, $\text{MED}(r_i^2(\hat{\boldsymbol{\beta}})) \to \text{MED}(e^2)$ in probability where $\text{MED}(e^2)$ is the population median of the squared error e^2. Since the LMS estimator is consistent, the probability that $\hat{\boldsymbol{\beta}}$ has a smaller median squared residual than the biased estimator $\hat{\boldsymbol{\beta}}_{k,B}$ goes to 1 as $n \to \infty$. Hence $\hat{\boldsymbol{\beta}}_H$ is asymptotically equivalent to $\hat{\boldsymbol{\beta}}$. QED

Breakdown for MLD

Let $\boldsymbol{W} = \boldsymbol{X}$ and let $\boldsymbol{W}_d^n$ denote the data matrix where any d of the cases have been replaced by arbitrarily bad contaminated cases. The following result shows that a multivariate location estimator T basically "breaks down" if the d outliers can make the median Euclidean distance $\text{MED}(\|\boldsymbol{w}_i - T(\boldsymbol{W}_d^n)\|)$ arbitrarily large where $\boldsymbol{w}_i^T$ is the ith row of $\boldsymbol{W}_d^n$. Thus a multivariate location estimator T will not break down if T can not be driven out of some ball of (possibly huge) radius R about the origin.

Proposition 6. If nonequivariant estimators (that have a breakdown value of greater than 1/2) are excluded, then a multivariate location estimator has a breakdown value of d_T/n iff d_T is the smallest number of arbitrarily bad cases that can make the median Euclidean distance $\text{MED}(\|\boldsymbol{w}_i - T(\boldsymbol{W}_{d_T}^n)\|)$ arbitrarily large.

Proof. Note that for a fixed data set $\boldsymbol{W}_d^n$ with ith row $\boldsymbol{w}_i$, if the multivariate location estimator $T(\boldsymbol{W}_d^n)$ satisfies $\|T(\boldsymbol{W}_d^n)\| = M$ for some constant M, then the median Euclidean distance $\text{MED}(\|\boldsymbol{w}_i - T(\boldsymbol{W}_d^n)\|) \le \max_{i=1,\ldots,n} \|\boldsymbol{x}_i - T(\boldsymbol{W}_d^n)\| \le \max_{i=1,\ldots,n} \|\boldsymbol{x}_i\| + M$ if $d < n/2$. Similarly, if $\text{MED}(\|\boldsymbol{w}_i - T(\boldsymbol{W}_d^n)\|) = M$ for some constant M, then $\|T(\boldsymbol{W}_d^n)\|$ is bounded if $d < n/2$. QED

Since the coordinatewise median $\text{MED}(\boldsymbol{X})$ is a HB estimator of multivariate location, it is also true that a multivariate location estimator T will not break down if T can not be driven out of some ball of radius R about MED$(\boldsymbol{X})$. Hence (MED$(\boldsymbol{X}), \boldsymbol{I}_p$) is a HB estimator of MLD. The proof of the following result shows that if the classical estimator $(\overline{\boldsymbol{x}}_J, \boldsymbol{S}_J)$ is applied to a subset J of $c_n \approx n/2$ cases $\boldsymbol{z}_j$ such that the $\boldsymbol{z}_j$ are not arbitrarily far from the coordinatewise median in Euclidean distance, then $(\overline{\boldsymbol{x}}_J, \boldsymbol{S}_J)$ is a HB estimator of MLD.

Proposition 7 (Olive 2004). Let J consist of the c_n cases $\boldsymbol{x}_i$ such that $\|\boldsymbol{x}_i - \text{MED}(\boldsymbol{X})\| \leq \text{MED}(\|\boldsymbol{x}_i - \text{MED}(\boldsymbol{X})\|)$. Then the classical estimator $(\overline{\boldsymbol{x}}_J, \boldsymbol{S}_J)$ applied to J is a HB estimator of MLD.

Proof. Note that $\overline{\boldsymbol{x}}_J$ is HB by Proposition 6. From numerical linear algebra, it is known that the largest eigenvalue of a $p \times p$ matrix $\boldsymbol{C}$ is bounded above by $p \max |c_{i,j}|$ where $c_{i,j}$ is the (i,j) entry of $\boldsymbol{C}$. See Datta (1995, p. 403). Denote the c_n cases by $\boldsymbol{z}_1, ..., \boldsymbol{z}_{c_n}$. Then the (i,j)th element $c_{i,j}$ of $\boldsymbol{C} \equiv \boldsymbol{S}_J$ is

$$c_{i,j} = \frac{1}{c_n - 1} \sum_{k=1}^{c_n} (z_{i,k} - \overline{z}_k)(z_{j,k} - \overline{z}_j).$$

Hence the maximum eigenvalue λ_1 is bounded if fewer than half of the cases are outliers. Unless the percentage of outliers is high (higher than a value tending to 0.5 as $n \to \infty$), the determinant $|\boldsymbol{C}_{MCD}(c_n)|$ of the HB minimum covariance determinant (MCD) estimator is greater than 0. Thus $0 < |\boldsymbol{C}_{MCD}(c_n)| \leq |\boldsymbol{S}_J| = \lambda_1 \cdots \lambda_p$, and $\lambda_p > |\boldsymbol{C}_{MCD}(c_n)|/\lambda_1^{p-1} > 0$. QED

Additional Comments

Morgenthaler (1989) and Stefanski (1991) conjectured that high breakdown estimators with high efficiency are not possible. These conjectures have been shown to be false by

Proposition 5 and Olive (2007, $\oint$ 8.2 and 10.7).

For many elemental basic resampling algorithms, the number of elemental sets $K_n \equiv K$ does not depend on n. For a fixed data set with small p and an outlier proportion $\gamma < 0.5$, the probability that a clean elemental set is selected will be high if $K_n \equiv K \geq 3(2^d)$ where $d = p$ for MLR and $d = p + 1$ for MLD. Such estimators are sometimes called "high breakdown with high probability," although Proposition 3 shows that the resulting estimator has zero breakdown (asymptotically), *regardless of the MLR criterion.* Many authors, including Maronna and Yohai (2002) and Singh (1998), have mistaken "high breakdown with high probability" for "high breakdown."

There are several important points about breakdown that do not seem to be well known. First, a breakdown result is weaker than even a result such as an estimator being asymptotically unbiased for some population quantity such as $\boldsymbol{\beta}$. This latter property is useful since if the asymptotic distribution of the estimator is a good approximation to the sampling distribution of the estimator, and if many independent samples can be drawn from the population, then the estimator can be computed for each sample and the average of all the different estimators should be close to the population quantity. The breakdown value merely gives a yes or no answer to the question of whether the median absolute residual can be made arbitrarily large when the contamination proportion is equal to γ, and having a bounded median absolute residual does not imply that the high breakdown estimator is asymptotically unbiased or useful.

Secondly, the literature implies that the breakdown value is a measure of the global reliability of the estimator and is a lower bound on the amount of contamination needed to destroy an estimator. These interpretations are not correct since the complement of complete and total failure is *not* global reliability. The breakdown value d_T/n is actually an upper bound on the amount of contamination that the estimator can tolerate since the estimator can be made arbitrarily bad with d_T maliciously placed cases.

In particular, the breakdown value of an estimator tells nothing about more important properties such as consistency or asymptotic normality. Certainly we are reluctant to call an estimator robust if a small proportion of outliers can drive the median absolute residual to ∞, but this type of estimator failure is very simple to prevent.

1 REFERENCES

Datta, B.N. (1995). *Numerical Linear Algebra and Applications,* Brooks/Cole Publishing Company, Pacific Grove, CA.

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann,* (P.J. Bickel, K.A. Doksum and J.L. Hodges, eds.) 157-184. Wadsworth, Pacific Grove, CA.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics,* John Wiley and Sons, Inc., NY.

Hawkins, D.M., and Olive, D.J. (2002). Inconsistency of resampling algorithms for high breakdown regression estimators and a new algorithm. *J. Amer. Statis. Assoc.* (with discussion), 97, 136-159.

Maguluri, G., and Singh, K. (1997). "On the Fundamentals of Data Analysis," in *Robust Inference,* eds. Maddela, G.S., and Rao, C.R., Elsevier Science B.V., Amsterdam, 537-549.

Maronna, R.A., and Yohai, V.J., (2002). Comment on 'Inconsistency of resampling algorithms for high breakdown regression and a new algorithm' by D.M. Hawkins and D.J. Olive. *J. Amer. Statist. Assoc.* **97**, 154-155.

Morgenthaler, S. (1989), "Comment on Yohai and Zamar," *Journal of the American Statistical Association,* 84, 636.

Olive, D.J. (2004). A resistant estimator of multivariate location and dispersion. *Computat. Statist. Data Analys.* 46, 99-102.

Olive, D.J. (2005), "Two Simple Resistant Regression Estimators," *Computational Statistics and Data Analysis*, 49, 809-819.

Olive, D. J. (2007), *Applied Robust Statistics,* unpublished online text available from (www.math.siu.edu/olive).

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

Singh, K. (1998). Breakdown theory for bootstrap quantiles. *Ann. Statis.* **26** 1719-1732.

Stefanski, L.A. (1991), "A Note on High-Breakdown Estimators," *Statistics and Probability Letters,* 11, 353-358.