#1938R

The Library of the Department of Statistics
North Carolina State University

A NOTE ON HIGH-BREAKDOWN ESTIMATORS

by

Leonard A. Stefanski*
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

ABSTRACT

It is shown that regression-equivariant high-breakdown estimators necessarily possess the *exact-fit property* as defined by Yohai and Zamar (1987). Examples are given showing that estimators possessing the exact-fit property can exhibit unusual finite-sample behavior.

Key words: breakdown point, efficiency, equivariance, exact-fit property and robust regression.

* Leonard A. Stefanski is Associate Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203. Research was supported in part by the National Science Foundation.

1. INTRODUCTION

Consider the classical linear regression model $y = \alpha + \beta^T x + \epsilon$ where ϵ is a random error term, β and x are p-dimensional column vectors and α and y are scalars. Fitting this model to a set of data $Z_n = \{(y_i, x_i); i=1,\ldots,n\}$ entails estimating α and β. The fact that least-squares estimation is sensitive to outliers has prompted the search for regression estimators which are resistant to discrepant data. Some estimators proposed to date include Edgeworth's (1887) least-absolute values estimator, Huber's (1973, p. 800) M-estimators, the various generalized M-estimators of Mallows (1975), Hampel (1978), Krasker (1980), and Krasker & Welsch (1982), and most recently the "high-breakdown" estimators of Siegel (1982), Rousseeuw (1984), Rousseeuw & Yohai (1984), Yohai (1985) and Yohai and Zamar (1988). Roughly speaking an estimator has high breakdown if it can "resist" contamination of nearly fifty percent of the data.

In this note it is shown that all regression-equivariant high-breakdown estimators possess the exact-fit property defined by Yohai and Zamar (1988). This implies that all regression-equivariant high-breakdown estimators produce estimates of dubious value for certain data sets of the type exhibited by Oldford (1985) and Yohai and Zamar (1988). More interestingly, it also can be used to show that all high-breakdown regression-equivariant estimators can have arbitrarily low efficiency compared to least-squares in finite samples. Thus while Yohai and Zamar (1988) have shown that high-breakdown and high asymptotic efficiency are not incompatible the story is quite different in finite samples.

The class of high-breakdown estimators considered in this paper are those that are regression equivariant. By this it is meant that the estimators $T_n(Z_n)$ and $T_n(Z_n^*)$ are related via

$$T_n(Z_n^*) = m_{11} \left\{ \begin{array}{c:c} 1 & -C_2^T M_{22}^{-T} \\ \hdashline 0 & M_{22}^{-T} \end{array} \right\} T_n(Z_n) + \left\{ \begin{array}{c} c_1 - C_2^T M_{22}^{-T} M_{12}^T \\ M_{22}^{-T} M_{12}^T \end{array} \right\}$$

whenever $Z_n^* = \{(y_i^*, x_i^*);\ i=1,\dots,n\}$ is obtained from Z_n by a transformation of the form

$$\begin{pmatrix} y_i^* \\ x_i^* \end{pmatrix} = \begin{Bmatrix} m_{11} & M_{12} \\ 0 & M_{22} \end{Bmatrix} \begin{pmatrix} y_i \\ x_i \end{pmatrix} + \begin{pmatrix} c_1 \\ C_2 \end{pmatrix} \qquad (i=1,\dots,n),$$

where m_{11} and c_1 are scalars, M_{12} and C_2 are p-dimensional row and column vectors respectively, M_{22} is a pxp matrix and $m_{11}M_{22}$ is nonsingular.

It is also assumed that if there exist α_n and β_n such that $y_i = \alpha_n + \beta_n^T x_i$ $(i=1,\dots,n)$ and $Z_n = \{(y_i,x_i);\ i=1,\dots,n\}$ then $T_n(Z_n) = (\alpha_n,\beta_n^T)^T$.

The definition of breakdown point used in this paper is due to Donoho and Huber (1983) and employed by Rousseeuw (1984). Let $T_n(Z_n)$ denote an estimator of $(\alpha,\beta^T)^T$ based on the data Z_n and let $Z_{n-m} \oplus W_m$ denote a data set containing n observations, of which n-m are selected from Z_n (and contained in Z_{n-m}) and m observations are selected arbitrarily (and contained in W_m). Define $D(T_n,n,Z_n,m) = \sup |T_n(Z_n) - T_n(Z_{n-m} \oplus W_m)|$ where the supremum is taken over all data sets W_m. Finally, define the breakdown point of T_n at Z_n as $\epsilon^*(T_n,n,Z_n) = \min\{m/n;\ D(T_n,n,Z_n,m) \text{ is infinite}\}$. The breakdown point is the smallest amount of contamination that can cause the estimator to take on values arbitrarily far from $T_n(Z_n)$.

2. REGRESSION EQUIVARIANCE AND BREAKDOWN POINT

Note that the definition of $\epsilon^*(T_n,n,Z_n)$ depends only on the finiteness of $D(T_n,n,Z_n,m)$, thus we could also define the breakdown point as

$\epsilon^*(T_n,n,Z_n) = \min\{m/n;\ I(T_n,n,Z_n,m) = 1\}$ where $I(T_n,n,Z_n,m)$ is zero if $D(T_n,n,Z_n,m)$ is finite and one if not. Now when T_n is regression equivariant, $I(T_n,n,Z_n^*,m) = I(T_n,n,Z_n,m)$ and thus $\epsilon^*(T_n,n,Z_n) = \epsilon^*(T_n,n,Z_n^*)$, i.e., the breakdown point is invariant with respect to the transformation (1).

To see this note that when T_n is equivariant there exists a nonsingular matrix M and a vector V such that $T_n(Z_n^*) = MT_n(Z_n) + V$ and thus

$$\begin{aligned} D(T_n,n,Z_n^*,m) &= \sup_{W_m^*} \left|T_n(Z_n^*) - T_n(Z_{n-m}^* \oplus W_m^*)\right| \\ &= \sup_{W_m^*} \left|MT_n(Z_n) + V - MT_n(Z_{n-m} \oplus W_m) - V\right| \\ &= \sup_{W_m^*} \left|M\{T_n(Z_n) - T_n(Z_{n-m} \oplus W_m)\}\right| , \end{aligned}$$

from which it follows that $D(T_n,n,Z_n^*,m)$ is finite if and only if $D(T_n,n,Z_n,m)$ is finite.

Now suppose that Z_n consists of n coplanar points, i.e., $y_i = \alpha_n + \beta_n^T x_i$ $(i=1,\ldots,n)$ for some α_n and β_n. Thus by assumption $T_n(Z_n) = (\alpha_n,\beta_n^T)^T$ and it follows that under the transformation

$$\begin{pmatrix} y_i^* \\ x_i^* \end{pmatrix} = \begin{pmatrix} 1 & -\beta_n^T \\ 0 & I \end{pmatrix} \begin{pmatrix} y_i \\ x_i \end{pmatrix} + \begin{pmatrix} -\alpha_n \\ 0 \end{pmatrix}$$

that $T_n(Z_n^*) = (0,0^T)^T$, Z_{n-m}^* is contained in the plane y=0 and $T_n(Z_{n-m}^* \oplus W_m^*) = T_n(Z_{n-m} \oplus W_m) - (\alpha_n,\beta_n^T)^T$. By the invariance of $I(T_n,n,Z_n,m)$ we have that $D(T_n,n,Z_n,m)$ is finite if and only if $D(T_n,n,Z_n^*,m)$ is finite. Now consider a second transformation from (y_i^*,x_i^*) to (y_i^{**},x_i^{**}) given by $y_i^{**} = \lambda y_i^*$, $x_i^{**} = x_i^*$. The effect on the estimator is multiplication by λ. Again using the invariance of T_n we have that $D(T_n,n,Z_n^*,m)$ is finite if and only if $D(T_n,n,Z_{n-m}^{**},m)$ is finite and, noting that $Z_{n-m}^* = Z_{n-m}^{**}$ (since $y_i^* \equiv 0$), we have

$$D(T_n,n,Z_{n-m}^{**},m) = \sup_{W_m^{**}} |T_n(Z_{n-m}^{**} \oplus W_m^{**})|$$

$$= \sup_{\lambda, W_m^*} |\lambda| \left| T_n(Z_{n-m}^* \oplus W_m^*) \right| .$$

The last quantity is finite if and only if $T_n(Z_{n-m}^* \oplus W_m^*) = (0,0^T)^T$ for all W_m^*. Thus when Z_n is coplanar we have an alternate interpretation of the breakdown point: $\epsilon^*(n,T_n,Z_n) = m^*(n)/n$ where $m^*(n)$ is the smallest number of observations in Z_n which must be altered to change the estimator $T_n(Z_n)$ to some finite value other than $(\alpha_n,\beta_n^T)^T$. In many cases $\epsilon_n^*(n,T_n,Z_n)$ does not depend on Z_n except through the assumption that $|T_n(Z_n)|$ is unique and finite and thus to determine the breakdown point of T_n we can consider its breakdown point at any set of n coplanar points. In this case if the minimum number of observations which must be altered to produce any change in the estimator is $m^*(n)$, then the breakdown point is $m^*(n)/n$. A similar result was mentioned by Donoho and Rousseeuw at the 1985 Oberwolfach Workshop on Robustness.

This observation allows us to elaborate on a property of high-breakdown estimators first noted by Rousseeuw (1984) and formally defined by Yohai & Zamar (1988). Both authors show that their estimators, which have breakdown points of 1/2 + o(n), have the property that if any subset of more than half the observations is coplanar then the estimator determines that plane. Yohai & Zamar (1986) call this the *exact-fit property*. The preceding analysis shows that among regression-invariant estimators the exact-fit property is a consequence of high breakdown. If a regression invariant estimator has breakdown point $m^*(n)/n$ and any subset of the observations of size $n-m^*(n) + 1$ or larger is coplanar then the estimator is determined by that plane and completely ignores the remaining $m^*(n)-1$ observations. This extends the notion of exact-fit to estimators with breakdown points less than 1/2 and will be called the exact-fit property or order $m^*(n)/n$.

3. BREAKING HIGH-BREAKDOWN ESTIMATORS

The results in Section 2 suggest how to construct data sets in which high-breakdown estimators produce suspect results. If an entire data set is well fit by one model while a subset of the data of size $k > n\{1 - \epsilon^*(T_n,n,Z_n)\}$ is well fit by another model then it is possible that a high breakdown estimator will model the local as opposed to the global trend in the data. One such example can be found in Yohai and Zumar (1988).

See Figure 1 for a second example. Figure 1 illustrates a data set exhibiting a trend corresponding to the line $y \approx 0$ plus two wild outliers and is similar to a scatter plot presented by Oldford (1983). By appropriate choice of position we can force a high-breakdown estimator to yield the model $y=x$ although a safe bet is that most statisticians would be more comfortable with a procedure which produced the model $y \approx 0$ and identified two outliers.

Dennis Boos of North Carolina State University suggested the following method of constructing examples for multiple linear regression models. For example, for $p=2$ let A, B and C be three arbitrary but distinct points in $\mathbb{R}^3$. They uniquely determine a plane, call it ABC. Now select any four additional distinct points, D, E, F and G, which are coplanar and which deviate from ABC by an arbitrarily small amount. The seven points A-G can be modeled arbitrarily well by the plane ABC, yet since more than half the points are coplanar a high-breakdown, regression-equivariant estimator will ignore the points A, B and C and fit the plane determined by D-G, which can be made to differ substantially from ABC. In a similar manner it is possible to construct a counter example of size n=2p+3 when x is p-dimensional.

Of course there is a natural and forceful counter-argument to these counterexamples: such behavior occurs with negligible frequency in practice. However the examples used to illustrate the benefits of high-breakdown estimators over other robust estimators also occur in practice with nearly equally small frequency. Percent contamination rarely exceeds 15%-20% and is typically less than 10%. In light of this fact 50%-breakdown estimators seem like overkill. However the poor breakdown properties of M-estimators make them less than fully acceptable. Perhaps emphasis should be on finding estimators possessing moderately good breakdown properties (10%-25%) as well as other desirable properties, such as bounded influence. Whether or not this is possible is unknown at present.

It is now shown that the exact-fit property also implies arbitrarily low efficiency with respect to least squares in finite samples. A conjecture of this fact has appeared in print since the first submission of this paper, Morgenthaler (1989).

Consider a simple linear regression model $Y_i = \alpha + \beta X_i + \epsilon_i$, $i=1,\ldots,n$. As before let $Z_n = \{(Y_i,X_i);\ i=1,\ldots,n\}$ and let $T_n(\cdot)$ be any estimator possessing the exact-fit property. Let $W_n(\cdot)$ denote the least squares estimator. Let $\beta_T(Z_n)$ and $\beta_W(Z_n)$ denote the slope estimators determined by $T_n(Z_n)$ and $W_n(Z_n)$ respectively. Below it is shown that $T_n(\cdot)$ can have arbitrarily low efficiency with respect to $W_n(\cdot)$ in finite samples.

For simplicity only the case n=4 is considered. Also, let $\alpha = \beta = 0$ without loss of generality. Then $Y_i = \epsilon_i$, i=1,...,4. Let $x_i = 0$, $x_2 = 1$, $x_3 = 2$ and $x_4 = x$, to be determined later.

Assume initially that the common distribution of $\epsilon_1,\ldots,\epsilon_4$ is supported on {-1,0,1}, where each of the three support points has positive mass. For this error distribution it follows readily that $-2/x \leq \beta_W(Z_n) \leq 2/x$. Thus $\sigma_W^2 = E\{\beta_W^2(Z_n)\}$ can be made arbitrarily small by taking x sufficiently large.

Since $T_n(\cdot)$ has the exact-fit property it is easy to see that $pr\{\beta_T(Z_n) = 1\} \geq pr(\epsilon_1 = -1,\ \epsilon_2 = 0,\ \epsilon_3 = 1) = p^* > 0$ where p^* does not depend on x. Thus no matter what the value of x, $\beta_T(Z_n)$ is bounded away from zero in probability. Thus so too is $\sigma_T^2 = E\{\beta_T^2(Z_n)\}$. It follows that σ_W^2/σ_T^2 can be made arbitrarily small by choosing x large enough.

Now assume that $T_n(\cdot)$ is continuous at the point Z^* defined below with respect to the norm $\|\cdot\|$, given by $\|M\| = \Sigma|m_{ij}|$ when M is a rectangular matrix. Let $Z^* = \{(a,0),\ (a+h,1),\ (a+2h,2),\ (\epsilon_4,x)\}$ and $Z = \{(\epsilon_1,0)\ (\epsilon_2,1)\ (\epsilon_3,2)\ (\epsilon_4,x)\}$, where a and $h \neq 0$ are fixed constants. Let $0 < \epsilon < |h|/2$ be given. By continuity of $T_n(\cdot)$, $|\beta_T(Z) - \beta_T(Z^*)| < \epsilon$ whenever $\|Z - Z^*\| < \delta$. By the exact fit property $\beta_T(Z^*) = h$. Thus it follows that if $\|Z - Z^*\| = |\epsilon_1 - a| + |\epsilon_2 - a - h| + |\epsilon_3 - a - 2h| < \delta$, then $|\beta_T(Z) - h| < \epsilon$, which in turn implies that $|\beta_T(Z)| > |h|/2 > 0$. Thus

$$pr\{|\beta_T(Z)| > |h|/2\} > pr\{|\epsilon_1 - a| + |\epsilon_2 - a - h| + |\epsilon_3 - a - 2h| < \delta\} .$$

The right hand side above does not depend on x. Thus whenever the error distribution is such that $(\epsilon_1,\epsilon_2,\epsilon_3)$ lies near some nondegenerate arithmetic progression with nonzero probability it follows that $\beta_T(Z)$ is bounded away from zero in probability uniformly in x and thus so too is $\sigma_T^2 = E\{\beta_T^2(Z)\}$. Provided $E(\epsilon_i^2) < \infty$, it still follows that $\sigma_W^2 = E\{\beta_W^2(Z)\}$ tends to zero as x increases. Thus again σ_W^2/σ_T^2 can be made arbitrarily small by letting x increase without bound.

Any continuous distribution will put positive probability near some nondegenerate arithmetic progression and the argument above applies quite generally. Furthermore, it extends easily to arbitrary n. Finally note that the argument works for any estimator possessing the exact-fit property of any order exceeding 1/n. Thus it can be stated quite generally that any estimator possessing the exact fit property of any order > 1/n has arbitrarily low efficiency with respect to least-squares in finite samples.

I conducted a small simulation study to investigate the inefficiency of a particular high-breakdown estimator, the least trimmed squares, LTS, estimator defined in Rousseeuw (1984). Since computation of high breakdown estimators is time consuming in large samples and since the phenomenon under study is inherently small-sample in nature, only samples of size n=9 were considered. In this case the LTS estimator is the regression line obtained by fitting least squares lines to each of the $\binom{9}{5}$ subsets of $\{Y_i, X_i \;\; i=1,\ldots,9\}$ and choosing the line having the best fit in terms of residual variation; see Rousseeuw (1984) for further explanation.

Two experiments were performed. In the first $X_i = i$, $i=1,\ldots,8$, $X_9 = 9$, 12, 24 and $Y_1, \ldots, Y_n$ were generated from a standard linear model with normal errors. The relative efficiencies for estimating slope reported in Table 1(a) are based on 7500 simulated data sets. As X_9 increases the superiority of least squares becomes more noticeable. However, even with the equally-spaced design, least-squares is approximately four times more efficient than LTS.

In the second experiment, $X_i = i$, $i=1,\ldots,9$, and $Y_1,\ldots, Y_n$ were generated from a standard linear model with errors having a contaminated normal distribution. The fraction of contamination was fixed at .30, and kurtosis was set at 0, 3, 6 respectively. The relative efficiencies for estimating slope in Table 1(b) are based on 7500 simulated data sets. As kurtosis increases efficiency of the LTS estimator improves. The break-even point is approximately at a kurtosis of 5.6, as suggested by linear interpolation.

REFERENCES

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

Edgeworth, F. Y. (1887), "On Observations Relating to Several Quantities," *Hermathena*, 6, 279-285.

Hampel, F. R. (1978), "Optimally Bounding the Gross-Error-Sensitivity and the Influence of Position in Factor Space," *Proceedings of the Statistical Computing Section, American Statistical Association*, 59-64.

Huber, P. J. (1973), "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics*, 1, 799-821.

Krasker, W. S. (1980), "Estimation in Linear Regression Models With Disparate Data Points," *Econometrica*, 48, 1333-1346.

Krasker, W. S. and Welsch, R. E. (1982), "Efficient Bounded-Influence Regression Estimation," *Journal of the American Statistical Association*, 77, 595-604.

Mallows, C. L. (1975), "On Some Topics in Robustness," unpublished memorandum, Bell Telephone Laboratories, Murray Hill, NJ.

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

Oldford, R. W. (1983), "A Note on High Breakdown Regression Estimators," Technical Report, Massachusetts Institute of Technology, Center for Computational Research in Economics and Management Science.

Rousseeuw, P. J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw, P. J. and Yohai, V. J. (1984), "Robust Regression by Means of S-Estimators," in *Robust and Nonlinear Time Series*, Franke, Handle and Martin, eds., *Lecture Notes in Statistics*, No. 26, Springer-Verlag, New York.

Siegel, A. F. (1982), "Robust Regression Using Repeated Medians," *Biometrika*, 69, 242-244.

Yohai, V. J. (1985), "High Breakdown-Point and High Efficiency Robust Estimates for Regression," Technical Report No. 66, Department of Statistics, University of Washington, Seattle, WA.

Yohai, V. J. and Zamar, R. (1987), "High Breakdown-Point Estimates of Regression by Means of the Minimization of an Efficient Scale," *Journal of the American Statistical Association*, 83, 406-413.

Legend for Figure 1.

Figure 1. A data set for which high-breakdown estimators produce suspect estimates; trend induced by outliers is modelled and not the global trend $y \approx 0$.

Figure 1

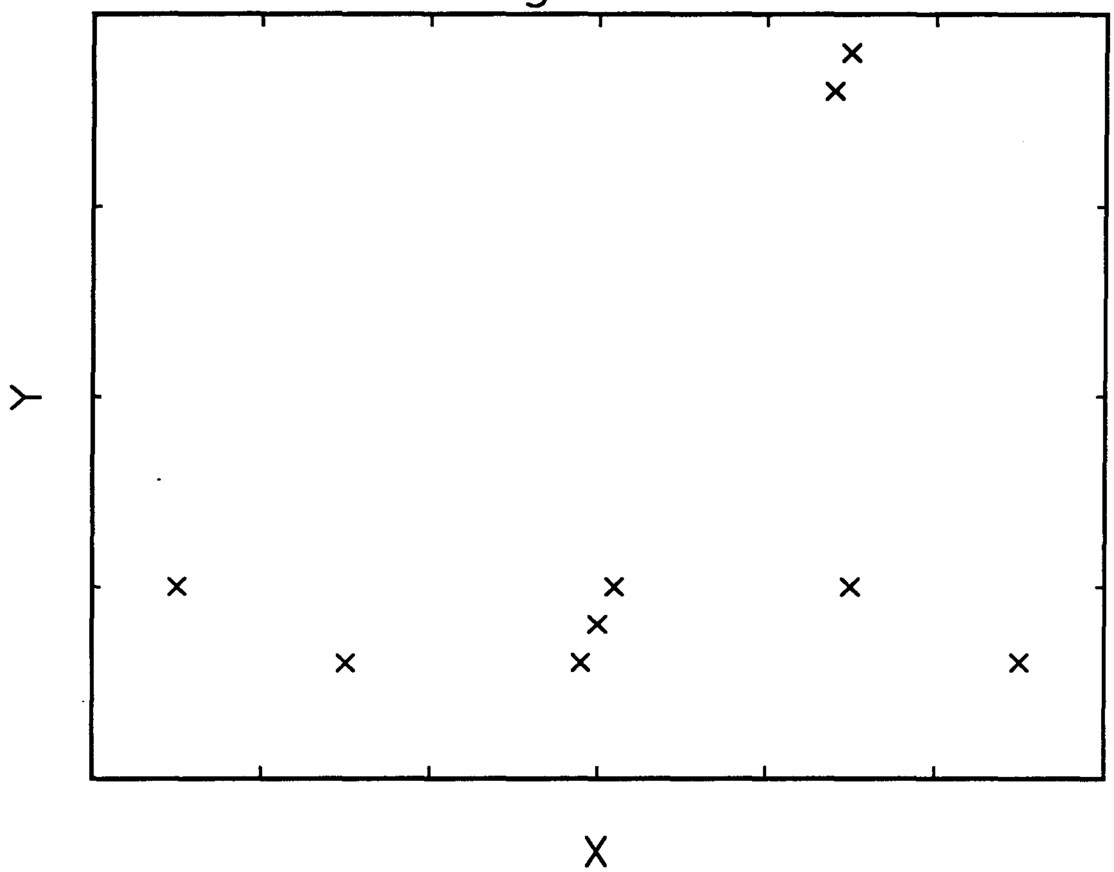

Table 1. Efficiency of least squares to least trimmed squares, LTS, slope estimators.

(a)	X_9 =	9	12	24
	*Relative Efficiency =	3.8	4.6	15.6
		(0.1)	(0.2)	(0.3)
(b)	Kurtosis =	0	3	6
	*Relative Efficiency =	3.9	2.5	0.8
		(0.1)	(0.1)	(0.1)

* Relative Efficiency estimated as $\{s.d.(\hat{\beta}_{LTS})/s.d.(\hat{\beta}_{LS})\}^2$ based on 7500 replications; approximate standard errors given in parentheses.