High Breakdown Multivariate Estimators

David J. Olive

Southern Illinois University, Carbondale, IL 62901-4408

and

Douglas M. Hawkins

University of Minnesota, Minneapolis, MN 55455-0493.

January 9, 2008

Abstract

In the literature, estimators for regression or multivariate location and dispersion that have been shown to be both consistent and high breakdown are impractical to compute. This paper gives easily computed high breakdown robust $\sqrt{n}$ consistent estimators, and the applications for these estimators are numerous.

For regression, the response plot of the fitted values versus the response is shown to be an effective tool for detecting outliers and influential cases.

KEY WORDS: minimum covariance determinant estimator, multivariate location and dispersion, outliers, robust regression.

1. INTRODUCTION

The *multiple linear regression (MLR) model* is $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ 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 $\boldsymbol{\Sigma}$. The observations $\boldsymbol{x}_i$ for $i = 1, ..., n$ are collected in an $n \times p$ matrix $\boldsymbol{X}$ with n rows $\boldsymbol{x}_1^T, ..., \boldsymbol{x}_n^T$. An important MLD model is the elliptically contoured $EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ distribution with probability density function

$$f(\boldsymbol{z}) = k_p |\boldsymbol{\Sigma}|^{-1/2} g[(\boldsymbol{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{\mu})]$$

where $k_p > 0$ is some constant and g is some known function. See Johnson (1987, pp. 107-108). The multivariate normal (MVN) distribution is a special case, and $\boldsymbol{x}$ is "spherical about $\boldsymbol{\mu}$" if $\boldsymbol{x}$ has an $EC_p(\boldsymbol{\mu}, c\boldsymbol{I}_p, g)$ distribution where $c > 0$ is some constant and $\boldsymbol{I}_p$ is the $p \times p$ identity matrix.

Many of the most used estimators in statistics are semiparametric. The least squares (OLS) estimator

$$\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y} \tag{1}$$

is a semiparametric MLR estimator. If the e_i are iid with mean 0 and variance σ^2, then there is a central limit type theorem for OLS. For multivariate analysis, the classical estimator $(\overline{\boldsymbol{x}}, \boldsymbol{S})$ is the sample mean and sample covariance matrix where

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \quad \text{and} \quad \boldsymbol{S} = \frac{1}{\mathrm{n}-1}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\boldsymbol{x}_{\mathrm{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\mathrm{i}} - \overline{\boldsymbol{x}})^{\mathrm{T}}. \tag{2}$$

Many classical procedures originally meant for the MVN distribution are semiparametric in that the procedures also perform well on a much larger class of EC distributions.

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 $\boldsymbol{C}(\boldsymbol{X})$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

$$D_i^2 = D_i^2(T(\boldsymbol{X}), \boldsymbol{C}(\boldsymbol{X})) = (\boldsymbol{x}_i - T(\boldsymbol{X}))^T \boldsymbol{C}^{-1}(\boldsymbol{X})(\boldsymbol{x}_i - T(\boldsymbol{X})) \tag{3}$$

for each observation $\boldsymbol{x}_i$. Notice that the Euclidean distance of $\boldsymbol{x}_i$ from the estimate of center $T(\boldsymbol{X})$ is $D_i(T(\boldsymbol{X}), \boldsymbol{I}_p)$. The classical Mahalanobis distance uses $(T, \boldsymbol{C}) = (\overline{\boldsymbol{x}}, \boldsymbol{S})$.

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 $0 \leq \lambda_p(\boldsymbol{C}) \leq \cdots \leq \lambda_1(\boldsymbol{C})$ denote the eigenvalues of the dispersion estimator $\boldsymbol{C}$. The breakdown value of $\boldsymbol{C}$ is the smallest value of γ needed to drive the smallest eigenvalue to zero or the largest eigenvalue to ∞.

High breakdown (HB) statistics have $\gamma \rightarrow 0.5$ as $n \rightarrow \infty$, and an important goal of high breakdown robust statistics is to produce easily computed semiparametric MLR and MLD estimators that perform well when the classical estimators perform well, but are also useful for detecting some important types of outliers: cases that lie far away from the bulk of the data.

The published literature for HB MLR or MLD estimators contains one or more major flaws: either i) the estimator is impractical to compute or ii) the estimator is practical to compute but has not been shown to be both high breakdown and consistent.

Computational complexity is discussed in Bernholt (2005, 2006) and Bernholt and Fischer (2004). Of the MLD and MLR estimators that have been shown to be both consistent and HB, perhaps the fastest is the least median of squares (LMS) estimator that has complexity $O(n^p)$.

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.

Many criteria for screening the attractors have been suggested. See Rousseeuw (1984) for the following four criteria. For MLR the LMS criterion is the median squared residual, or more generally, the LMS(c_n) criterion is $Q_{LMS}(\boldsymbol{b}) = r^2_{(c_n)}(\boldsymbol{b})$ where $r^2_{(1)} \leq \cdots \leq r^2_{(n)}$ are the ordered squared residuals. The least trimmed sum of squares LTS(c_n) criterion is $Q_{LTS}(\boldsymbol{b}) = \sum_{i=1}^{c_n} r^2_{(i)}(\boldsymbol{b})$. For MLD, the attractor is the classical estimator $(\overline{\boldsymbol{x}}_{k,j}, \boldsymbol{S}_{k,j})$ computed from a subset of c_n cases. The minimum covariance determinant MCD(c_n) criterion is the determinant $\det(\boldsymbol{S}_{k,j})$. The volume of the hyperellipsoid

$$\{\boldsymbol{z} : (\boldsymbol{z} - \overline{\boldsymbol{x}}_{k,j})^T \boldsymbol{S}_{k,j}^{-1} (\boldsymbol{z} - \overline{\boldsymbol{x}}_{k,j}) \leq h^2\} \text{ is equal to } \frac{2\pi^{\mathrm{p}/2}}{\mathrm{p}\Gamma(\mathrm{p}/2)} \mathrm{h}^{\mathrm{p}} \sqrt{\det(\boldsymbol{S}_{\mathrm{k,j}})}, \tag{4}$$

see Johnson and Wichern (1988, pp. 103-104). The minimum volume ellipsoid MVE(c_n) criterion is $h^p \sqrt{\det(\boldsymbol{S}_{\mathrm{k,j}})}$ where $h = D_{(c_n)}(\overline{\boldsymbol{x}}_{k,j}, \boldsymbol{S}_{k,j})$.

Hawkins and Olive (2002) showed that if K randomly selected elemental starts are used and concentration is used to produce the attractors, then the best attractor is not consistent if K and k are fixed and free of n. Hence no matter how the attractor is chosen, the resulting estimator is not consistent. The proof is simple given the results of He and Portnoy (1992) and Lopuhaä (1999) who show that if a start $\boldsymbol{b}$ or $(T, \boldsymbol{C})$ is a consistent estimator of $\boldsymbol{\beta}$ or $(\boldsymbol{\mu}, s\boldsymbol{\Sigma})$, then the attractor is a consistent estimator of $\boldsymbol{\beta}$ or $(\boldsymbol{\mu}, a\boldsymbol{\Sigma})$ where $a, s > 0$ are some constants. Also the constant a does not depend on s and the attractor and the start have the same rate. If the start is inconsistent, then so is the attractor. The classical estimator applied to a randomly drawn elemental set is an inconsistent estimator, so the K starts and the K attractors are inconsistent. The final estimator is an attractor and thus inconsistent.

If concentration is iterated to convergence so that k is not fixed, then it has not been proven that the attractor is inconsistent if elemental starts are used. It is possible to produce consistent estimators if K_n is allowed to increase to ∞, but for MLR the rate of the algorithm is bounded above by $K_n^{1/p}$.

This theory has been largely ignored. For example, Maronna, Martin and Yohai (2006, pp. 198-199) use $K = 500$ and $k = 1$ to create MLD estimators and state that no theoretical results for their inconsistent method are known. Hubert, Rousseeuw and Van Aelst (2007) have the following quote.

It turns out that most of the currently available highly robust multivariate estimators are difficult to compute, which makes them unsuitable for the analysis of large and/or high-dimensional datasets. Among the few exceptions is the minimum covariance determinant estimator (MCD) of Rousseeuw (1984, 1985). The MCD is a highly robust estimator of multivariate location and scatter that can be computed efficiently with the FAST-MCD algorithm of Rousseeuw and Van Driessen (1999).

FAST-MCD is a concentration estimator that uses 500 elemental starts. Since 5 starts are iterated until convergence, the estimator has not been proven to be inconsistent, but also has not been shown to be consistent. FAST-MCD and MCD are completely different estimators. This technical error is very common in the literature: the theory is known or derived for a robust estimator that is impractical to compute, so an algorithm is used to compute a completely different estimator. Only the theory of the algorithm estimator actually used is of interest, but this theory is not given.

Although Maronna and Zamar (2002, p. 309) claim that their orthogonalized

Gnanadesikan-Kettenring (OGK) estimator is consistent and HB, they fail to provide the proofs. The OGK estimator is faster than FAST-MCD and is not computed with an elemental concentration algorithm.

Sections 2 and 3 modify the concentration algorithms of Hawkins and Olive (1999) resulting in easily computed HB $\sqrt{n}$ consistent estimators. Section 4 compares FAST-MCD and OGK with the new multivariate estimators in a small simulation study.

2. PRACTICAL HB REGRESSION AND MLR OUTLIER DETECTION

Olive (2005) showed that an MLR estimator is high breakdown if the median absolute residual stays bounded under high contamination. (Notice that if $\|\hat{\boldsymbol{\beta}}\| = \infty$, then $\text{MED}(|r_i|) = \infty$, and if $\|\hat{\boldsymbol{\beta}}\| = M$ then $\text{MED}(|r_i|)$ is bounded if fewer than half of the cases are outliers.) For MLR, breakdown is more of a Y-outlier property than an $\boldsymbol{x}$-outlier property. If the Y_i's are fixed, arbitrarily large $\boldsymbol{x}$-outliers tend to drive the slope estimates to 0, not ∞. If the LTS criterion is used, concentration insures that the criterion function of the $c_n \approx n/2$ absolute residuals gets smaller. Hence LTS concentration algorithms that use a HB start are HB.

Assume the MLR model contains a constant β_1. To make a HB MLR estimator with good properties, simply use OLS as an attractor and the following easily computed HB inconsistent attractor. Let $\boldsymbol{b}_k$ be the attractor from the start consisting of OLS applied to the c_n cases with Y's closest to the median of the Y_i and let $\hat{\boldsymbol{\beta}}_{k,B} = 0.99\boldsymbol{b}_k$. Then $\hat{\boldsymbol{\beta}}_{k,B}$ is a HB biased estimator of $\boldsymbol{\beta}$ (biased if $\boldsymbol{\beta} \neq \boldsymbol{0}$, see Olive 2005). The notation CLTS will be used if the LTS criterion is used in the concentration algorithm.

Theorem 1 shows that the HB CLTS and LTS(0.5) estimators are completely different. CLTS is simple to compute and has 100% Gaussian efficiency while LTS(0.5) is impractical to compute and has 7.1% Gaussian efficiency. The proof of the theorem is simple since He and Portnoy (1992) showed that if the start is a consistent estimator of $\boldsymbol{\beta}$ with rate n^{δ}, so is the attractor.

Theorem 1. Suppose that the CLTS algorithm uses K randomly selected elemental starts (e.g., K = 500) and the attractors $\hat{\boldsymbol{\beta}}_{OLS}$ and $\hat{\boldsymbol{\beta}}_{k,B}$. Then the resulting estimator is a HB $\sqrt{n}$ consistent estimator that is asymptotically equivalent to $\hat{\boldsymbol{\beta}}_{OLS}$.

Proof. The CLTS estimator is HB by the remarks above. The LTS estimator is consistent by Mašīček (2004) or Čížek (2006). As $n \to \infty$, consistent estimators $\hat{\boldsymbol{\beta}}$ satisfy $Q_{LTS}(\hat{\boldsymbol{\beta}})/n - Q_{LTS}(\boldsymbol{\beta})/n \to 0$ in probability. Since $\hat{\boldsymbol{\beta}}_{k,B}$ is a biased estimator of $\boldsymbol{\beta}$, with probability tending to one, OLS will have a smaller criterion value. With probability tending to one, OLS will also have a smaller criterion value than the criterion value of the attractor from a randomly drawn elemental set (by He and Portnoy 1992, also see Remark 4 in Hawkins and Olive 2002). Since K random elemental sets are used, the CLTS estimator is asymptotically equivalent to OLS. □

The outlier resistance of CLTS is not much better than that of other LTS concentration algorithms. Since for MLR the HB property is a Y-outlier property, $\boldsymbol{x}$-outliers may have small residuals even if HB estimators are used. After examining many "benchmark" data sets, we found that Cook's distances CD_i from Cook (1977) tend to be larger for influential cases than the Peña (2005) S_i statistics. These influence diagnostics are also ineffective in the presence of $\boldsymbol{x}$-outliers.

For detection of outliers and influential cases, it is crucial to make the residual plot of

$\hat{Y}$ vs r and the response plot of $\hat{Y} = \boldsymbol{x}^T \hat{\boldsymbol{\beta}}$ vs Y with the identity line with zero intercept and unit slope added as a visual aid. Vertical deviations from the identity line are the residuals $r_i = Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$.

Olive and Hawkins (2005) also showed that the two plots are crucial for visualizing the MLR model and for examining lack of fit. If $n > 10p$ and if the plotted points scatter about the identity line and the $r = 0$ line in an evenly populated band, then the MLR model with iid e_i where VAR$(e_i) = \sigma^2$ may be reasonable. Deviations from the evenly populated band suggest that something is wrong with the MLR model, and are often easily detected even if OLS is used. The following two examples help illustrate the above remarks.

Example 1. Gladstone (1905-6) records the brain weight and various head measurements for 267 individuals. Consider predicting *brain weight* using six head measurements (head *height, length, breadth, (size)*$^{1/3}$, *cephalic index* and *circumference*) as predictors. There are five infants (cases 230, and 254-257) of age less than 7 months that are $\boldsymbol{x}$-outliers. Nine toddlers were between 7 months and 3.5 years of age, three of whom appear to be $\boldsymbol{x}$-outliers (cases 232, 258, and 260). Figures 1 and 2 show the OLS response and residual plots. For this example the $\boldsymbol{x}$-outliers are "good leverage points" in that a response plot using the OLS fit without the eight $\boldsymbol{x}$-outliers has the leverage points scatter about the identity line. Cases 118 and 234 had the largest Cook's distances. Cook's distances are ineffective because the residuals and classical Mahalanobis distances are not large for the cluster of infants. Influence diagnostics are the most effective when there is a single cluster about the identity line.

Suppose the iid error MLR model with constant variance is appropriate for this data. Then MSE = $\hat{\sigma}^2$ = 6381.182. To use the response plot to visualize the MLR model, suppose the brain weight given fit = 1200 is of interest. Mentally examine the response plot for a narrow vertical strip about fit = 1200, perhaps from 1175 to 1225. The cases in the narrow strip have mean near 1200 since they fall close to the identity line. Similarly, when then fit = w for w between 400 and 1500, the cases have brain weights near w on average. If the errors are iid $N(0, \sigma^2)$, then $Y|\boldsymbol{x}^T\hat{\boldsymbol{\beta}} \approx N(\boldsymbol{x}^T\hat{\boldsymbol{\beta}}, 6381.182)$.

Example 2. Buxton (1920, p. 232-5) gives 20 measurements of 88 men. Consider predicting *stature* using an intercept, *head length, nasal height, bigonal breadth*, and *cephalic index.* One case was deleted since it had missing values. Five individuals, numbers 61-65, were reported to be about 0.75 inches tall with head lengths well over five feet! Figures 3 and 4 show the OLS response plot and residual plot for the Buxton data. Although an index plot of Cook's distance CD_i may be useful for flagging influential cases, the index plot provides no direct way of judging the model against the data. As a remedy, cases in the plots with $CD_i > \min(0.5, 2p/n)$ were highlighted. Notice that the OLS fit passes through the outliers, but the response plot is resistant to Y–outliers since Y is on the vertical axis. Also notice that although the outlying cluster is far from $\overline{Y}$ only two of the outliers had large Cook's distance. Hence masking occurred for both Cook's distances and for OLS residuals, but not for OLS fitted values.

3. PRACTICAL HB MLD ESTIMATORS

There are many applications for easily computed HB consistent estimators of multivariate location and dispersion. In addition to outlier detection, the robust estimator can be plugged in for the classical estimator to produce robust estimators for multivariate procedures. Hubert, Rousseeuw and Van Aelst (2007) consider many methods including *discriminant analysis*, *principal component regression* and *partial least squares.* Also see, for example, Croux and Haesbroeck (2003) for *binary regression*; Branco, Croux, Filzmoser, and Oliviera (2005) for *canonical correlations*; Pison, Rousseeuw, Filzmoser, and Croux (2003) for *factor analysis*; He, Fung and Zhu (2005) for *generalized partial linear models*; Willems, Pison, Rousseeuw, and Van Aelst (2002) for *analogs of Hotelling's* T^2 *test*; Rousseeuw, Van Aelst, Van Driessen and Agulló (2004) for *multivariate regression*; Hubert, Rousseeuw, and Vanden Branden (2005) and Maronna (2005) for *principal components.*

Unfortunately, computation and theory for HB estimators has not kept up with the applications. The above "robust methods" typically use a "robust estimator" which is impractical to compute or which has not been shown to be both HB and consistent. Hence the procedures are outlier diagnostics rather than HB robust methods. If $n > 20p$ and $p \leq 40$, the easily computed HB $\sqrt{n}$ consistent estimators described below can be used in place of the classical estimator to produce HB robust methods.

Suppose K is a fixed positive integer and there are K consistent estimators $(T_j, \boldsymbol{C}_j)$ of $(\boldsymbol{\mu}, a\ \boldsymbol{\Sigma})$ for some constant $a > 0$, each with the same rate n^δ. If $(T_A, \boldsymbol{C}_A)$ is an estimator obtained by choosing one of the K estimators, then $(T_A, \boldsymbol{C}_A)$ is a consistent estimator of $(\boldsymbol{\mu}, a\ \boldsymbol{\Sigma})$ with rate n^δ by Pratt (1959). Theorem 4 below and the following lemma show Lopuhaä (1999) can be used with Pratt (1959) with $a = a_{MCD}$ to provide simple proofs for MLD concentration algorithms.

Assumption (E1): Assume that $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid $EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ with nonsingular $\text{Cov}(\boldsymbol{x}_i)$

$= c_X \boldsymbol{\Sigma}$ for some constant $c_X > 0$.

Lemma 2. Assume that (E1) holds and that $(T, \boldsymbol{C})$ is a $\sqrt{n}$ consistent estimator of $(\boldsymbol{\mu}, s\boldsymbol{\Sigma})$ where the constant $s > 0$. Then the classical estimator $(\overline{\boldsymbol{x}}_{m,j}, \boldsymbol{S}_{m,j})$ computed with the $c_n \approx n/2$ of cases with the smallest distances $D_i(T, \boldsymbol{C})$ is a $\sqrt{n}$ consistent estimator of $(\boldsymbol{\mu}, a_{MCD}\boldsymbol{\Sigma})$.

Proof. The result follows by Lopuhaä (1999) if $a = a_{MCD}$. But the MCD estimator is a $\sqrt{n}$ consistent estimator of $(\boldsymbol{\mu}, a_{MCD}\boldsymbol{\Sigma})$ by Butler, Davies and Jhun (1993). If the MCD estimator is the start, then it is also the attractor by Rousseeuw and Van Driessen (1999) who show that concentration does not increase the MCD criterion. Hence $a_{MCD}\boldsymbol{\Sigma} = a\boldsymbol{\Sigma}$. $\square$

If the MLD estimator $(T, \boldsymbol{C})$ tracks the data, then T will not break down if T can not be driven out of some ball of radius R about the origin. To see this, let $\boldsymbol{W}_d^n$ denote the data matrix with ith row $\boldsymbol{w}_i^T$ where any d of the cases have been replaced by arbitrarily bad contaminated cases. So $\boldsymbol{X} = \boldsymbol{W}_0^n$. If T satisfies $\|T(\boldsymbol{W}_d^n)\| = M$ for some constant M, then the median Euclidean distance $\text{MED}(\|\boldsymbol{w}_i - T(\boldsymbol{W}_d^n)\|) \leq \max_{i=1,\ldots,n} \|\boldsymbol{x}_i - T(\boldsymbol{W}_d^n)\|$ $\leq \max_{i=1,\ldots,n} \|\boldsymbol{x}_i\| + M$ 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 $d < n/2$. Since the coordinatewise median $\text{MED}(\boldsymbol{X})$ is a HB estimator, the origin can be replaced by $\text{MED}(\boldsymbol{X})$.

Recall that the sample median $\text{MED}(Y_i) = Y((n+1)/2)$ is the middle order statistic if n is odd. Thus if $n = m + d$ where m is the number of clean cases and $d = m - 1$ is the number of outliers so $\gamma \approx 0.5$, then the sample median can be driven to the max or min of the clean cases. The jth element of $\text{MED}(\boldsymbol{X})$ is the sample median of the jth predictor.

Hence with $m-1$ outliers, MED($\boldsymbol{X}$) can be driven to the "coordinatewise covering box" of the m clean cases. The boundaries of this box are at the min and max of the clean cases from each predictor, and the lengths of the box edges equal the ranges R_i of the clean cases for the ith variable. If $d \approx m/2$ so that $\gamma \approx 1/3$, then the MED($\boldsymbol{X}$) can be moved to the boundary of the much smaller "coordinatewise IQR box" corresponding the 25th and 75th percentiles of the clean date. Then the edge lengths are approximately equal to the interquartile ranges of the clean cases.

Note that $D_i(\text{MED}(\boldsymbol{X}), \boldsymbol{I}_p) = \|\boldsymbol{x}_i - \text{MED}(\boldsymbol{X})\|$ is the Euclidean distance of $\boldsymbol{x}_i$ from MED($\boldsymbol{X}$). Let $\mathcal{C}$ denote the set of m clean cases. If $d \leq m-1$, then the minimum distance of the outliers is larger than the maximum distance of the clean cases if the distances for the outliers satisfy $D_i > B$ where

$$B^2 = \max_{i \in \mathcal{C}} \|\boldsymbol{x}_i - \text{MED}(\boldsymbol{X})\|^2 \leq \sum_{i=1}^p R_i^2 \leq p(\max R_i^2).$$

Next we define three easily computed HB $\sqrt{n}$ consistent MLD estimators and then give the corresponding theory. The new CMVE and FCH estimators have greater outlier resistance than the Olive (2004) median ball algorithm (MBA) estimator because location information from T is used as well as dispersion information from $\boldsymbol{C}$. The CMVE estimator uses concentration to produce two attractors, but makes use of the MVE criterion to choose the final attractor. The FCH estimator is so named because it is fast, consistent and HB.

The CMVE, MBA and FCH estimators use the same two attractors. The first attractor is the Devlin, Gnanadesikan and Kettenring (1981) DGK estimator that uses the classical estimator as the start. The second attractor is the median ball (MB) estimator that uses the classical estimator computed from the cases with $D_i(\text{MED}(\boldsymbol{X}), \boldsymbol{I}_p) \leq \text{MED}(D_i(\text{MED}(\boldsymbol{X}), \boldsymbol{I}_p))$ as a start. Thus the start $(T_{0,M}, \boldsymbol{C}_{0,M}) = (\overline{\boldsymbol{x}}_{0,M}, \boldsymbol{S}_{0,M})$ is the classical estimator applied after trimming $M\%$ of the cases furthest in Euclidean distance from $\text{MED}(\boldsymbol{X})$ for $M \in \{0, 50\}$. The Mth attractor is $(T_{k,M}, \boldsymbol{C}_{k,M}) = (\overline{\boldsymbol{x}}_{k,M}, \boldsymbol{S}_{k,M})$. Let $(\overline{\boldsymbol{x}}_{-1,50}, \boldsymbol{S}_{-1,50}) = (\text{MED}(\boldsymbol{X}), \boldsymbol{I}_p)$. Then the median ball estimator $(\overline{\boldsymbol{x}}_{k,50}, \boldsymbol{S}_{k,50})$ is also the attractor of $(\text{MED}(\boldsymbol{X}), \boldsymbol{I}_p)$. The MBA estimator uses the attractor with the smallest determinant as does the FCH estimator if $\|\overline{\boldsymbol{x}}_{k,0} - \text{MED}(\boldsymbol{X})\| \leq \text{MED}(D_i(\text{MED}(\boldsymbol{X}, \boldsymbol{I}_p))$. If the DGK location estimator $\overline{\boldsymbol{x}}_{k,0}$ has a greater Euclidean distance from $\text{MED}(\boldsymbol{X})$ than half the data, then FCH uses the median ball attractor. Let $(T_A, \boldsymbol{C}_A)$ be the attractor used. Then the estimator $(T_F, \boldsymbol{C}_F)$ takes $T_F = T_A$ and

$$\boldsymbol{C}_F = \frac{\text{MED}(D_i^2(T_A, \boldsymbol{C}_A))}{\chi^2_{p,0.5}} \boldsymbol{C}_A \tag{5}$$

where $\chi^2_{p,0.5}$ is the 50th percentile of a chi–square distribution with p degrees of freedom and F is the MBA or FCH estimator. CMVE is like FCH but the MVE criterion is used.

Example 3. Tremearne (1911) recorded *height* = x[,1] and *height while kneeling* = x[,2] of 112 people. Figure 5a shows a scatterplot of the data. Case 3 has the largest Euclidean distance of 214.767 from $\text{MED}(\boldsymbol{X}) = (1680, 1240)^T$, but if the distances correspond to the contours of a covering ellipsoid, then case 44 has the largest distance. The start $(\overline{\boldsymbol{x}}_{0,50}, \boldsymbol{S}_{0,50})$ is the classical estimator applied to the "half set" of cases closest to $\text{MED}(\boldsymbol{X})$ in Euclidean distance. The circle (hypersphere for general p) centered at $\text{MED}(\boldsymbol{X})$ that covers half the data is small because the data density is high near $\text{MED}(\boldsymbol{X})$. The median Euclidean distance is 59.661 and case 44 has Euclidean distance 77.987. Hence the intersection of the sphere and the data is a highly correlated clean ellipsoidal region. The Rousseeuw and Van Driessen (1999) DD plot is a plot of classical distances (MD) vs "robust" distances (RD). Figure 5b shows the DD plot using the median ball estimator. Notice that both the classical and MB estimators give the largest distances to cases 3 and 44. Notice that case 44 could not be detected using marginal methods.

Olive (2002) showed that if a consistent robust estimator is scaled as in (5), then the plotted points in the DD plot will cluster about the identity line if the data is MVN and about some other line through the origin if the data is not MVN but is EC with a nonsingular covariance matrix. Since multivariate procedures tend to perform well for EC data, the DD plot is useful even if outliers are not present. The median ball estimator in Figure 5b was not scaled.

In MLD simulations, sometimes the attractor can be based on a clean half set even if the half set corresponding to the start contains outliers. The MBA and FCH estimators needed $k = 5$ concentration steps while DGK needed $k = 10$.

As the dimension p gets larger, outliers that can not be detected by marginal methods (case 44 in Example 3) become harder to detect. When $p = 3$ imagine that the clean data is a baseball bat with one end at the SW corner of the bottom of the box (corresponding to the coordinate axes) and one end at the NE corner of the top of the box. If the outliers are a ball, there is much more room to hide them in the box than in a covering rectangle when $p = 2$.

The MB estimator has outlier resistance similar to $(\text{MED}(\boldsymbol{X}), \boldsymbol{I}_p)$ for distant outliers but, as shown in Example 3, can be much more effective for detecting certain types of outliers that can not be found by marginal methods. For EC data, the MB estimator is best if the data is spherical about $\boldsymbol{\mu}$ or if the data is highly correlated with the major axis of the highest density region $\{\boldsymbol{x}_i : D_i^2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \leq h^2\}$. Olive (2004) showed that if the data distribution is EC but not "spherical about $\boldsymbol{\mu}$," then for $m \geq 0$, $\boldsymbol{S}_{m,50}$ underestimates the major axis and overestimates the minor axis of the highest density region. Concentration reduces but fails to eliminate this bias. Hence the estimated highest density hyperellipsoid based on the attractor is "shorter" in the direction of the major axis and "fatter" in the direction of the minor axis than estimated regions based on consistent estimators. The following lemma is closely related to a result in Olive (2004).

Lemma 3. $(\overline{\boldsymbol{x}}_{0,50}, \boldsymbol{S}_{0,50})$ is a HB estimator of MLD.

Proof. Let $\det(\boldsymbol{C}) = |\boldsymbol{C}|$. Arcones (1995) showed that $\overline{\boldsymbol{x}}_{0,50}$ is a $\sqrt{n}$ consistent HB estimator of $\boldsymbol{\mu}$. Or use the fact that $\overline{\boldsymbol{x}}_{0,50}$ can not get arbitrarily far from MED($\boldsymbol{X}$) if the number of outliers $d < n/2$. 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 $\boldsymbol{C} \equiv \boldsymbol{S}_{0,50}$ 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 $d < n/2$. Since the MCD estimator is HB, $0 < |\boldsymbol{C}_{MCD}(c_n)| \leq |\boldsymbol{S}_{0,50}| = \lambda_1 \cdots \lambda_p$, and $\lambda_p \geq |\boldsymbol{C}_{MCD}(c_n)|/\lambda_1^{p-1} > 0$ even for d almost as large as $n/2$. $\square$

It is very difficult to drive the determinant of the dispersion estimator from a concentration algorithm to zero if at least one of the starts is nonsingular since the attractor $\boldsymbol{C}_A$ and the MCD estimator are both classical estimators applied to c_n cases. Even for d

almost as large as $n/2$, $0 < \det(\boldsymbol{C}_{\text{MCD}}) \leq \det(\boldsymbol{C}_{\text{A}})$.

The following theorem shows that the MBA and FCH estimators have good statistical properties.

Theorem 4. Suppose (E1) holds. If $(T_A, \boldsymbol{C}_A)$ is the attractor that minimizes $\det(\boldsymbol{S}_{k,M})$, then $(T_A, \boldsymbol{C}_A)$ is a HB $\sqrt{n}$ consistent estimator of $(\boldsymbol{\mu}, a_{MCD}\boldsymbol{\Sigma})$. Hence the MBA and FCH estimators are HB $\sqrt{n}$ consistent estimators of $(\boldsymbol{\mu}, c\boldsymbol{\Sigma})$ where $c = 1$ for multivariate normal data.

Proof. Under (E1) the FCH and MBA estimators are asymptotically equivalent since $\|T_{k,0} - \text{MED}(\boldsymbol{X})\| \to 0$ in probability. The estimator is HB since $0 < \det(\boldsymbol{C}_{\text{MCD}}) \leq \det(\boldsymbol{C}_{\text{A}}) \leq \det(\boldsymbol{S}_{0,50}) < \infty$ by Lemma 3 if up to nearly 50% of the cases are outliers. If the distribution is spherical about $\boldsymbol{\mu}$, then the result follows from Pratt (1959) and Lemma 2 since both starts are $\sqrt{n}$ consistent. Otherwise, the estimator with $M = 50$ trims too much data in the direction of the major axis and hence the resulting attractor is not estimating the highest density region. Hence $\boldsymbol{S}_{k,50}$ is not estimating $a_{MCD}\boldsymbol{\Sigma}$. But the DGK estimator $\boldsymbol{S}_{k,0}$ is a $\sqrt{n}$ consistent estimator of $a_{MCD}\boldsymbol{\Sigma}$ by Lemma 2 and $\|\boldsymbol{C}_{MCD} - \boldsymbol{S}_{k,0}\| = O_P(n^{-1/2})$. Thus the probability that the DGK attractor minimizes the determinant goes to one as $n \to \infty$, and $(T_A, \boldsymbol{C}_A)$ is asymptotically equivalent to the DGK estimator $(\overline{\boldsymbol{x}}_{k,0}, \boldsymbol{S}_{k,0})$. The scaling (5) makes $c = 1$ for MVN data. $\square$

The proof for CMVE is nearly identical: the CMVE volume is bounded by that of MVE and MB, and the DGK estimator can be used to estimate the highest density minimum volume region while MB volume is too large for nonspherical EC distributions.

Example 4. The estimators can also be useful when the data is not EC. The Gladstone (1905-6) data has 12 variables on 267 persons after death. Head measurements were *breadth, circumference, head height, length* and *size* as well as *cephalic index* and *brain weight*. *Age*, *height* and three categorical variables *ageclass* (0: under 20, 1: 20-45, 2: over 45), *sex* and *cause* of death (1: acute, 2: not given, 3: chronic) were given. Figure 6 shows the DD plots for the FCH, CMVE, FAST-MCD (FMCD) and MB estimators. CMVE used MB while FCH used DGK. The plots were very similar and six outliers correspond to the six infants in the data set. In spite of the categorical data, the classical and robust distances were highly correlated. In S-PLUS 2000, the FMCD estimator is singular for this data set.

4. SIMULATIONS

A simple simulation for outlier resistance is to generate outliers and count the percentage of times the minimum distance of the outliers is larger than the maximum distance of the clean cases. Then the outliers can be separated from the clean cases with a horizontal line in the DD plot. The simulation used 100 runs and $n = 200$. If $\gamma = 0.2$ then the first 40 cases were outliers. The clean cases were MVN: $\boldsymbol{x} \sim N_p(\boldsymbol{0}, diag(1, 2, ..., p))$. Outlier types were 1) a point mass $(0, ..., 0, pm)^T$ at the major axis, 2) a point mass $(pm, 0, ..., 0)^T$ at the minor axis and 3) $\boldsymbol{x} \sim N_p(pm\boldsymbol{1}, diag(1, 2, ..., p))$ where $\boldsymbol{1} = (1, ..., 1)^T$.

Maronna and Zamar (2002) suggest that a point mass orthogonal to the major axis may be least favorable for OGK, but for FAST-MCD and MBA a point mass at the major axis will cause a lot of difficulty because an ellipsoid with very small volume can cover half of the data by putting the outliers at one end of the ellipsoid and the clean data in the other end. This half set will produce a classical estimator with very small determinant by (4). Rocke and Woodruff (1996) suggest that outliers with a mean shift are hard to detect. A point mass is used although for large γ and moderate p the point mass causes numerical difficulties in that R software will declare that the sample covariance matrix is singular.

Notice that the clean data can be transformed to a $N_p(\mathbf{0}, \boldsymbol{I}_p)$ distribution by multiplying $\boldsymbol{x}_i$ by $\text{diag}(1, 1/\sqrt{2}, ..., 1/\sqrt{p})$. The counts for affine equivariant estimators such as DGK and FAST-MCD will not be changed. Notice that the point mass at the minor axis $(pm, 0, ..., 0)^T$ is not changed by the transformation, but the point mass at the major axis becomes $(0, ..., 0, pm/\sqrt{p})^T$, which is much harder to detect.

The results of the simulation are shown in Table 1. The counts for the classical estimator were always 0 and thus omitted. The simulations suggest that for fast MLD estimators, the HB MCD and MVE dispersion criteria are not adequate for screening attractors: a HB location criterion is also needed. This can be seen in Table 1 for $p = 20$ and $\gamma = 0.2$ where for a point mass at the major axis, the MCD criterion needs $pm = 10000$ for MBA and 4000 for FMCD to have a small chance of detecting the outliers, but $pm = 50$ for FCH and CMVE. The point mass outliers make the DGK determinant small (though larger than the MCD determinant by definition), but pull the DGK location estimator away from $\text{MED}(\boldsymbol{X})$. Note that FCH performance dominated MBA and was usually better than OGK. CMVE was nearly always better than OGK. For a mean shift and small p and γ the elemental FAST-MCD estimator was somewhat better than CMVE, MB, MBA and FCH. If γ is large enough then CMVE, MBA, FCH and MB dominate FAST-MCD. MB was never worse than OGK, but OGK did seem to behave like a HB estimator in that it could detect distant outliers.

The simulation suggests that marginal methods for detecting outliers should not be abandoned. We suggest making a DD plot with the $\sqrt{n}$ consistent HB FCH estimator as an EC diagnostic. Make the MB DD plot to check for outliers. Other methods that do not have proven theory can also be used as outlier diagnostics. For $p \leq 10$ make a scatterplot matrix of the variables. The plots should be ellipsoidal if the EC assumption is reasonable. Dot plots of individual predictors with superimposed histograms are also useful. For large n the histograms should be approximately symmetric if the EC assumption is reasonable.

5. CONCLUSIONS

Robust HB estimators of MLR and MLD should be i) $\sqrt{n}$ consistent for a large class of distributions, ii) easy to compute, iii) effective at detecting certain types of outliers and iv) high breakdown. Although almost all of the literature focuses either on i) and iv) or on ii) and iii), this paper shows that it is simple to construct estimators satisfying i)–iv) provided that $n > 20p$ and $p \leq 40$. These results represent both a computational and theoretical breakthrough in the field of HB MLR and MLD.

The new FCH and CMVE estimators use information from both HB location and dispersion criteria and are more effective at screening attractors than estimators such as MBA and FMCD that only use the MCD or MVE dispersion criterion. The RCMVE, RMBA and RFCH estimators are reweighted versions of CMVE, MBA and FCH that may perform better for small n, and they are HB $\sqrt{n}$ consistent estimators by Lopuhaä (1999).

For multiple linear regression, the OLS response and residual plots are useful for detecting outliers and influential cases and these plots should be made for any multiple linear regression analysis. Information from influence diagnostics can be incorporated by highlighting cases with large values of the diagnostic, and fits from both classical and HB estimators could be used to make the plots.

The collection of easily computed "robust estimators" for MLR and MLD that have not been shown to be both HB and consistent is enormous, but without theory the methods should be classified as outlier diagnostics rather than robust statistics.

Examine the estimator on many "benchmark data sets." FCH was examined on 30 such data sets. Outlier performance was competitive with estimators such as FAST-MCD. For any given estimator, it is easy to find outlier configurations where the estimator fails. For the modified wood data of Rousseeuw (1984), MB detected the planted outliers but FCH used DGK while CMVE used MB. For another data set, 2 clean cases had larger MB distances than 4 of 5 planted outliers that FAST-MCD can detect. For small p, elemental methods can be used as outlier diagnostics for consistent HB methods.

Simulations were done in *R*. Programs are in the collection of functions *rpack.txt* at (www.math.siu.edu/olive/ol-bookp.htm). The `robustbase` library was downloaded from (www.r-project.org/#doc) to compute OGK, and the `MASS` library was used to compute FAST-MCD. The `rpack` function *mldsim* was used to produce Table 1.

The `R` implementation of FCH is much faster than OGK which is much faster than FAST-MCD. Functions *covdgk*, *covmba* and *rmba* compute the scaled DGK, MBA and RMBA estimators while *covfch* and *cmve* are used to compute FCH, RFCH, CMVE and RCMVE.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grants DMS 0202922, DMS 0600933, DMS 0306304, DMS 9803622 and ACI 9619020.

REFERENCES

Arcones, M. A. (1995), "Asymptotic Normality of Multivariate Trimmed Means," *Statistics and Probability Letters,* 25, 43-53.

Bernholt, T. (2005), "Computing the Least Median of Squares Estimator in Time $O(n^d)$," *Proceedings of ICCSA 2005*, LNCS 3480, 697-706.

Bernholt, T. (2006), "Robust Estimators are Hard to Compute," technical report available from (http://ls2-www.cs.uni-dortmund.de/∼bernholt/ps/tr52-05.pdf).

Bernholt, T., and Fischer, P. (2004), "The Complexity of Computing the MCD-Estimator," *Theoretical Computer Science*, 326, 383-398.

Branco, J.A., Croux, C., Filzmoser, P., and Oliviera, M.R. (2005), "Robust Canonical Correlations: a Comparative Study," *Computational Statistics*, 20, 203-229.

Butler, R. W., Davies, P. L., and Jhun, M. (1993), "Asymptotics for the Minimum Covariance Determinant Estimator," *The Annals of Statistics,* 21, 1385-1400.

Buxton, L.H.D. (1920), "The Anthropology of Cyprus," *The Journal of the Royal Anthropological Institute of Great Britain and Ireland,* 50, 183-235.

Čížek, P. (2006), "Least Trimmed Squares Under Dependence," *Journal of Statistical Planning and Inference*, 136, 3967-3988.

Cook, R.D. (1977), "Deletion of Influential Observations in Linear Regression," *Technometrics,* 19, 15-18.

Croux, C., and Haesbroeck, G. (2003), "Implementing the Bianco and Yohai Estimator for Logistic Regression," *Computational Statistics and Data Analysis,* 44, 273-295.

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

Devlin, S. J., Gnanadesikan, R., and Kettenring, J. R. (1981), "Robust Estimation of Dispersion Matrices and Principal Components," *Journal of the American Statistical Association,* 76, 354-362.

Gladstone, R. J. (1905-6), "A Study of the Relations of the Brain to the Size of the Head," *Biometrika,* 4, 105-123.

Hawkins, D. M., and Olive, D. J. (1999), "Improved Feasible Solution Algorithms for High Breakdown Estimation," *Computational Statistics and Data Analysis,* 30, 1-11.

Hawkins, D. M., and Olive, D. J. (2002), "Inconsistency of Resampling Algorithms for High Breakdown Regression Estimators and a New Algorithm," (with discussion), *Journal of the American Statistical Association,* 97, 136-159.

He, X., Fung, W.K., and Zhu, Z.Y. (2005), "Robust Estimation in Generalized Partial Linear Models for Clustered Data," *Journal of the American Statistical Association ,* 100, 1176-1184.

He, X., and Portnoy, S. (1992), "Reweighted LS Estimators Converge at the Same Rate as the Initial Estimator," *The Annals of Statistics,* 20, 2161-2167.

Hubert, M., Rousseeuw, P. J., and Van Aelst, S. (2007), "High Breakdown Multivariate Methods," *Statistical Science*, to appear.

Hubert, M., Rousseeuw, P.J., and Vanden Branden, K. (2005), "ROBPCA: a New Approach to Robust Principal Component Analysis," *Technometrics,* 47, 64-79.

Johnson, M. E. (1987), *Multivariate Statistical Simulation,* New York: Wiley.

Johnson, R. A., and Wichern, D. W. (1988), *Applied Multivariate Statistical Analysis,* 2nd ed., Englewood Cliffs, NJ: Prentice Hall.

Lopuhaä, H. P. (1999), "Asymptotics of Reweighted Estimators of Multivariate Location and Scatter," *The Annals of Statistics,* 27, 1638-1665.

Maronna, R. (2005), "Principal Components and Orthogonal Regression Based On Robust Scales," *Technometrics*, 47, 264-273.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006), *Robust Statistics: Theory and Methods*, Hoboken, NJ: Wiley.

Maronna, R. A., and Zamar, R. H. (2002), "Robust Estimates of Location and Dispersion for High-Dimensional Datasets," *Technometrics,* 44, 307-317.

Mašíček, L. (2004), "Optimality of the Least Weighted Squares Estimator," *Kybernetika*, 40, 715-734.

Olive, D. J. (2002), "Applications of Robust Distances for Regression," *Technometrics,* 44, 64-71.

Olive, D. J. (2004), "A Resistant Estimator of Multivariate Location and Dispersion," *Computational Statistics and Data Analysis*, 46, 99-102.

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

Olive, D.J., and Hawkins, D.M. (2005), "Variable Selection for 1D Regression Models," *Technometrics*, 47, 43-50.

Peña, D. (2005), "A New Statistic for Influence in Regression," *Technometrics*, 47, 1-12.

Pison, G., Rousseeuw, P.J., Filzmoser, P., and Croux, C. (2003), "Robust Factor Analysis," *Journal of Multivariate Analysis,* 84, 145-172.

Pratt, J. W. (1959), "On a General Concept of 'in Probability'," *The Annals of Mathematical Statistics,* 30, 549-558.

Rocke, D. M., and Woodruff, D. L. (1996), "Identification of Outliers in Multivariate Data," *Journal of the American Statistical Association,* 91, 1047-1061.

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

Rousseeuw, P.J., Van Aelst, S., Van Driessen, K., and Agulló, J. (2004), "Robust Multivariate Regression," *Technometrics*, 46, 293-305.

Rousseeuw, P. J., and Van Driessen, K. (1999), "A Fast Algorithm for the Minimum Covariance Determinant Estimator," *Technometrics,* 41, 212-223.

Tremearne, A. J. N. (1911), "Notes on Some Nigerian Tribal Marks," *Journal of the Royal Anthropological Institute of Great Britain and Ireland,* 41, 162-178.

Willems, G., Pison, G., Rousseeuw, P.J., and Van Aelst, S. (2002), "A Robust Hotelling Test," *Metrika*, 55, 125-138.

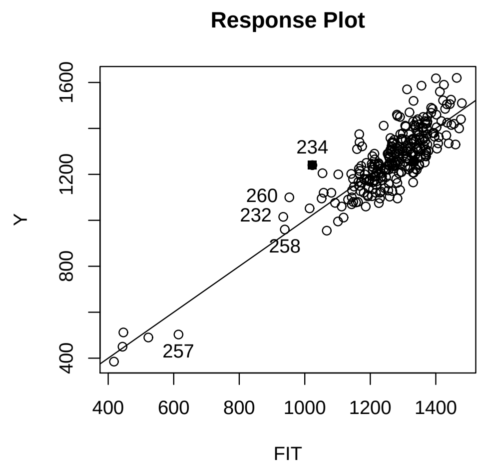

Figure 1: Gladstone Data

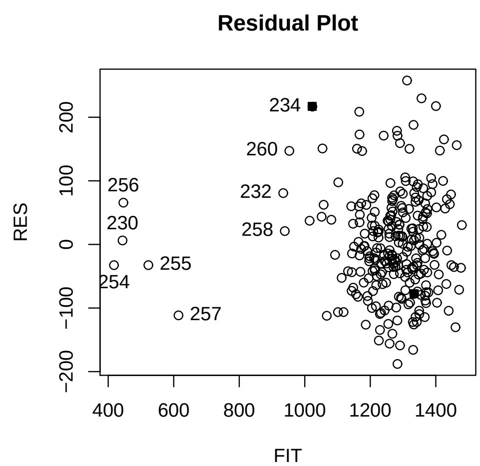

Figure 2: Gladstone Data

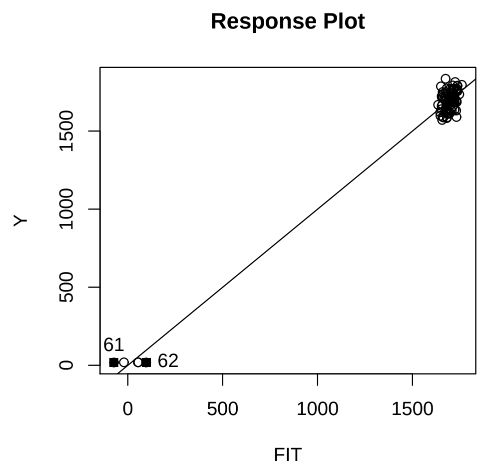

Figure 3: Buxton Data

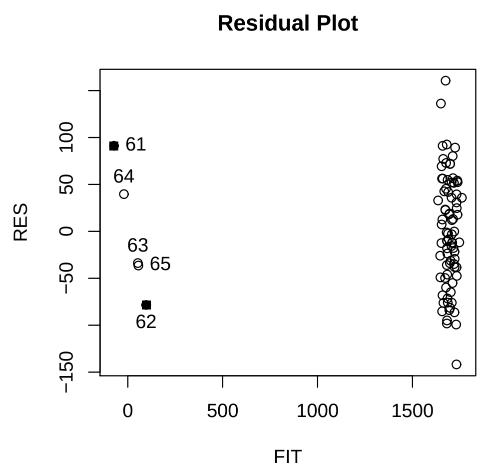

Figure 4: Buxton Data

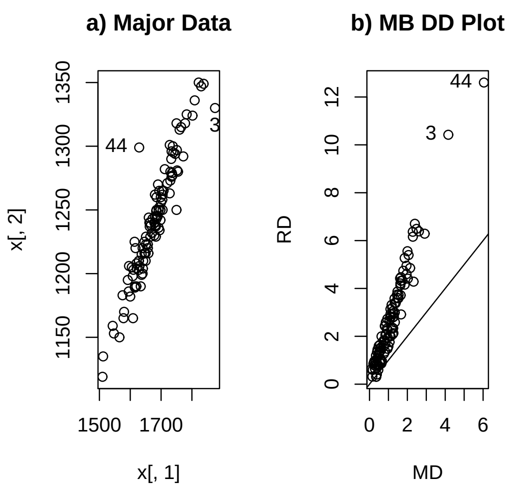

Figure 5: Plots for Major Data

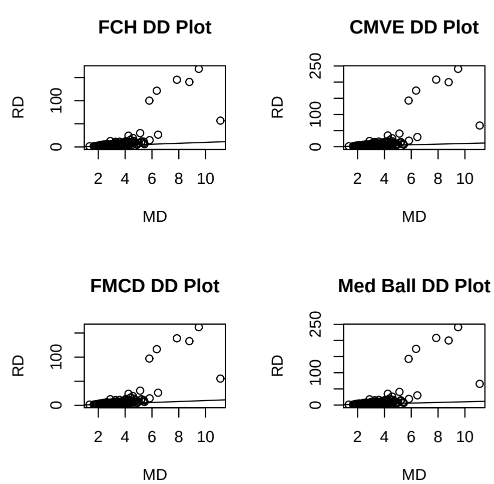

Figure 6: DD Plots for Gladstone Data

Table 1: Percentage of Times Outliers Were Detected

p	γ	type	pm	MBA	FCH	DGK	OGK	FMCD	CMVE	MB
5	.2	1	15	0	100	0	0	0	100	100
10	.2	1	20	0	4	0	0	0	16	96
20	.2	1	30	0	0	0	0	0	1	61
20	.2	1	50	0	100	0	0	0	100	100
20	.2	1	100	0	100	0	22	0	100	100
20	.2	1	4000	0	100	0	100	31	100	100
20	.2	1	10000	24	100	0	100	100	100	100
5	.2	2	15	97	100	0	71	100	100	100
10	.2	2	20	0	58	0	71	0	97	100
20	.2	2	30	0	0	0	99	0	76	100
20	.2	2	50	0	100	0	100	0	100	100
20	.2	2	100	0	100	0	100	0	100	100
20	.2	2	4000	96	100	0	100	100	100	100
5	.2	3	5	88	88	87	5	97	92	91
10	.2	3	5	92	92	84	2	100	92	94
20	.2	3	5	85	85	1	0	99	66	85
40	.4	3	20	38	38	0	0	0	40	100
40	.4	3	30	77	97	0	59	0	91	100
40	.4	3	40	91	100	0	100	0	100	100