

# **HIGH BREAKDOWN ESTIMATION OF NONLINEAR REGRESSION PARAMETERS**

by

**Arnold J. Stromberg  
University of Minnesota**

**Technical Report No. 541**

**February 1990**

**Revised June 1990**

## HIGH BREAKDOWN ESTIMATION OF NONLINEAR REGRESSION PARAMETERS

Arnold J. Stromberg<sup>1</sup>

### ABSTRACT

A new technique for computing high breakdown estimates in linear and nonlinear regression is presented. The technique is applied to compute Rousseeuw's (1984) least median of squares estimate and Yohai's MM-estimate (1987) in both simulations and examples. A method for computing the exact value of the LMS for certain nonlinear regression functions is presented and used to compare various approximations to the LMS estimate. Asymptotic standard errors for MM-estimates are compared with bootstrapped and Monte Carlo standard errors. Examples show that the deficiencies of the LMS algorithm pointed out in Cook and Weisberg (1989) are not due to their use of the PROGRESS algorithm but are inherent in the LMS estimate. The MM-estimate is shown to be a high breakdown estimate that is useful in identifying outliers and model deviations in both linear and nonlinear regression.

---

<sup>1</sup>Arnold Stromberg is Visiting Assistant Professor, Department of Applied Statistics, University of Minnesota, St. Paul, MN 55108. This work presents and extends material from his Ph.D. dissertation. He was supported by NSF Grant DMS-8701201. The author thanks David Ruppert and R. Dennis Cook for their helpful comments.

## 1. INTRODUCTION

In this paper I present a new technique for computing high breakdown estimates in linear and nonlinear regression. The technique is applied to compute Rousseeuw's (1984) least median of squares estimate and Yohai's MM-estimate (1987) in both simulations and examples. MM-estimates, which can have both a high breakdown point and high asymptotic efficiency if the errors are normally distributed, are computed to investigate Simpson and Ruppert's (1990) claim that a high efficiency estimate is important for data analysis and outlier detection as well as estimation of regression coefficients. The results presented in this paper substantiate Simpson and Ruppert's claim by showing the usefulness of MM-estimators in identifying outliers and model deviations in both linear and nonlinear regression.

High breakdown estimators have been used in linear regression to identify outliers and discover problems with masking that other diagnostic techniques missed (Atkinson 1986; Rousseeuw and von Zomeren 1990). On the other hand, Cook and Weisberg (1989) give examples showing that using Rousseeuw's PROGRESS algorithm as an approximation to the LMS estimate can find outliers when none are present and when outliers are present, identify them incorrectly. They also point out the inability of the PROGRESS approximation to the LMS estimate to identify deviations from the proposed model.

It seems possible that the problems pointed out by Cook and Weisberg are due to the deficiencies of the PROGRESS algorithm and not the least median of squares estimator. For multiple linear regression, my algorithm produces estimates of the LMS estimate with much smaller median squared residuals than the PROGRESS algorithm but these improved estimates do not seem to improve the performance of the estimator. For the class of nonlinear regression models given by  $g(\alpha + \beta x)$  where  $g$  is monotone, I am able to compute the exact value of the LMS estimate, but doing so does not improve the performance of the estimator.

One might conjecture that the problems demonstrated by Cook and Weisberg are inherent in all high breakdown methods, but I show that the MM-estimate overcomes these

problems and retains a high breakdown point. I am also able to show that the MM-estimate needs only an approximation to the LMS estimate as a starting value, and thus there is no need to go to the computational expense of computing a very accurate estimate of the LMS estimate.

The definitions of the LMS and MM estimators can easily be generalized for use in nonlinear regression, but several important issues must be addressed in order to justify their use in practice. Stromberg and Ruppert (1989) analyze breakdown properties in nonlinear regression. Stromberg (1989) gives a proof of the consistency of  $\hat{\theta}_{\text{LMS}}$  in nonlinear regression which justifies its use as the high breakdown estimate needed to compute a nonlinear MM-estimate. This paper deals with the use and computation of the LMS and MM estimators in nonlinear regression. The ideas presented here could be modified to study the performance of other high breakdown estimates.

Section 2 presents an overview of the LMS and MM estimators in linear regression. Because the computation of the exact fit to  $p$  data points is far from trivial in nonlinear regression, the PROGRESS algorithm, which computes a large number of exact fits, must be modified for use in nonlinear regression. My algorithm and a method for computing the exact value of the LMS estimate for nonlinear regression functions of the form  $g(\alpha + \beta x)$  where  $g$  is monotone are presented in section 3. In section 4, I present simulation results comparing the performance of several estimates of  $\hat{\theta}_{\text{LMS}}$  with and without outliers in the data. I also compare the estimates of  $\hat{\theta}_{\text{LMS}}$  as starting values for computing  $\hat{\theta}_{\text{MM}}$ . Section 5 discusses a method for computing standard errors for MM-estimates. The method is based on the asymptotic distribution of MM-estimates. The asymptotic standard errors are shown to be close to bootstrapped and Monte Carlo standard errors as long as the proportion of outliers is not too large. Finally, in section 6, I present several examples involving the use of high breakdown estimates in practice.

## 2. High Breakdown Estimation in Linear Regression

Rousseeuw (1984) defines the least median of squares estimator, denoted  $\hat{\theta}_{\text{LMS}}$ , for the linear regression model given by:

$$Y_i = x_i^T \theta + \epsilon_i \quad 1 \leq i \leq n \quad (2.1)$$

where the  $x_i$ 's are known  $p$ -dimensional vectors of explanatory variables, the  $\epsilon_i$  are independent with a common distribution  $F$  that is symmetric and strongly unimodal.  $\hat{\theta}_{\text{LMS}}$  is then defined by:

$$\arg \inf_{\theta \in \Theta} \text{med}_{1 \leq i \leq n} (Y_i - x_i^T \theta)^2 \quad (2.2)$$

where the median is defined as the  $\lceil \frac{n}{2} \rceil + 1^{\text{st}}$  order statistic and  $\lceil \cdot \rceil$  is the greatest integer function. He proves that  $\hat{\theta}_{\text{LMS}}$  exists and that if any  $p$  observations determine a unique value of  $\hat{\theta}_{\text{LMS}}$ , then the finite sample breakdown point of the LMS method in linear regression is

$\frac{\lceil \frac{n}{2} \rceil - p + 2}{n}$ . Kim and Pollard (1990) have recently proved that the rate of convergence for the least median of squares estimator in linear regression is  $O_p(n^{-1/3})$ . This implies that the asymptotic efficiency of the least median of squares estimator is zero.

Simulations presented in section 4 and other places (e.g., Rousseeuw and Leroy, 1987) verify the low efficiency of LMS compared this least squares for finite samples with normal errors. Using LMS as a starting value for an MM estimate preserves the high breakdown point of the LMS estimate and solves the efficiency problem. For completeness, I now review the definition of an MM estimate.

A rho function (adapted from Yohai, 1987) satisfies the following conditions:

- (i)  $\rho(0) = 0$ . (ii)  $\rho(-u) = \rho(u)$ . (iii)  $0 \leq u \leq v$  implies  $\rho(u) \leq \rho(v)$ . (iv)  $\rho$  is continuous.
- (v)  $0 < \sup(\rho(u)) < \infty$ . (vi)  $\rho(u) < \sup(\rho(u))$  and  $0 \leq u < v$  implies  $\rho(u) < \rho(v)$ .

Given the sample  $u = (u_1, u_2, \dots, u_n)$ , The scale M-estimate  $s(u)$  is the value of  $s$  that solves:

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{u_i}{s}\right) = B. \quad (2.3)$$

where  $B$  can be taken to be the expected value of  $\rho(u)$  with respect to the standard normal distribution.

Yohai (1987) defines the MM-estimate by the following three stage procedure:

1. Compute a consistent high breakdown estimate,  $\hat{\theta}_{HB}$ , of  $\theta_0$ .
2. Use the residuals  $r_i(\hat{\theta}_{HB}) = y_i - x_i^T \hat{\theta}_{HB}$ ,  $1 \leq i \leq n$  and (2.3) to compute a scale M-estimate which I denote  $\hat{s}_n$ . Let the rho function in (2.3), denoted  $\rho_0$ , be such that

$$\frac{B}{A} = 0.5 \tag{2.4}$$

where  $A = \max(\rho_0(u))$ . Huber (1981) shows that this scale M-estimate has breakdown point 0.5.

3. Let  $\rho_1$  be a rho function such that  $\rho_1(u) \leq \rho_0(u)$  and  $\sup(\rho_1(u)) = \sup(\rho_0(u)) = A$ .

The MM-estimate,  $\hat{\theta}_{MM}$ , is any solution of

$$\frac{\partial}{\partial \theta} \sum_{i=1}^n \rho_1\left(\frac{r_i(\theta)}{\hat{s}_n}\right) = 0 \tag{2.5}$$

such that

$$\sum_{i=1}^n \rho_1\left(\frac{r_i(\hat{\theta}_{MM})}{\hat{s}_n}\right) \leq \sum_{i=1}^n \rho_1\left(\frac{r_i(\hat{\theta}_{HB})}{\hat{s}_n}\right).$$

Yohai shows that the MM-estimator will have the breakdown point of the estimator used in stage one, and efficiency under normal errors determined by the choice of  $\rho_1$ . He then shows that the high breakdown estimate and  $\rho_1$  can be chosen so that the resulting MM-estimator has breakdown point 0.5 and arbitrarily high efficiency.

Because it is the most studied 50% breakdown estimator, I choose to use the least median of squares estimate ( $\hat{\theta}_{LMS}$ ) as the high breakdown estimate computed in stage 1 above. The objective function used to define  $\hat{\theta}_{LMS}$  is nondifferentiable and has many local minima, and thus the computation of  $\hat{\theta}_{MM}$  is difficult. Rousseeuw and Leroy (1987) use the

PROGRESS algorithm to approximate  $\hat{\theta}_{LMS}$ . The algorithm can be summarized as follows: First calculate the exact fit to  $p$  points, denote it  $\hat{\theta}_{ex}$ , then calculate the median residual at  $\hat{\theta}_{ex}$ . Ideally the procedure is repeated for the  $\binom{n}{p}$  possible  $p$  element subsets and the value of  $\hat{\theta}_{ex}$  producing the lowest median residual is called the least median of squares estimate. If repeating the procedures  $\binom{n}{p}$  times is computationally difficult, Rousseeuw and Leroy suggests a different method for choosing the number of subsamples. If the proportion of outliers is  $\epsilon$ , then the number of subsamples can be chosen to ensure that, with high probability at least one of the subsamples contains none of the outliers. They note that for large  $\frac{n}{p}$ , this probability is approximated by:

$$1 - (1 - (1 - \epsilon)^p)^k \quad (2.6)$$

where  $k$  is the number of subsamples.

They suggest that  $k$  could be chosen to ensure that (2.6) is at least .95, but in their algorithm they chooses  $k$  differently.  $k$  is chosen to depend on  $p$  as follows.

<u>p</u> :	1	2	3	4	5	$\geq 6$
<u>max k</u> :	500	1000	1500	2000	2500	3000

For small values of  $p$ , their choice virtually assures the choice of at least one subset containing no outliers. They appear to be choosing  $k$  to provide a better estimate of the true value of the least median of squares estimate. Note that although it saves on computation time, the cap of  $k = 3000$  implies an increasing risk as  $p$  increases of no subsamples avoiding all the outliers.

### 3. COMPUTING HIGH BREAKDOWN ESTIMATES IN NONLINEAR REGRESSION

There are several reasons why I chose to modify the PROGRESS algorithm to find an estimate of the LMS estimate in the nonlinear regression setting defined by replacing  $x_i^T \theta$  in (2.1) with  $h(x_i, \theta)$  for some known regression function  $h$ . First of all, choosing a large number of subsamples would require too much computation time in nonlinear regression. Also, since no attempt is made to minimize the objective function, the procedure rarely yields even a local

minimum. In fact, since  $k$  can be no more than 3000, the algorithm doesn't compute a consistent estimator of  $\theta_0$ .

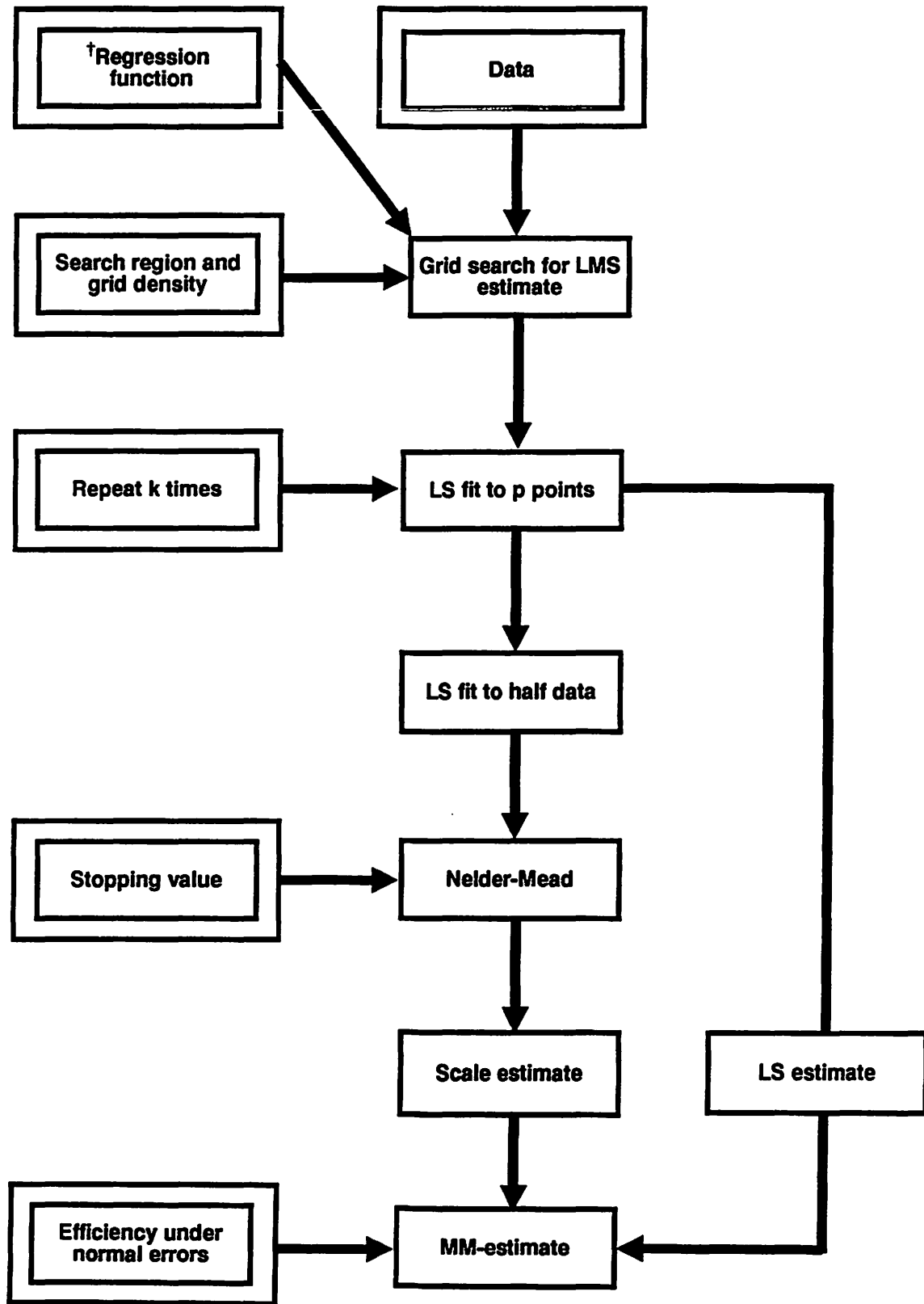
My algorithm is a multi-stage procedure. At each stage an attempt is made to improve the current best estimate, denoted  $\hat{\theta}$ , of  $\hat{\theta}_{LMS}$ . The algorithm is summarized in figure 1.

Since reasonable starting values are critical when dealing with nonlinear equations, I begin my algorithm by performing a grid search over a user-defined region and grid density. Initially, let  $\hat{\theta}$  be one of the points on the grid and compute the median squared residual at  $\hat{\theta}$ . The squared residuals at another grid point,  $\theta_{gp}$ , is compared with the median squared residual at  $\hat{\theta}$  and if more than half the residuals at  $\theta_{gp}$  are less than the median squared residual at  $\hat{\theta}$ , then the median squared residual at  $\theta_{gp}$  is computed and it becomes the new value of  $\hat{\theta}$ . This process is then repeated for all of the grid points. Note that the sort for computing the median residual is avoided at most grid points, and thus this grid search will be only slightly more complex than a grid search for a starting value for computing the least squares estimate.

Next  $\hat{\theta}$  is used as a starting value for finding the least squares fit, denoted  $\hat{\theta}_{LS}$ , to  $p$  randomly selected points. If the median squared residual at  $\hat{\theta}_{LS}$  is less than the median squared residual at  $\hat{\theta}$ , then  $\hat{\theta}_{LS}$  replaces  $\hat{\theta}$  as the current estimate of  $\hat{\theta}_{LMS}$ . This procedure is then repeated  $k$  times where  $k$  is specified by the user. The default method for computing  $\hat{\theta}_{LS}$  uses the Newton-Raphson method with  $\hat{\theta}$  as a starting value. If  $\hat{\theta}_{LS}$  can be found algebraically, users may want to save computation time by modifying the algorithm to use the algebraic solution. The default number of least squares fits is chosen to ensure that (2.6) is at least .99 when  $\epsilon = 50$  percent. At the same time the least squares fits can be used to find a starting value for finding the least squares estimate to the whole data set.

The next stage takes advantage of the fact that  $\hat{\theta}_{LMS}$  is basically trying to find a good fit to half of the data.  $\hat{\theta}$  is used as a starting value for calculating the least squares fit, denoted  $\hat{\theta}_{LS}^*$ , for data points such that  $r_i^2(\hat{\theta}) \leq \text{med}_{1 \leq i \leq n} r_i^2(\hat{\theta})$ . If  $\text{med}_{1 \leq i \leq n} r_i^2(\hat{\theta}_{LS}^*) < \text{med}_{1 \leq i \leq n} r_i^2(\hat{\theta})$ , then  $\hat{\theta}_{LS}^*$  replaces  $\hat{\theta}$  as the current estimate of  $\hat{\theta}_{LMS}$ .





†Double frames represent user inputs, single frames are steps performed by the program.

Figure 1. MM-estimation algorithm for estimating nonlinear regression parameters.

In order to find a still better estimate of  $\hat{\theta}_{LMS}$ , the current value of  $\hat{\theta}$  is then used as the starting value for an algorithm designed to minimize  $\text{med}_{1 \leq i \leq n} r_i^2(\theta)$ . Under reasonable conditions, the continuity of  $\text{med}_{1 \leq i \leq n} r_i^2(\theta)$  is proved in Stromberg (1989), but  $\text{med}_{1 \leq i \leq n} r_i^2(\theta)$  is not differentiable since as the median squared residual switches from one residual to another, the derivatives from the left and right will not generally be the same. Because the Nelder-Mead Simplex Algorithm (1965, Nelder and Mead) does not rely on derivatives of the objective function, I use it to minimize  $\text{med}_{1 \leq i \leq n} r_i^2(\theta)$ . The algorithm evaluates the objective function on a  $p+1$  dimensional simplex, where  $p$  is the dimension of  $\theta$ , and thus it never computes derivatives. I use a stopping criterion suggested by Dennis and Woods (1986)<sup>2</sup>. If  $v_i$ ,  $1 \leq i \leq p+1$  are the rows (vertices) of the simplex, then

$$\frac{1}{\max(1, \|v_1\|)} \max_{2 \leq i \leq p+1} \|v_i - v_1\| \quad (3.1)$$

is a measure of the relative size of the simplex. When (3.1) is less than the user specified stopping value or 1000 iterations of the algorithm have occurred, the Nelder-Mead algorithm is stopped. My experience has led us to chose the default value of  $10^{-4}$  which seems to be adequate if  $\hat{\theta}_{LMS}$  is being used as a starting value for  $\hat{\theta}_{MM}$ . A smaller stopping value will produce a slightly better estimate of  $\hat{\theta}_{LMS}$  at the expense of substantially more computation time. The algorithm is made significantly faster by taking advantage of the fact that if less than half the residuals at  $\theta'$  are less than  $\text{med}_{1 \leq i \leq n} r_i^2(\theta)$ , then  $\text{med}_{1 \leq i \leq n} r_i^2(\theta') > \text{med}_{1 \leq i \leq n} r_i^2(\theta)$ . The Nelder-Mead algorithm's choice for the minimum of  $\text{med}_{1 \leq i \leq n} r_i^2(\theta)$  is then final estimate of  $\hat{\theta}_{LMS}$ .

Yohai's (1987) MM-estimator for linear regression parameters can be generalized to provide a high breakdown, highly efficient under normal errors estimator of nonlinear

---

<sup>2</sup>The Nelder-Mead algorithm given in Dennis and Woods' paper is incorrect. Press, Flannery, Teukolsky, and Vetterling (1986, p289-293) present the correct algorithm.

regression parameters. In Yohai's Remark 2.3 (1987, p. 645) he suggests that for rho function  $\rho$ ,  $\rho_0$  and  $\rho_1$  can be chosen such that  $\rho_0(u) = \rho(\frac{u}{k_0})$  and  $\rho_1(u) = \rho(\frac{u}{k_1})$ .  $k_0$  is chosen so that (2.4) holds and thus ensures a high breakdown estimator.  $k_1$  is chosen to ensure high efficiency under normal errors.

My algorithm uses the function found by integrating Hampel's re-descending psi function with constants denoted a, b, and c as the rho function, denoted  $\rho_H(\cdot)$ , to determine  $\rho_0$  and  $\rho_1$ . Since (2.4) must hold,  $k_0$  should satisfy  $\frac{a}{4}(b - a + c) = \int_{-\infty}^{\infty} \rho_H(\frac{u}{k_0})f(u)du$ , where f denotes the standard normal density. If the traditional values of a = 1.5, b = 3.5 and c = 8 are chosen then numerical integration yields  $k_0 = 0.212$ . From Yohai's theorem 4.1 (1987), the asymptotic efficiency under normal errors is given by:

$$\frac{\left[ \int_{-ck_1}^{ck_1} \psi_1'(u)f(u)du \right]^2}{\int_{-ck_1}^{ck_1} \psi_1^2(u)f(u)du}$$

where f is the standard normal density. Using (a,b,c) = (1.5, 3.5, 8), numerical integration yields Table 1.

---

Table 1

$k_1$  for Various Efficiencies

Desired Efficiency(%):	80	85	90	95	96	97	98	99
$k_1$	.4950	.5704	.6877	.9014	.9687	1.0524	1.1642	1.3402

---

The scale M-estimate,  $\hat{s}_n$ , is then computed using the Newton-Raphson method to solve (2.3) using  $\rho = \rho_0 = \rho_H(\frac{\cdot}{k_0})$ , where a=1.5, b=3.5, c=8, and  $k_0=0.212$ . To improve the

chances of convergence, the Newton-Raphson algorithm is modified to allow squeezing, which means that if an iteration doesn't provide a better solution of the objective function then the step is divided by 2 and then checked again for improvement in the objective function. If ten divisions produce no improvement, then the current estimate is declared the solution.

Finally, (2.5) is solved using Newton-Raphson with squeezing where  $\rho_1 = \rho_H(\frac{c}{k_1})$ ,  $a=1.5$ ,  $b=3.5$ ,  $c=8$ , and  $k_1$  is determined by the user specified efficiency under normal errors. The algorithm minimizes the objective function using  $\hat{\theta}_{LMS}$  and then the least squares fit to all the data as starting values. Whichever starting value yields a lower value of the objective function is the final MM-estimate. The default choice for the efficiency of the MM-estimate is .95. Choosing a higher efficiency risks forcing the MM-estimate to the least squares estimate which is undesirable when outliers are present. With no outliers present, setting the efficiency at .95 will yield an estimate very close to the least squares estimate. The algorithm is coded using the Gauss Programming Language, Version 1.49b. (A version of the algorithm in XLISP-STAT (Tierney, 1989) will be available soon.) Persons interested in obtaining a copy should contact the author.

At this point I generalize part of the "Main Lemma" of Steele and Steiger (1986) and its proof. This will provide a method for computing the exact value of the least median of squares estimate for nonlinear regression functions of the form  $g(\alpha + \beta x_i)$  where  $g$  is continuous and monotonically increasing or monotonically decreasing. I use notation developed by Steele and Steiger.

The goal is to identify possible minima for the objective function

$$f(\alpha, \beta) = \text{med}_{1 \leq i \leq n} (y_i - g(\alpha + \beta x_i))^2 = \text{med}_{1 \leq i \leq n} r_i^2(\alpha, \beta)$$

I assume that the parameter space for  $(\alpha, \beta)$  is  $\mathfrak{R}^2$  and that the data  $(x_i, y_i)$ ,  $1 \leq i \leq n$  are in general position, meaning that any two data points determine a unique  $(\alpha, \beta)$ . This implies

that the  $x_i$ 's must be unique. Let  $\ell_{\alpha,\beta} = \{(x,y): y = g(\alpha + \beta x)\}$ . If  $x_i < x_j < x_k$  and  $r_i(\alpha,\beta) = -r_j(\alpha,\beta) = r_k(\alpha,\beta)$ , then  $\ell_{\alpha,\beta}$  equioscillates with respect to  $i, j$ , and  $k$ . A consequence of the general position assumption and the monotonicity of  $g$  is that there is one and only  $\ell_{\alpha,\beta}$  that equioscillates with respect to any three distinct data points. Note that any  $\ell_{\alpha,\beta}$  divides  $\{1,2, \dots, n\}$  into

$$B_{\alpha,\beta} = \{i: r_i^2(\alpha,\beta) > f(\alpha,\beta)\}$$

$$M_{\alpha,\beta} = \{i: r_i^2(\alpha,\beta) = f(\alpha,\beta)\}$$

$$S_{\alpha,\beta} = \{i: r_i^2(\alpha,\beta) < f(\alpha,\beta)\}.$$

### Theorem

If  $(\alpha^*, \beta^*)$  is a local minimum of  $f(\alpha, \beta)$ , then  $\ell_{\alpha^*, \beta^*}$  equioscillates with respect to three data points.

### Proof

First assume that  $|M_{\alpha^*, \beta^*}| = 1$ . Fix  $\beta^*$ . The monotonicity of  $g$  implies that by shifting  $\alpha^*$  in the appropriate direction we can decrease the median sized squared residual. For a small enough shift the continuity of the residuals implies that the points indexed by  $S$  will remain unchanged. Thus  $f(\alpha, \beta)$  will be decreased, contradicting the fact that  $(\alpha^*, \beta^*)$  is a local minimum. If  $|M_{\alpha^*, \beta^*}| = 2$ , then the monotonicity of  $g$  and the general position assumption imply that by shifting  $\alpha^*$  and  $\beta^*$  in the appropriate directions we can reduce the median squared residuals equally. Again, for a small enough shift in the parameters, the continuity of residuals ensures that the points indexed by  $S$  remain unchanged. Thus  $f(\alpha, \beta)$  will be reduced which violates the assumption of the theorem. If  $|M_{\alpha^*, \beta^*}| = 3$  and  $\ell_{\alpha^*, \beta^*}$  does not equioscillate, then the same argument implies that  $(\alpha^*, \beta^*)$  can not be a local minimum. Thus the conclusion of the theorem must hold.  $\square$

The theorem implies that the exact value of the LMS estimate can be found by computing, either numerically or algebraically, each of the  $\binom{n}{3}$  possible local minima and then comparing their median squared residuals to find the global minimum.

#### 4. Simulation results

Monte Carlo simulations can be used to study the behavior of  $\hat{\theta}_{LMS}$  and  $\hat{\theta}_{MM}$  in practice. The general procedure used to generate the data sets is as follows: Given the regression function  $h$ , a data point  $(x_i, y_i)$  is generated at  $\theta = \theta_0$  by letting  $x_i$  be randomly generated from a given distribution for  $X_i$ , and  $y_i = h(x_i, \theta_0) + \epsilon_i$ , where  $\epsilon_i$  is randomly generated from a given distribution for  $E_i$ . When outliers are present in the data set, they are generated as above except that  $y_i$  is shifted by a specified amount. The points chosen to be outliers will have  $x$  values close together, since outliers of this type are difficult for many robust estimators to handle.

I will consider three models. The first is simple linear regression where

$$y_i = \alpha_0 + \beta_0 x_i + \epsilon_i \quad 1 \leq i \leq n$$

and  $(\alpha_0, \beta_0) = (1, 1)$ . The  $x_i$ 's are randomly generated from a uniform(0,10) distribution, and the  $\epsilon_i$ 's are randomly generated from a standard normal distribution. When outliers are present, they are generated at  $x_{(15)}$  through  $x_{(26)}$  by letting their  $y$  values be increased by 40. The second model considered is the Michaelis-Menten model. I use the parameterization suggested by Ratkowsky (1983), namely

$$y_i = \frac{\alpha_0}{e^{\beta_0 + X_i}} + \epsilon_i \quad 1 \leq i \leq n.$$

$(\alpha_0, \beta_0)$  is set to (10, 0). The  $x_i$ 's are randomly generated from a uniform(0,10) distribution, and the  $\epsilon_i$ 's are randomly generated from a standard normal distribution. When outliers are present, they are generated at  $x_{(10)}$  through  $x_{(21)}$  by letting their  $y$  values be increased by 40. Finally I consider the isomerization model used in Carr (1960) and in Bates and Watts (1988). The model is given by:

$$y_i = \frac{X_{i2} - X_{i3} / 1.632}{\alpha_0 + \beta_0 X_{i1} + \gamma_0 X_{i2} + \delta_0 X_{i3}} + \epsilon_i \quad 1 \leq i \leq n$$

where  $(\alpha_0, \beta_0, \gamma_0, \delta_0) = (1.00, .05, .03, .10)$ . The  $x_{i1}$ 's were generated from a uniform(100,500) distribution, the  $x_{i2}$ 's from a uniform(100,300) distribution, and the  $x_{i3}$ 's from a uniform(0,100) distribution. The  $\epsilon_i$ 's were generated from a normal distribution with mean zero and variance .001. When outliers are present, they are generated at  $x_{(15)1}$  through  $x_{(26)1}$  by letting their y values be increased by 1.

To compute the estimates for a given regression function and data set, I use a slightly modified version of my algorithm. Since the data is generated from a specified  $\theta_0$ , a grid search could easily be specified that would find a value very near the actual value of the least median of squares estimate. To avoid this problem and to speed computation time, I skip the grid search step of my algorithm and use the algebraic solution to the least squares fit to p points. I use  $10E-4$  as the stopping value for the Nelder-Mead algorithm and the efficiency of the MM-estimate is set to .95. The number of least squares fits to p points will be specified. For each model, I generate 500 data sets with sample size 30 and then compute the estimates, I then move 40% of the y values so they are outliers and recompute the estimates.

Table 2 summarizes the performance of  $\hat{\theta}_{LMS}$  and  $\hat{\theta}_{MM}$  relative to  $\hat{\theta}_{LS}$  when no outliers are present for all three models. Since all the estimators appeared to be unbiased, the table presents the relative efficiency of the estimators compared with the least squares estimator. The relative efficiency is the ratio of the sample variances of the two estimates for the same simulated data sets. For the simple linear regression model, the exact value of the LMS estimate is used. By reparameterizing the Michaelis-Menten model, it can be seen that it is of the form  $g(\alpha + \beta x)$  where g is monotonically decreasing, and thus the theorem presented in section 3 can be used to compute the exact value of the LMS estimate. For the isomerization model I use my algorithm with 108 least squares fits as the estimate of  $\hat{\theta}_{LMS}$ . Note that for all three models the relative efficiency of the variances of the least median estimates to the least squares estimates is about .2 while the MM estimates have relative efficiency of about .9.

Table 2

Efficiencies (Relative to  $\hat{\theta}_{LS}$ ) of  $\hat{\theta}_{MM}$  and  $\hat{\theta}_{LMS}$  with no outliers present  
for various Regression Models

Estimator/Model	<u>Simple Linear Regression</u>	<u>Michaelis-Menten</u>	<u>Isomerization</u>
$\hat{\theta}_{LMS}$	(.219, .201) <sup>†</sup>	(.189, .182)	(.136, .141, .143, .158)
$\hat{\theta}_{MM}$	(.926, .924)	(.919, .935)	(.951, .910, .975, .869)

<sup>†</sup>Each table entry is the ratio of the sample variance of the high breakdown estimates for 500 trials to the sample variance of the least squares estimates for the same 500 trials.

When severe outliers are present in the data, simulations show that  $\hat{\theta}_{MM}$  outperforms  $\hat{\theta}_{LMS}$  in terms of variance of the estimates. The sample means and sample variances of the following estimates are presented in table 3:

- $\hat{\theta}_{LMS}$  : The exact value of the least median of squares estimate.
- $\hat{\theta}_{MM}$  : The MM-estimate using  $\hat{\theta}_{LMS}$  as a starting value.
- $\hat{\theta}_{LS}$  : The least squares estimate
- $\hat{\theta}_{M25}$  : The MM-estimate using my algorithm with 25 least squares fits.
- $\hat{\theta}_{M47}$  : The MM-estimate using my algorithm with 47 least squares fits.
- $\hat{\theta}_{M108}$  : The MM-estimate using my algorithm with 108 least squares fits.

Table 3 reveals a number of interesting results. As is to be expected the least squares estimates are heavily influenced by the outliers. Note that both  $\hat{\theta}_{LMS}$  and  $\hat{\theta}_{MM}$  ignore the presence of the outliers and thus I can compare their performances by looking at the variances of the estimates. In both simple linear regression and the Michaelis-Menten model the variance of the  $\hat{\theta}_{MM}$  estimates is approximately half of the variance of the  $\hat{\theta}_{LMS}$  estimates. Next note that the performance of  $\hat{\theta}_{M25}$  for both simple linear regression and the Michaelis-Menten



model is almost as good as  $\hat{\theta}_{MM}$ . This seems to indicate that there is no particular need to compute the exact value of the  $\hat{\theta}_{LMS}$  even when it is possible. In practice this should reduce much of the computational difficulty of high breakdown estimates.

Other simulation results (Stromberg, 1989) show that if none of subsamples of  $p$  points contain all "good" points, the estimate of  $\hat{\theta}_{LMS}$  will be poor enough that the MM-estimate will be far from the  $\hat{\theta}_{MM}$ . As is indicated by the improved performance of  $\hat{\theta}_{M108}$  over  $\hat{\theta}_{M47}$  in table 3, the number of least squares fits should increase with the dimension of the parameter space. It seems likely that the number of least squares fits that provide a sufficiently good approximation to the least median of squares estimate will also be affected by the model, the sample size, the proportion of outliers, and the estimated variance.

Table 3

Performance Summary of Regression Estimators for various models  
in the presence of 40% outliers

Estimator\Model	<u>Simple Linear Regression</u>	<u>Michaelis-Menten</u>	<u>Isomerization</u>
$\hat{\theta}_{LMS}$	(.96942, 1.00486) <sup>†</sup> (0.3231, 0.0124) <sup>††</sup>	(10.0560, -.090605) (0.6789, 0.1439)	
$\hat{\theta}_{MM}$	(0.97166, 1.00824) (0.1479, 0.0056)	(10.0483, 0.00873) (0.3297, 0.0712)	
$\hat{\theta}_{LS}$	(0.71500, 4.24160) (11.213, 0.1736)	(25.4779, 3.5733) (3.5733, 0.8132)	(2.50036, .037859, .033551, .088539) (1.05E0, 3.65E-6, 2.08E-5, 7.76E-5)
$\hat{\theta}_{M25}$	(0.96812, 1.01269) (0.1497, 0.0093)	(10.0867, 0.01259) (0.6058, 0.0669)	
$\hat{\theta}_{M47}$			(1.02202, .049788, .030141, .099801) (9.04E-2, 3.64E-6, 2.92E-6, 8.16E-6)
$\hat{\theta}_{M108}$			(1.02020, .049998, .029918, .099951) (3.81E-2, 2.03E-7, 5.63E-7, 2.27E-6)

<sup>†</sup> Mean of parameter estimates for 500 trials.

<sup>††</sup> Variance of parameter estimates for the same 500 trials.

## 5. Standard Errors

I now consider standard errors for MM-estimates. Using Yohai's (1987) Theorem 4.1, it is reasonable to approximate the asymptotic variance of MM-estimates by the diagonal of the covariance matrix:

$$\frac{[\frac{1}{n-p}][\hat{s}_n]^2 \sum_{i=1}^n [\psi(\frac{r_i(\theta)}{\hat{s}_n})]^2}{[\frac{1}{n} \sum_{i=1}^n \psi'(\frac{r_i(\theta)}{\hat{s}_n})]^2} \mathbf{I}^{-1} \Bigg|_{\theta = \hat{\theta}_{MM}}$$

where the  $jk^{\text{th}}$  element of  $\mathbf{I}$ ,  $j, k \in \{1, 2, \dots, p\}$  is  $\sum_{i=1}^n \frac{\partial[h(x_i, \theta)]}{\partial \theta_j} \frac{\partial[h(x_i, \theta)]}{\partial \theta_k}$  and  $\psi$  is the derivative of  $\rho$ .

This estimate of the variance of MM-estimates has several drawbacks. In most regression settings it has breakdown point  $1/n$  since allowing one  $x_i$  to go to infinity can usually drive the standard errors to infinity. Morgenthaler (1989) argues convincingly that the finite sample efficiency of asymptotically efficient high breakdown estimates like the MM-estimate may be small if the design contains high leverage points. In order to study the performance of using (5.1) to approximate standard errors for finite samples, I computed bootstrap estimates of the standard errors for the first three Michaelis-Menten data sets generated for the preceding Monte Carlo study. 20% outliers were generated at  $x_{(10)}$  through  $x_{(15)}$  by allowing their  $y$  values to be increased by 40. (See table 4.)

The bootstrap standard errors were computed as Efron (1982, p. 35,36) suggests. The procedure is summarized and applied to MM-estimates as follows:

- (1) Compute  $\hat{\theta}_{MM}$  for a given data set and regression function  $h$ .
- (2) Let  $\hat{F}$  put mass  $1/n$  at each  $\hat{\epsilon}_i = y_i - h(x_i, \hat{\theta}_{MM})$ .
- (3) Draw a bootstrap data set  $Y_i^* = h(x_i, \hat{\theta}_{MM}) + \epsilon_i^*$  where the  $\epsilon_i^*$  are i.i.d. from  $\hat{F}$ .
- (4) Compute  $\hat{\theta}_{MM}^*$  for the bootstrap data set.
- (5) Repeat steps 3 and 4  $B$  times, resulting in bootstrap replicates  $\hat{\theta}_{MM}^{*1}, \hat{\theta}_{MM}^{*2}, \dots, \hat{\theta}_{MM}^{*B}$ .

(6) The bootstrap standard errors can be estimated by the square of the main diagonal of the covariance matrix given by:  $\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_{MM}^{*b} - \hat{\theta}_{MM}^*)(\hat{\theta}_{MM}^{*b} - \hat{\theta}_{MM}^*)^T$  where  $\hat{\theta}_{MM}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_{MM}^{*b}$ .

Table 4

Comparison of Asymptotic Standard Errors with Bootstrap Standard Errors

Michaelis-Menten Model

No Outliers      20% outliers      40% outliers

Monte Carlo Standard Errors:      (0.5178, 0.2584)      (0.6644, 0.3797)      (0.5742, 0.2668)

---

Data Set #1	MM-estimates	(10.464, 0.2823)	(10.409, 0.3148)	(10.342, 0.2795)
	Asymptotic SE:	(0.4507, 0.2289)	(0.5466, 0.2735)	(0.6535, 0.3364)
	Bootstrap SE:	(0.5131, 0.2489)	(0.6916, 0.3048)	(82.302, 3.7855)

---

Data Set #2	MM-estimates	(10.942, 0.3761)	(11.045, 0.3555)	(11.105, 0.3634)
	Asymptotic SE:	(0.6779, 0.2554)	(0.7243, 0.2730)	(0.8709, 0.5324)
	Bootstrap SE:	(0.8961, 0.2876)	(1.0212, 0.3221)	(1418.9, 13.918)

---

Data Set #3	MM-estimates	(10.239, -0.0015)	(10.185, -.0059)	(10.337, 0.0405)
	Asymptotic SE:	(0.4536, 0.2272)	(0.5427, 0.2739)	(0.7764, 0.5410)
	Bootstrap SE:	(0.4504, 0.2187)	(0.4906, 0.2313)	(80.138, 4.1669)

---

Note from table 4 that the bootstrap standard errors are quite close to the Monte Carlo and asymptotic standard errors for up to 20% outliers but that the bootstrap standard errors are much larger for more than 20% outliers. At least in these cases, a reasonable explanation for the inflated bootstrap standard errors with more than 20% outliers is the bias created by the fact that at least some of bootstrap data sets will contain more than 50% outliers, and thus the MM-estimate will fit the outliers for these data sets. A simple binomial calculation shows that if 10% outliers are present then the probability that a bootstrap data set will contain at least 50% outliers is  $3.56 \times 10^{-8}$ . With 20% outliers, the probability is  $2.31 \times 10^{-4}$ . In either case the chance of seeing more than a few bootstrap data sets containing 50% outliers is very small, and thus the bootstrap standard errors are close to the asymptotic standard errors. On the other hand, with 30% outliers present, the chance that a particular bootstrap data set contains at least 50% outliers is .0169. This means that at least a few of the bootstrap data sets will contain 50% outliers, and thus the bootstrapped standard errors

will be much larger than the asymptotic or monte carlo standard errors. In conclusion, it seems that Morgenthaler's concern over finite sample efficiency is not a problem in these examples. Perhaps this is because the data sets do not contain high enough leverage points to see the effect he discusses. In any case, it is valuable to know that with up to 20% outliers in the data, all three estimates of the standard error are similar.

## 6. Examples

### Growth of Prices in China

Consider the Annual Rates of Growth of Prices in China data set analyzed in by Rousseeuw and Leroy (1987, p.51). The data set is in Table 5. Rousseeuw and Leroy chose to use a simple linear regression model. The least squares estimate is (-1049.468, 24.845) which obviously provides a poor fit to the data.

---

**Table 5**  
**Rates of Growth of Average Prices in the Main Cities of Free China from 1940 to 1948**

<u>Year</u>	<u>Growth of Prices</u>
1940	1.62
1941	1.63
1942	1.90
1943	2.64
1944	2.05
1945	2.13
1946	1.94
1947	15.50
1948	364.00

Source: Simkin (1978)

---

Rousseeuw and Leroy (1987) report that using PROGRESS, their estimate of  $\hat{\theta}_{LMS}$  is (-2.468, 0.102). Since a simple linear regression model is being used,  $\hat{\theta}_{LMS}$  can be computed exactly, this yields the estimate (-3.4218, 0.1250). Note that even in this very simple case the exact value of  $\hat{\theta}_{LMS}$  is surprisingly far from the PROGRESS solution, indicating the difficulty

in computing  $\hat{\theta}_{LMS}$  using the PROGRESS algorithm. Using my algorithm with the default settings yields the global minimum for  $\hat{\theta}_{LMS}$ . Computing the MM-estimate with 95% efficiency under normal errors and 25 least squares fits, I find the MM-estimate (-1.3095, 0.0754). Virtually the same estimate is obtained using either the exact value of  $\hat{\theta}_{LMS}$  or the PROGRESS estimate for the estimate of  $\hat{\theta}_{LMS}$  needed to compute  $\hat{\theta}_{MM}$ . The three estimates for the first 7 data points are graphed in figure 2.

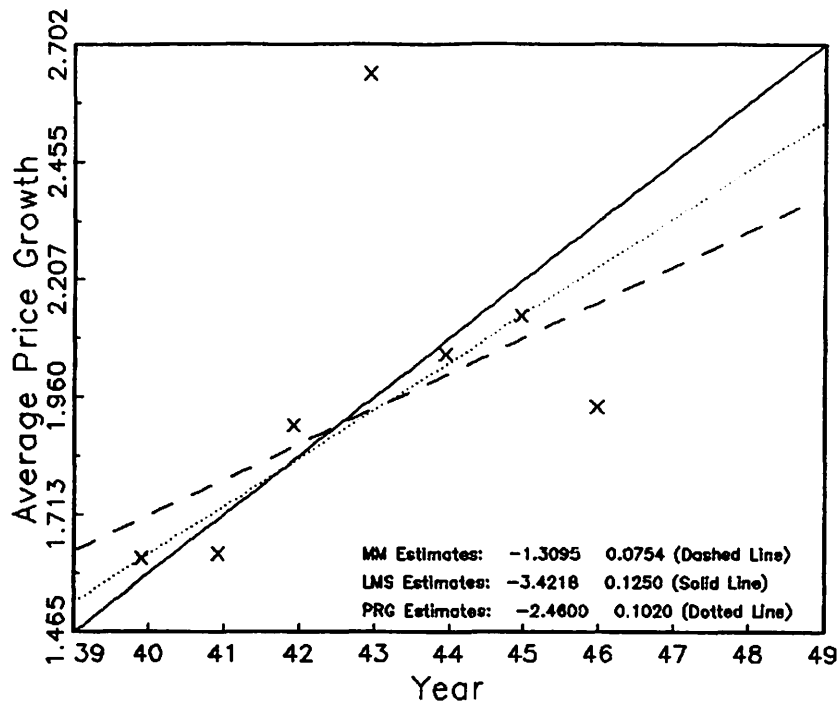


Figure 2. Plot showing high breakdown estimates for China data.

Note that the  $\hat{\theta}_{LMS}$  estimates are finding the best fit to 5 data points (since there are 9 total data points), and thus they ignore the data point at (46, 1.94). The MM-estimate allows this point to influence it, and therefore estimates the slope differently. If the efficiency is chosen to be 100%, the MM-estimate will be the least squares estimate. For this example we will look at the effect of the starting value and the choice of efficiency on the MM-estimate.

Table 6 presents the MM-estimates of the slope parameter for several different efficiencies and using the exact LMS estimate and the PROGRESS estimate.

Table 6

MM-estimates for Various Efficiencies using two different estimates of  $\hat{\theta}_{LMS}$

Efficiency (%)	MM-estimate of slope parameter starting at PROGRESS est.	MM-estimate of slope parameter starting at exact LMS value
1	.10305	.10281
5	.10431	.10552
10	.10628	.10684
20	.10726	.10698
30	.10623	.10075
40	.10104	.09471
50	.09549	.08851
60	.08889	.08502
70	.08504	.08046
80	.07945	.07548
90	.07536	.07536
95	.07536	.07536
99	.07536	.07536
99.9	.07536	.07536
99.99	.07536	.07536

As the efficiency of the MM-estimate approaches zeros, the estimate will be unstable, regardless of how  $\hat{\theta}_{LMS}$  is estimated. This is because two points can be fit exactly, the sum of  $\rho$ (standardized residuals) at the exact fit to any two data points will be local minimum and all such local minima will have the same value of the objective function. As the efficiency increases, more points are influencing the MM-estimate. It is interesting to note that the least squares estimate to the five data points selected by the LMS estimates is (-2.788, .10977) which is very close to the MM-estimates for low efficiencies. As the efficiency increases, the point at (46, 1.94) is given increasing weight. The least squares fit to these six points is (-1.3620, .07536). Finally the least squares estimate to the seven "good" points is (-1.2532, .07535). The fact that the addition of the seventh point has very little influence on the slope parameter of the least squares estimate explains the stability of the MM-estimates for high efficiencies. Note that for efficiencies above 90%, the slope estimates from table 6 are almost identical.

### Cloud Seeding Data

This data set (Cook and Weisberg 1982) summarizes the results of a cloud seeding experiment in Florida in 1975. On each of 24 days suitable for seeding, the following six explanatory variables were recorded:

A: "Action" was set to zero if no seeding took place and to one if seeding occurred.

T: "Time" was the number of days since the beginning of the experiment.

S: "Suitability" was a measure of the days suitability for seeding.

C: "Echo coverage" was the percent cloud coverage in the experimental area.

P: "Prewetness" was the log of total rainfall in the target area in the hour before seeding.

E: "Echo motion" was set to 1 for a moving radar echo and 2 for a stationary radar echo.

The response variable was  $\log(\text{rainfall})$  in a target area for a six hour period. Cook and Weisberg (1989) point out that Cook's distance labels point 1 (where the points are numbered 0,1,2, ..., 23) as highly influential, while the PROGRESS approximation to the LMS estimate labels it as a "good" leverage point because it has high leverage and a comparatively small residual. From the least squares residual plot (part of figure 3), they determine that points 6 and possibly 23 are outliers. The PROGRESS residual plot (also part of figure 3) indicates that point 6 is an outlier but not 23. Rousseeuw's method of reweighted least squares also identifies 3 additional outliers. Finally Cook and Weisberg (1989) show the need for an interaction term involving the action variable and at least one of the other explanatory variables. This is done by finding a "bowl" shape in the 3-dimensional plot (reproduced as part of figure 4) of  $x$  verses  $y$  verses  $z$  where:  $x$  is the difference between least squares fitted values for the full model and the least squares fitted values for the model without the action variable.  $y$  is the least squares residuals for the full model.  $z$  is the least squares fitted values for the model without the action variable. The corresponding plot (part of figure 4) for the PROGRESS approximation to the LMS residuals replaces the  $y$ -axis variable with the PROGRESS residuals. It shows little sign of the need for an interaction term.

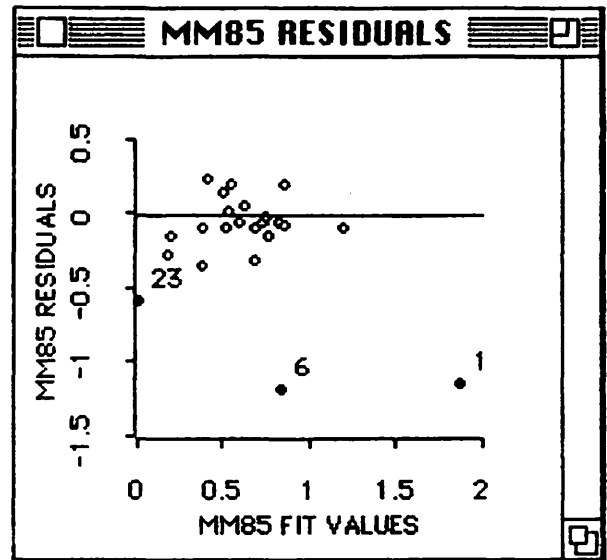
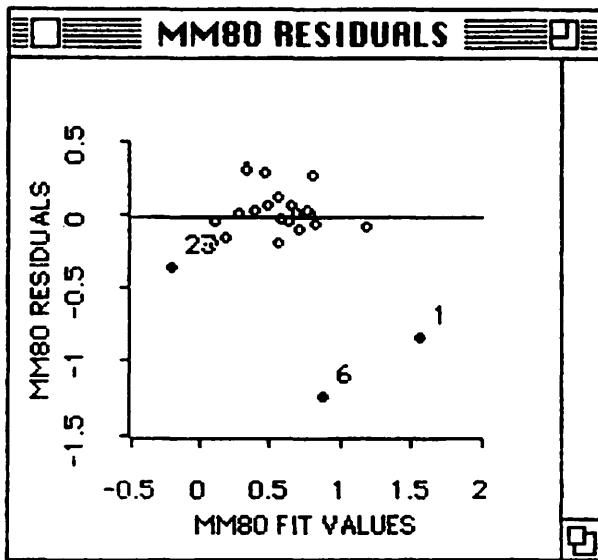
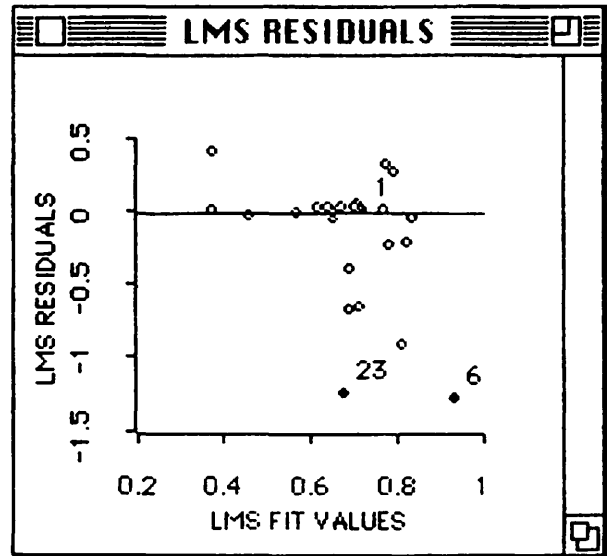
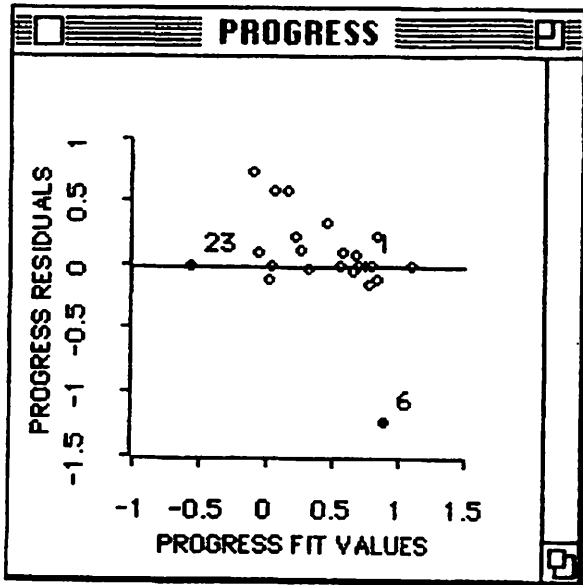


Figure 3 (part 1). Residual plots for cloud seeding data.



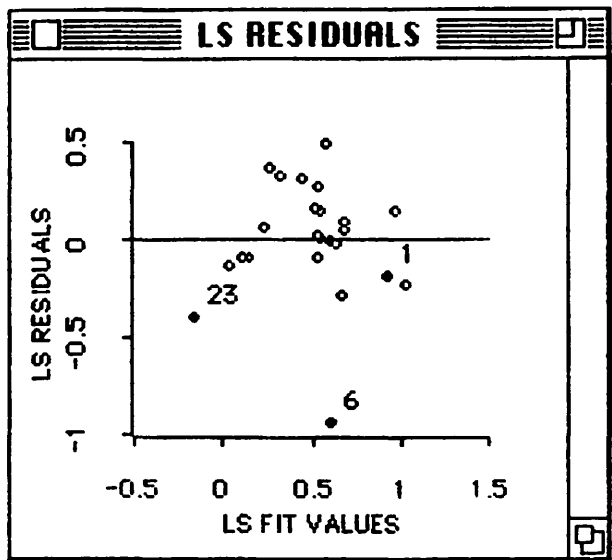
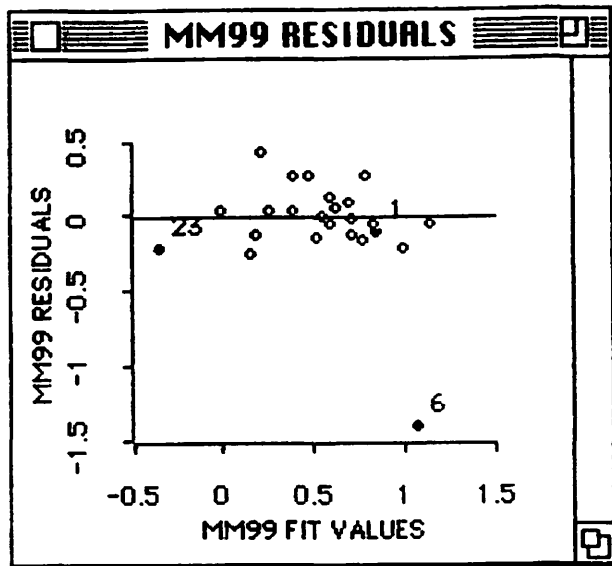
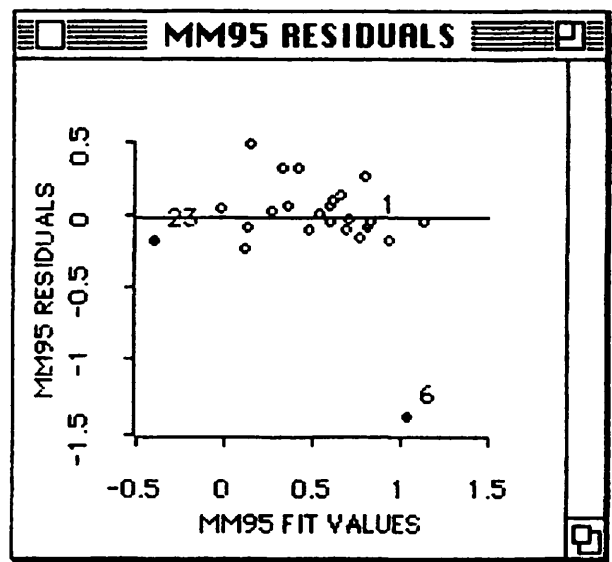
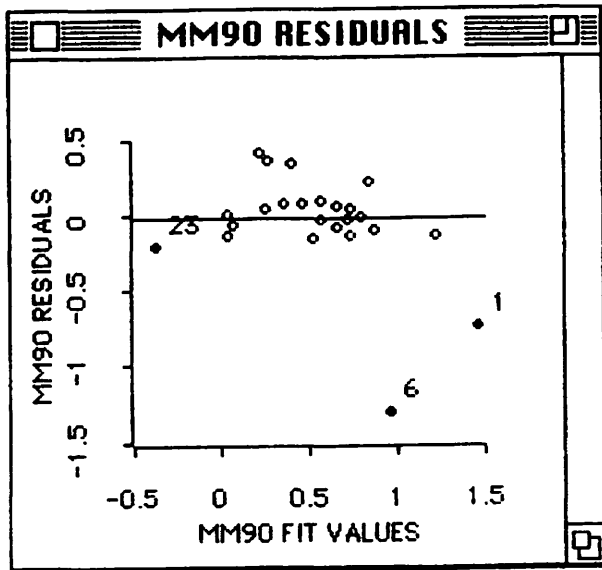


Figure 3 (part 2). Residual plots for cloud seeding data.

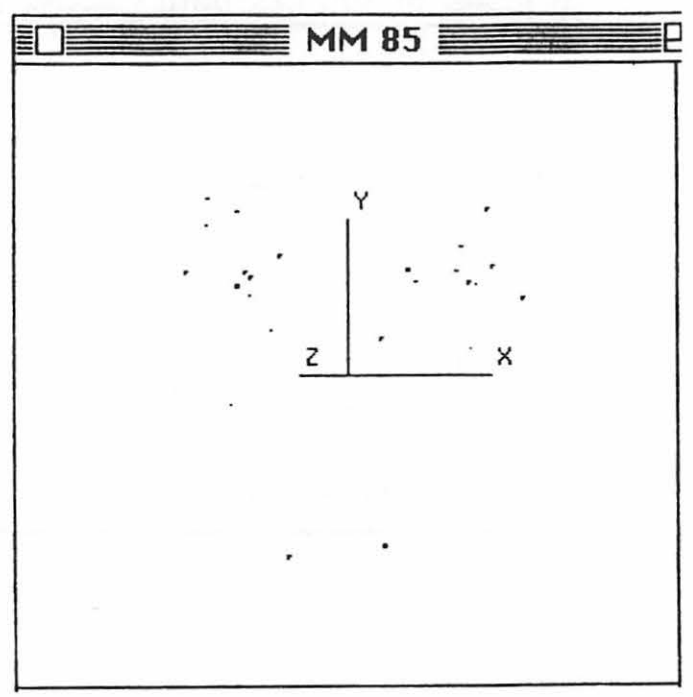
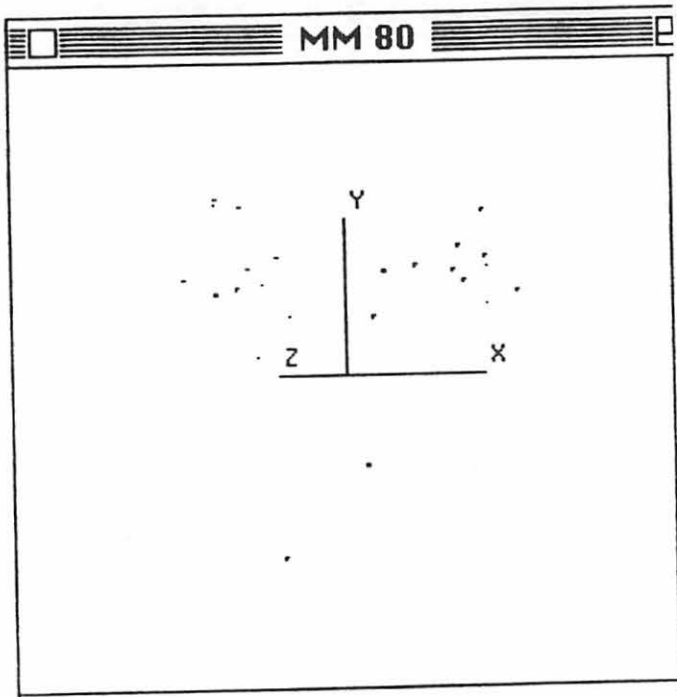
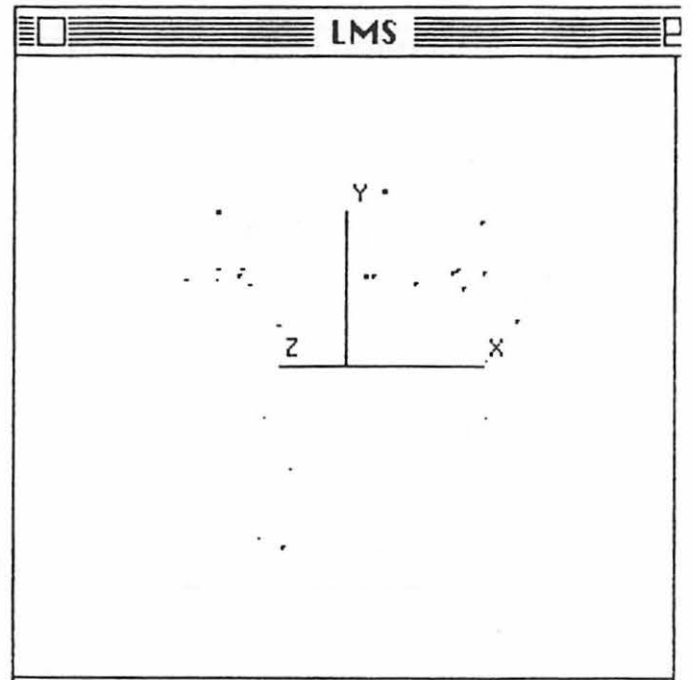
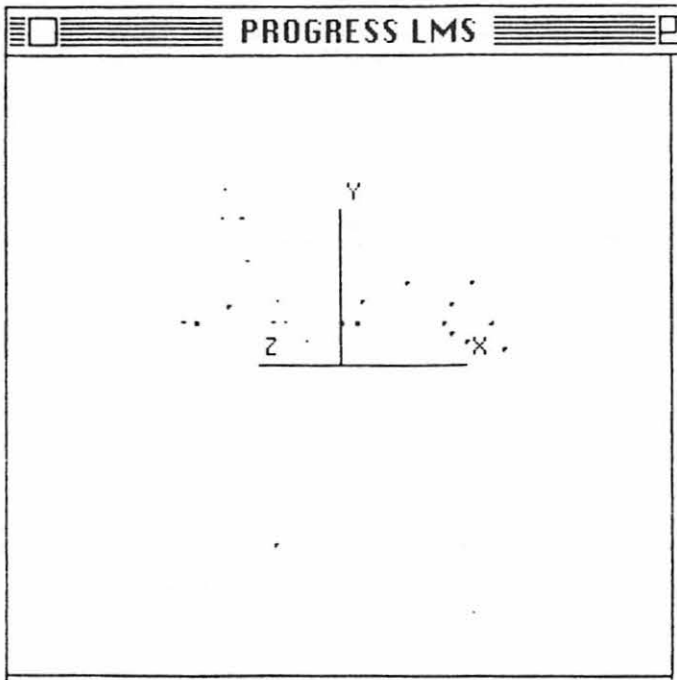


Figure 4 (part 1). 2-D projections of 3-D plots for cloud seeding data.

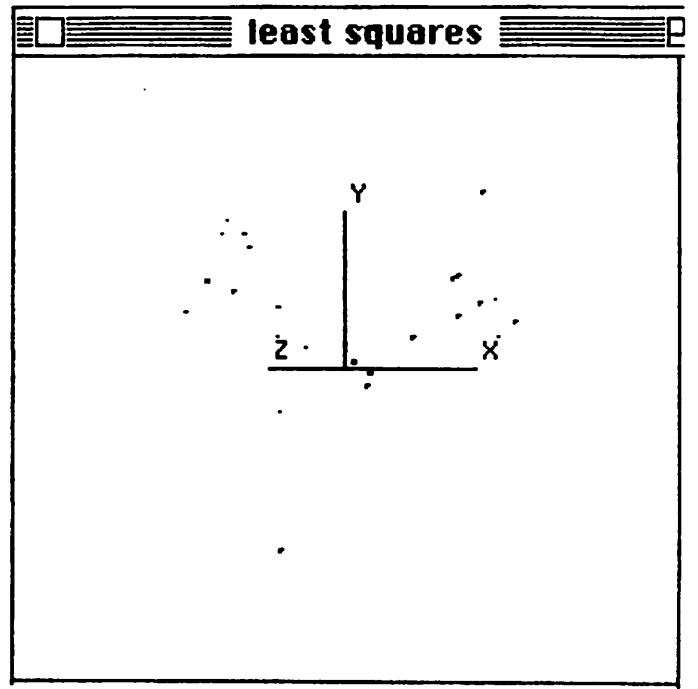
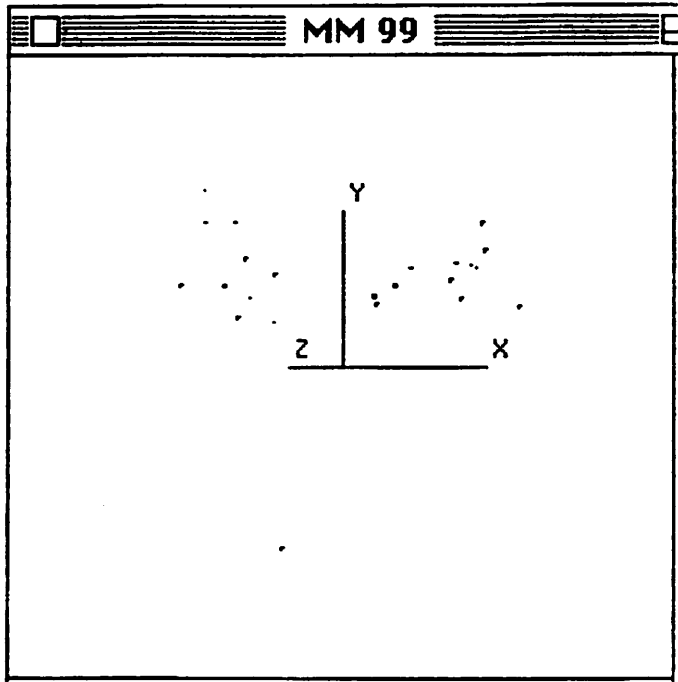
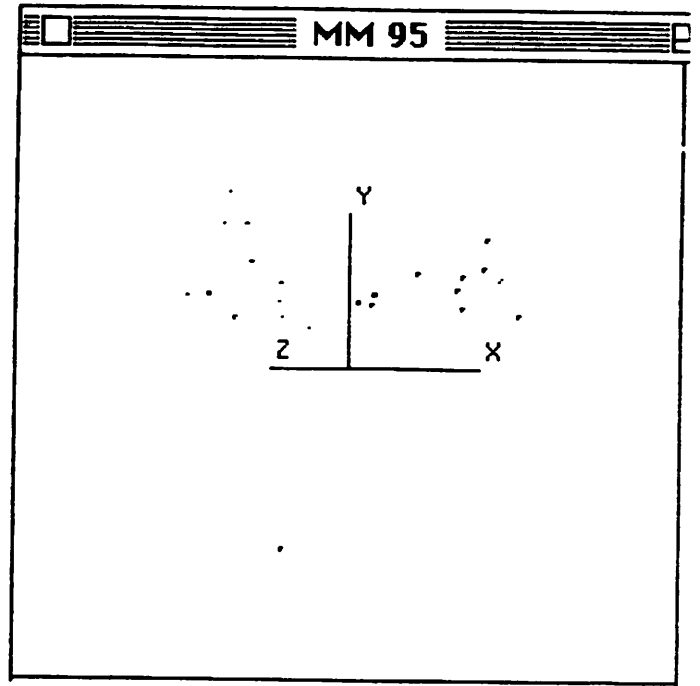
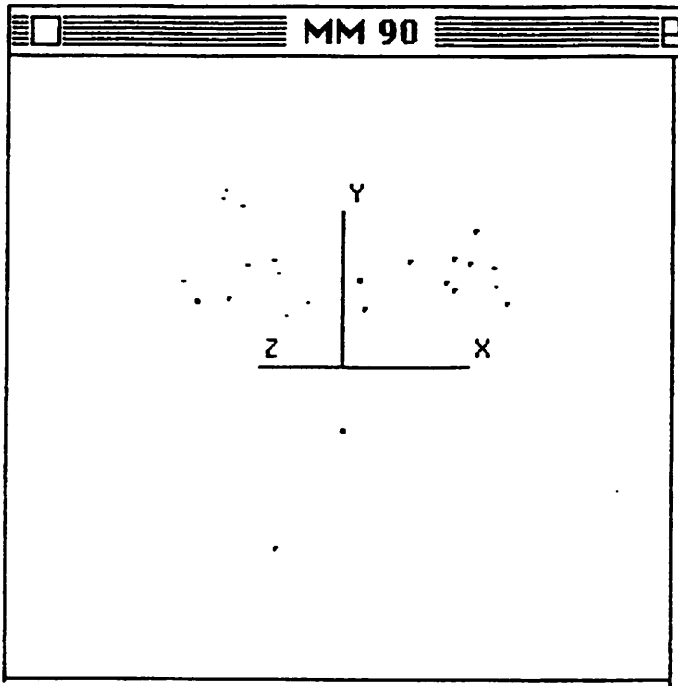


Figure 4 (part 2). 2-D projections of 3-D plots for cloud seeding data.

Figure 3 shows residual plots for various efficiencies of the MM estimate, starting with the PROGRESS approximation to the LMS solution and then an approximation to the LMS estimate given by my algorithm with 100 least squares fits and ending with the LS residual plot. The PROGRESS solution is based on 3549 least squares fits to groups of seven randomly selected points from the data; it produces a median residual of .0112. My algorithm, using just 100 least squares fits produces a median squared residual of .0024 and vastly different parameter estimates resulting in a very different residual plot. Although my algorithm produced an estimate with a smaller median square residual, the location of the exact value of the LMS estimate was not computed. Even if it was computed, experience tells us that is might not provide useful information.

The residuals that shift the most from plot to plot in figure 3 (1, 6, and 23) are labelled. All the plots show point six as an outlier. The behavior of point 23 is very interesting. In the PROGRESS solution it has zero residual while my approximation to the LMS estimate shows it as an outlier. The point moves a fair amount as the efficiency increases, but the residual remains small compared to its least squares residual even with the efficiency of the MM estimate set to .99. This leads one to suspect that point 23 may be outlying in the least squares plot because of the influence of point 6, which is the only point downweighted by the MM estimate with 99% efficiency. In fact, if the sixth data point is deleted, the resulting least squares residual plot does not indicate that point 23 is an outlier. (See figure 5, where it is labeled point 22 because point 6 has been deleted.) In the same figure the residual plot for the MM estimator with 95% efficiency with point 6 present but not graphed is presented. It is very similar to the plot of the least squares residuals when point 6 is deleted. In fact the parameter estimates are quite close; (1.56, .282, -.005, -.276, -.016, .539, .292) for the MM estimate with 95% efficiency for the entire data set and (1.59, .226, -.006, -.251, -.015, .530, .249) for the least squares estimate with point six deleted.

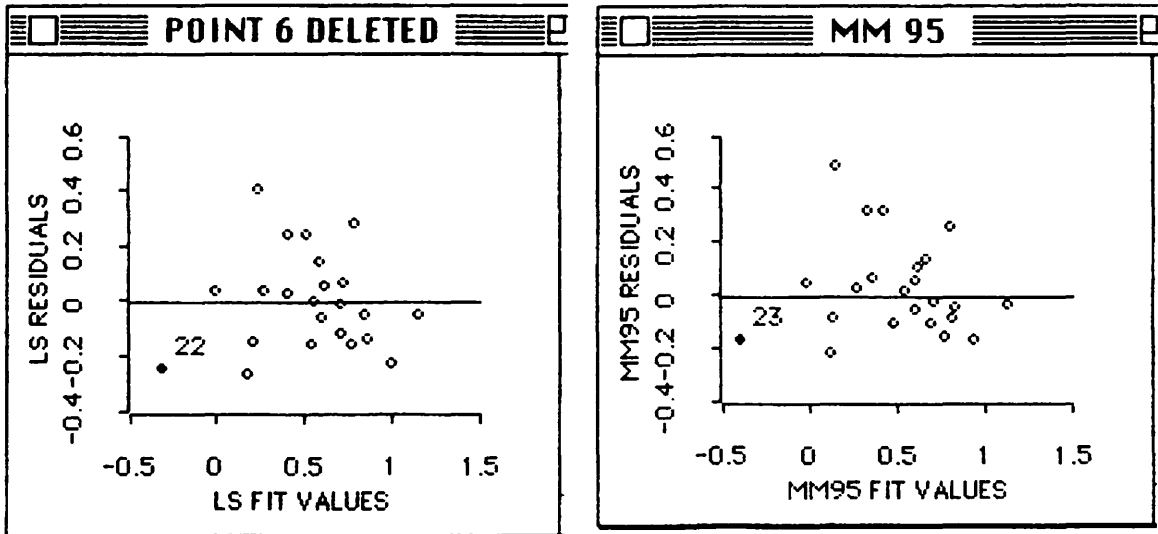


Figure 5. LS residual plot for cloud seeding data with point 6 deleted and MM 95 residual plot for entire data set with point 6 not graphed.

Next consider the 3-dimensional residual plots that are used by Cook and Weisberg to show the need for an interaction term in the model. The corresponding MM plots are presented in figure 4. Note, as Cook and Weisberg do, that the LMS estimate does not indicate clearly the need for the interaction term. Using my algorithm produces a smaller median squared residual but still no indication of the need for an interaction term. In general, as the efficiency increases, the plots increasingly indicate the need for the interaction term. It is important to understand if the outlying point 6 is causing the need for the interaction term. Point 6 does not influence any of the MM-estimates, and thus one would not expect the outlier to be causing the need for an interaction term. Figure 6 shows the 3-D least squares residual plot for the data set with point 6 deleted. As expected, the need for the interaction term is still evident. Figure 6 also depicts the 95% efficient MM estimate for the entire data set rescaled without point 6 plotted. Note the similarities in the plots.

At this point we consider the addition of the interaction terms involving the action covariate. The least squares estimates, their standard errors, and the p-values for testing each parameter equal to zero are presented in table 7. It appears that the action  $\times$  suitability interaction is important. The fact that the parameter estimate for that interaction term is

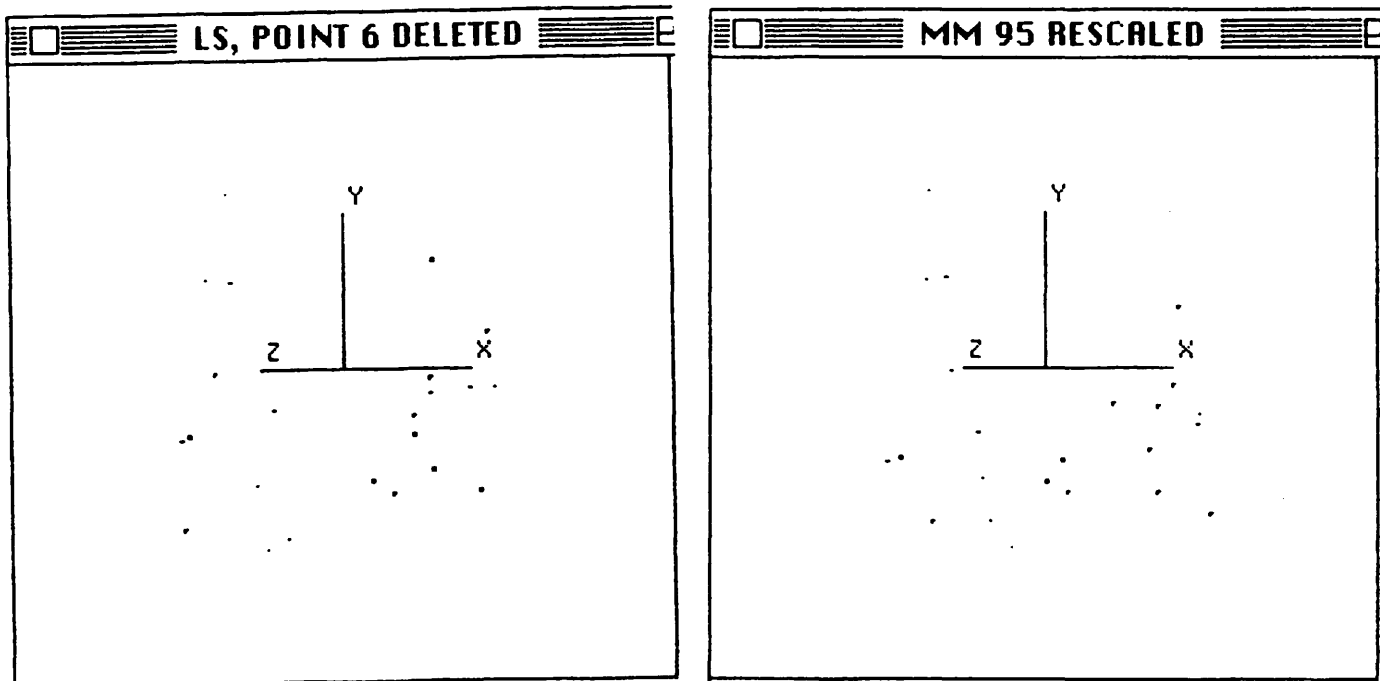


Figure 6. 2-D projections of 3-D plots for LS with point 6 deleted and MM 95 for entire data set with point 6 not graphed.

Table 7

Least Squares Regression Summary for full Model

<u>Variable</u>	<u>Estimate</u>	<u>Standard Error</u>	<u>p-value</u>
Intercept	-.2821	.4872	.5724
A	2.2120	.7994	.0160
D	-.0088	.0032	.0171
S	.1269	.1104	.2708
C	.0257	.0272	.3623
P	.4417	.2603	.1136
E	.5734	.2526	.0434
SA	-.4329	.1670	.0223
CA	-.0425	.0296	.1743
PA	-.1465	.4082	.7255
EA	-.2854	.3463	.4247

negative means that as suitability increasing the change in rainfall when seeding occurs decreases. This counterintuitive result could be due to large residuals at points 6 and 23. We can use the MM-estimate to reduce the affect of the outliers. Note that there are 24 data points and 11 parameters. The LMS estimate is attempting to minimize the 13th largest

residual. Since an exact fit can be found to 11 data points, the LMS estimate is likely to (and does) have a very small median residual. The estimate of scale for the MM-estimate, which is based on the LMS estimate, will therefore be very small. The small scale estimate means that even for very high efficiencies the MM-estimate will remain at the LMS-estimate and ignore half the data. We can still obtain useful information from the MM-estimate by setting the efficiency to be very large. In effect, we can compensate for the small scale estimate by inflating the constant used to determine the efficiency of the MM-estimate. The estimates for several values of  $k$  are given in table 8. Note that for  $k=50$ , the MM-estimate remains close to the LMS estimate. For  $k=400$ , the MM-estimate is very close to the least square estimate. For the lower efficiencies, many point are downweighted when compared to the least squares estimate where no points are downweighted. Thus as  $k$  increases, fewer and fewer points are downweighted. A logical choice for  $k$  would be one that downweights points 6 and 23 but no others. By this criterion,  $k=150$  would be a reasonable choice for  $k$ . It is interesting to note that all the estimates yield a negative coefficient for the  $A \times S$  variable. It is reasonable to conclude that the effect of the outliers is not causing the negative  $A \times S$  least squares coefficient. It is also important to note that the MM-estimate with  $k=150$  yields very similar estimates and standard errors to the least squares estimate with points 6 and 23 deleted. (See table 9.)

Table 8

MM Estimates for inflated efficiencies

Parameter	LMS Est.	$k=50$	$k=100$	$k=150$	$k=200$	$k=300$	$k=400$	LS Est.
Intercept	-1.5657	-1.5921	-.7944	.5116	-.0319	-.2627	-.2821	-.2821
A	3.9858	3.1905	2.5607	1.3423	1.9178	2.1788	2.212	2.212
D	.0025	-.0044	-.0057	-.0067	-.0076	-.0085	-.0088	-.0088
S	.2805	.3497	.2125	-.0288	.0830	.1254	.1269	.1269
C	.0942	.0970	.0583	.0223	.0289	.0281	.0257	.0257
P	-.1065	-.1159	.1301	.4552	.3893	.4115	.4417	.4417
E	.4306	.3701	.4158	.3485	.4455	.5369	.5734	.5734
SA	-.6087	-.6261	-.5114	-.2806	-.3909	-.4320	-.4329	-.4329
CA	.1245	-.1098	-.0728	-.0379	-.0451	-.0447	-.0425	-.0425
PA	1.0598	.4789	.2190	-.1207	-.0713	-.1092	-.1465	-.1465
EA	-.1704	-.0251	-.0937	-.0376	-.1442	-.2447	-.2854	-.2854

Table 9  
MM-estimates (k=150) and Standard Errors for full data set  
and

Least Squares estimates and Standard Errors for data set with points 6 and 23 deleted

<u>Parameter</u>	<u>MM Estimates</u>	<u>LS Estimates</u>
intercept	.5116 (.2988)	.4297 (.3670)
A	1.3423 (.4902)	1.3958 (.4678)
D	-.0067 (.0020)	-.0059 (.0015)
S	-.0288 (.0677)	-.0017 (.0778)
C	.0223 (.0167)	.0307 (.0138)
P	.4552 (.1597)	.3447 (.1340)
E	.3485 (.1571)	.2627 (.1243)
SA	-.2806 (.1024)	-.3089 (.0959)
CA	-.0379 (.0182)	-.0458 (.0148)
PA	-.1207 (.2503)	.0045 (.1954)
EA	-.0376 (.2123)	.0567 (.1637)

### Enzyme Kinetics

Table 10 presents data analyzed in Ruppert, Cressie and Carroll (1989). The data points are numbered column by column with the point being point 0. The data is from an enzyme kinetics study where substrate concentration (S) and inhibitor concentration (I) are independent variables and initial velocity ( $v_o$ ) is the response variable. As a model they consider a special case of the Michaelis-Menten model given by:

$$v_o = \frac{VS}{K(I) + S} \quad (5.1)$$

where V is the maximum velocity which is assumed to be independent of I and the Michaelis parameter (K) is allowed to depend on I.

Table 10  
Enzyme Kinetics Data from Becton Dickenson

<u>Substrate</u>	<u>Inhibitor Concentration</u>			
	<u>0</u>	<u>3</u>	<u>10</u>	<u>30</u>
25	.0328	.0153	.0087	.0039
50	.0510	.0327	.0146	
100	.0697	.0536	.0231	.0094
200	.0934	.0716	.0305	.0175
400	.0924	.0904	.0658	.0398



Note that if  $V$  depends on  $I$ , then (5.1) would be:

$$v_o = \frac{V(I)S}{K(I) + S}$$

which implies

$$\frac{1}{v_o} = \frac{K(I) + S}{V(I)S} = \frac{1}{V(I)} + \left(\frac{K(I)}{V(I)}\right)\left(\frac{1}{S}\right).$$

Thus  $\frac{1}{V(I)}$  and  $\frac{K(I)}{V(I)}$  can be estimated using the simple linear regression model. For fixed  $I$ , a Lineweaver-Burk plot includes the data points  $(\frac{1}{v_o(I)}, \frac{1}{S})$ , and the least squares fit to those points. It is common in practice to check the assumption that  $V$  is independent of  $I$  by superimposing Lineweaver-Burk plots for each inhibitor concentration. A common intercept is evidence for a common  $V$ . Ruppert, Cressie and Carroll point out that the apparently high intercept when  $I = 30$  (figure 7) is due to a high leverage point. The MM estimate with 95% efficiency (MM95, in figure 7) for  $I = 30$  ignores the high leverage point and yields an intercept of only .005. For the other inhibitor concentrations, the MM-estimates are very close to the least squares estimates. An investigator who computed the MM95 fit as well as the least squares fit would feel more comfortable with the common  $V$  assumption.

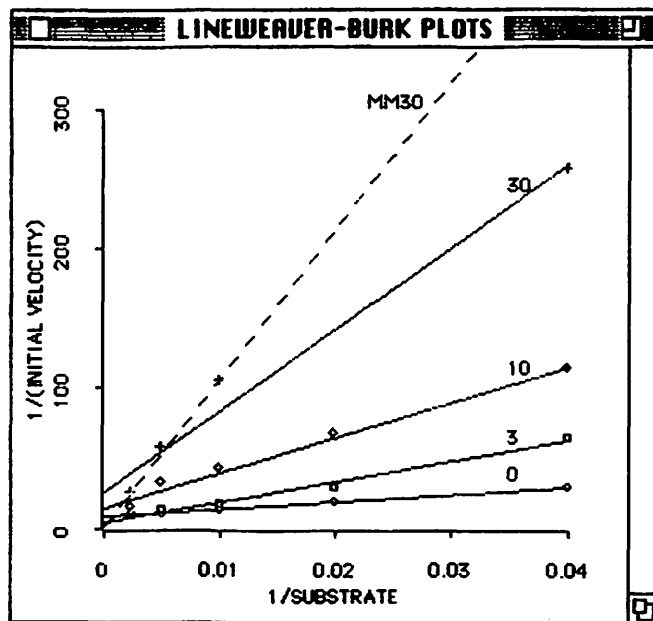


Figure 7. Lineweaver-Burk plots for each inhibitor concentration for enzyme kinetic data. (MM 95 estimate is dashed line.)

Fitting model (5.1) using least squares yields the residual plot given in figure 8. The Points 4, 13 and 14 might be considered outliers, but their status is uncertain because of the evident heteroscedasticity. Note that none of these points appear to be outliers in the Lineweaver-Burk plots in figure 7. The LMS residual plot suggests that points 4, 13 and 18 may be outliers, but since only an approximation to the LMS estimate is used, experience tells us that the true LMS residual plot may be quite different. The residual plot for MM-estimate with 95% efficiency shows the heteroscedasticity in the data set and also shows that point 14, which is somewhat outlying in the least square residual plot, should probably not be considered as an outlier. Points 4 and 13 are more clearly outlying in the MM 95 residual plot. High breakdown estimates have been criticized (Cook and Weisberg, 1989) for finding more outliers than are actually present in the data, and thus it is interesting that this is a case where the high breakdown MM-estimate finds fewer outliers than the least squares estimate.

The biochemist who provided this data set warned that deviations from the proposed model are common. The least squares plot indicates no such deviations and Ruppert, Cressie and Carroll find no evidence of a lack of fit in the proposed model. The MM-estimate residual plot suggests that perhaps some lack of fit exists, but no evidence for lack of fit is found in a plot of the residuals versus substrate concentrations. It seems that the possibility of lack of fit suggested by the MM-estimate residual plot is merely due to heteroskedasticity.

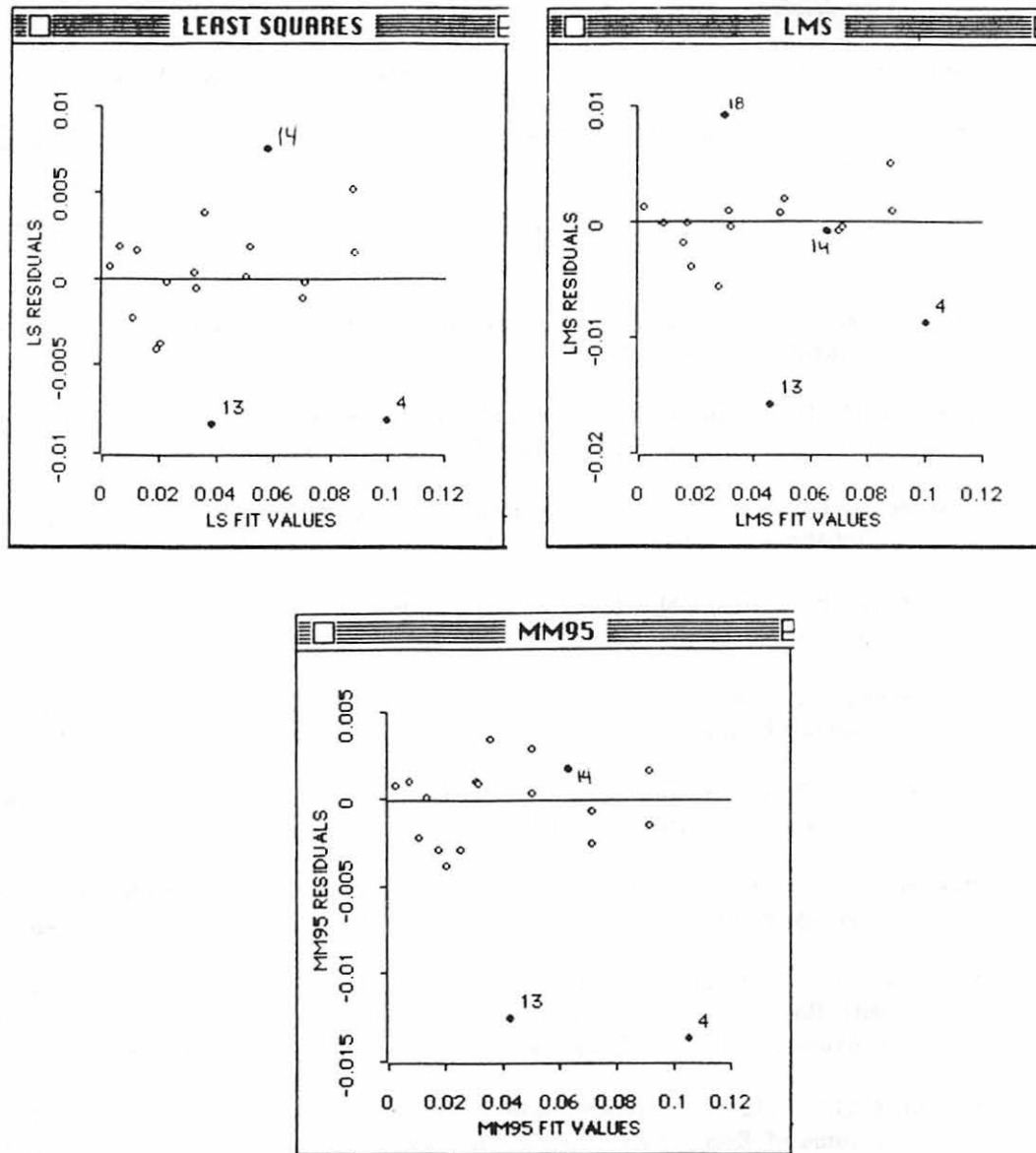


Figure 8. Residual plots for enzyme kinetic data. (LS, LMS, and MM95 estimates)

#### REFERENCES

- Atkinson, A. C. (1986). "Masking Unmasked," *Biometrika*, 73: 533-542.
- Cook, R.D. and S. Weisberg. (1982). *Residuals and Influence in Regression*. London: Chapman-Hall.
- Cook, R. D. and S. Weisberg (1989). "Regression Diagnostics: Robust *versus* Least Squares Residuals," Technical Report #538. University of Minnesota, School of Statistics.
- Dennis, J. E. and D.J. Woods (1986). "The Nelder-Mead Simplex Algorithm," *Optimization*

on *Microcomputers*, 116-122.

- Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley and Sons.
- Kim, J and D. Pollard (1990). "Cube Root Asymptotics," *The Annals of Statistics*, 18, 191-219.
- Morganthaler, S. (1989). "Comment on Yohai and Zamar," *Journal of the American Statistical Association*, 84, 636.
- Nelder, J.E. and R. Mead (1965). "A Simplex Method for Function Minimazation," *The Computer Journal*, 7:308.
- Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1986). *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press
- Rousseeuw, P.J. (1984). "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79: 871-880.
- Rousseeuw, P.J. and A.M. Leroy (1987). *Robust Regression and Outlier Detection*. New York: John Wiley and Sons.
- Rousseeuw, P.J., and B.C. van Zomeren (1990). "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association* (forthcoming).
- Ruppert, D., Cressie, N, and R.J. Carroll (1989). "A Transformation/Weighting Model for Estimating Michaelis-Menten Parameters," *Biometrics*, 45: 637-656.
- Ruppert, D. and D.G. Simpson (1990). Discussion of "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association* (forthcoming).
- Simpson, D.G., D. Ruppert, and R.J. Carroll (1989). "One-step GM-estimates for Regression with Bounded Influence and High Breakdown-point," Technical report #859. Cornell University, School of Operations Research and Industrial Engineering.
- Simkin, C.G.F. (1978). "Hyperinflation and Nationalist China," *Stability and Inflation, A Volume of Essays to Honor the Memory of A. W. H. Phillips*. New York: John Wiley and Sons.
- Steele, J.M. and W.L. Steiger (1986). "Algorithms and Complexity for Least Median of Squares Regression," *Discrete Applied Mathematics* 14:93-100.
- Stromberg, A.J. (1989). "Robust Efficient Estimation of Nonlinear Regression Parameters," Ph.D. dissertation. University of North Carolina, Chapel Hill.
- Stromberg, A.J. and D. Ruppert (1989). "Breakdown in Nonlinear Regression," Technical Report #539. University of Minnesota, School of Statistics.
- Tierney, L. (1988). "XLISP-STAT A Statistical Environment Based on the XLISP Language," Technical Report #528. University of Minnesota, School of Statistics.
- Yohai, V.J. (1987). "High Breakdown Point and High Efficiency Robust Estimates for Regression," *Annals of Statistics*, 15(2):642-656.