NBER WORKING PAPER SERIES

# A MONTE CARLO STUDY OF TWO ROBUST ALTERNATIVES TO LEAST SQUARES REGRESSION ESTIMATION

Richard W. Hill*
Paul W. Holland*

Working Paper No. 58

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

September 1974



* NBER Computer Research Center. Research supported in part by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research, Inc.

## Abstract

We give some Monte Carlo results on the performance of two robust alternatives to least squares regression estimation -- least absolute residuals and the one-step "sine" estimator. We show how to scale the residuals for the sine estimator to achieve constant efficiency at the Gaussian across various choices of X-matrix and give some results for the contaminated Gaussian distribution.

## Contents



## Figures



## Tables

## 1. Introduction and Notation

In this paper we discuss some preliminary results on the efficiency of two regression methods that have been suggested as robust alternatives to least squares. We assume the usual linear model,

$$\underset{\sim}{y} = \underset{\sim}{X}\beta + \underset{\sim}{e} \ , \tag{1-1}$$

where $\underset{\sim}{y}$ and $\underset{\sim}{e}$ are N×1 vectors, $\underset{\sim}{X}$ is N×p and $\beta$ is p×1. In our Monte Carlo study, $\underset{\sim}{e}$ is a random vector with independent and identically distributed coordinates from a density f(·), where f is either the Gaussian or one of a selected family of scale contaminated Gaussian densities. Our main interest is in the effect of f and $\underset{\sim}{X}$ on the behavior of estimators of $\beta$. We give some Monte Carlo results on these effects in sections 2 and 3 and discuss an implication they have in section 4.

Notation: We denote the ordinary least squares estimator of $\beta$ by $\hat{\beta}_{LS}$ or LS. The two alternatives to LS which we studied are denoted by LAR (or $\hat{\beta}_L$) and SIN1 (or $\hat{\beta}_S$).

LAR is the "least absolute residual" estimator so that $\hat{\beta}_L$ minimizes

$$\sum_i \left| y - \sum_j x_{ij} \beta_j \right| \ . \tag{1-2}$$

LAR has been studied extensively and goes by a variety of other names -- least lines, minimum absolute deviations (MAD), minimum sum of absolute errors (MSAE), least absolute deviations (LAD) and the

minimum 1-norm estimator. The estimator SIN1 is the "1-step SIN estimator starting from LAR" so that $\hat{\beta}_S$ is given by

$$\hat{\beta}_S = (X^T \langle w \rangle X)^{-1} X^T \langle w \rangle y \tag{1-3}$$

where $\langle w \rangle$ is a diagonal matrix whose diagonal entries are given by

$$w_i = \begin{cases} 1 & \text{if } r_i = 0 \\ \dfrac{\sin(r_i/\sigma)}{r_i/\sigma} & \text{if } 0 < r_i \leq \pi\sigma \\ 0 & \text{if } r_i > \pi\sigma \ . \end{cases} \tag{1-4}$$

In (1-4), $r_i$ is the $i^{\text{th}}$ residual from the LAR fit, i.e.

$$r = y - X\hat{\beta}_L \ , \tag{1-5}$$

and $\sigma$ is given by

$$\sigma = 2.1 \text{ median } \{|r_i|\} \ . \tag{1-6}$$

The SIN1 estimator is based on Andrews' "sine" estimator of location (AMT) from Andrews et al. [1972]. SIN1, as used here, consists of one step of iteratively reweighted least squares starting at LAR. The weights are based on the LAR residuals and the particular choice of scale given in (1-6) is adapted from the one used in Andrews et al. [1972] for the location problem. One result of our study is a modification of (1-6) that appears to be a modest improvement. We view SIN1 as a simple way to improve the LAR estimator and our simulation results indicate that it usually is.

In order to compare LAR and SIN1 with each other and with LS across various choices of X and f we use the following overall measure of the inefficiency of $\hat{\beta}$ relative to $\hat{\beta}_{LS}$ :

$$\text{iff}(\hat{\beta}) = E\left\| \hat{\beta} - \beta \right\|^2 / E\left\| \hat{\beta}_{LS} - \beta \right\|^2 \ . \tag{1-7}$$

For an unbiased estimator (1-7) becomes:

$$\text{iff}(\hat{\beta}) = \sum_j \text{Var}(\hat{\beta}_j) \; / \; \sum_j \text{Var}(\hat{\beta}_{jLS}) \; . \qquad (1\text{-}8)$$

When the number of regressor variables, p, equals 1 $\text{iff}(\hat{\beta})$ reduces to the reciprocal if the relative efficiency of $\hat{\beta}$ to $\hat{\beta}_{LS}$. The smaller the value of $\text{iff}(\hat{\beta})$, the better $\hat{\beta}$ is relative to LS.

The X-matrix: All of the X-matrices used in our Monte Carlo study are derived in various ways from a basic 20×6 matrix, VDATA1. First we describe the construction of VDATA1 and then indicate how the X-matrices used is the study are derived from it.

The 6 columns of VDATA1 were devided into 3 groups of 2. Columns 1 and 2 were chosen so that their scatter plot forms a perfect square centered about the origin. Thus the first two columns correspond to variables like those in a designed experiment. Columns 3 and 4 were chosen to be roughly independent bivariate Gaussian. Their scatter plot is given in Figure 1-1.

Figure 1-1 goes about here

Columns 5 and 6 were chosen to be roughly independent bivariate variables with outliers. Two independent Cauchy samples of size 20 were drawn and then the largest observations in each sample was moved in until they contributed 80 and 85% to the total sum of squares of their columns, respectively. The scatter plot of columns 5 and 6 is given in Figure 1-2.

Figure 1-2 goes about here

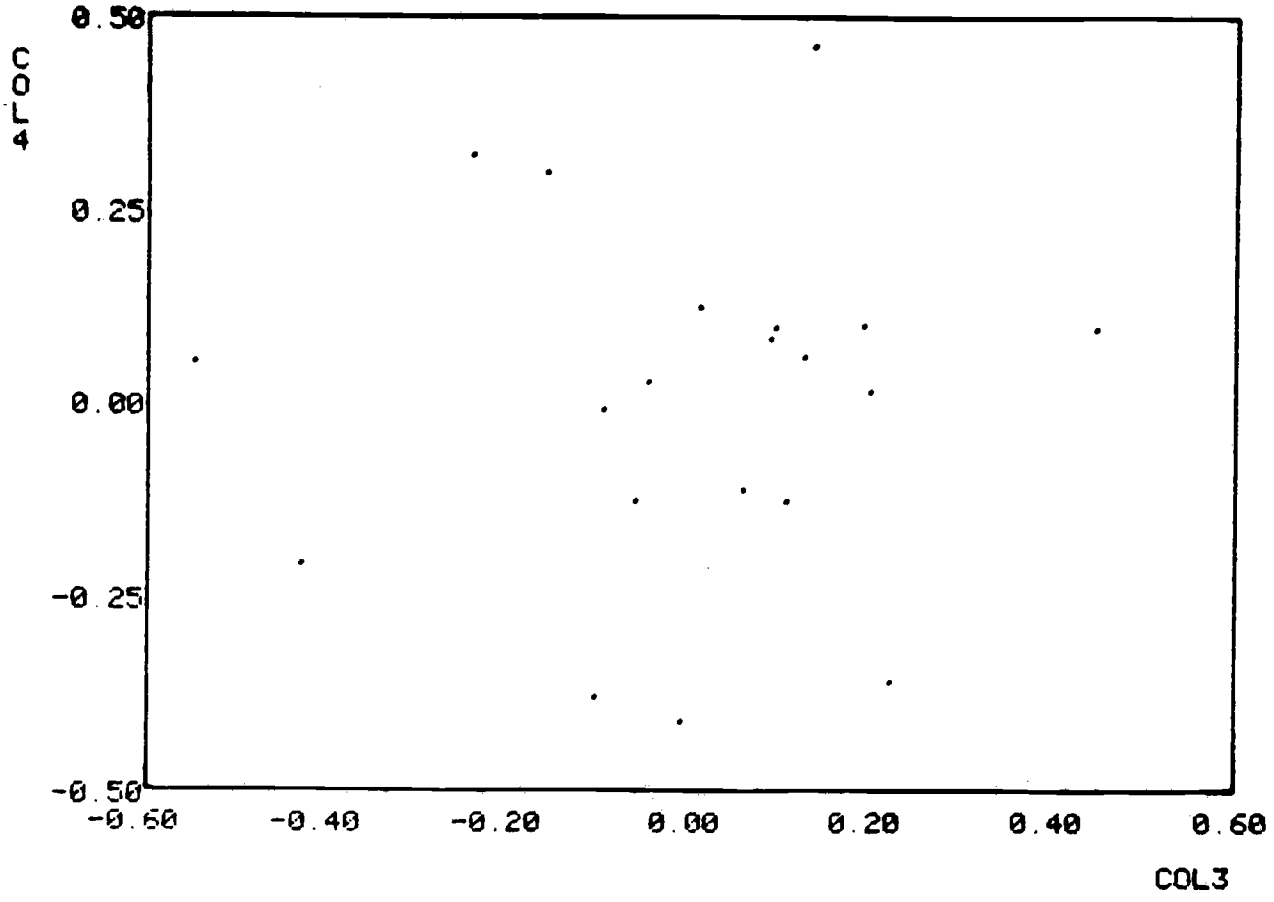


FIGURE 1-1. SCATTER PLOT OF COL3 AND COL4 FOR VDATA1

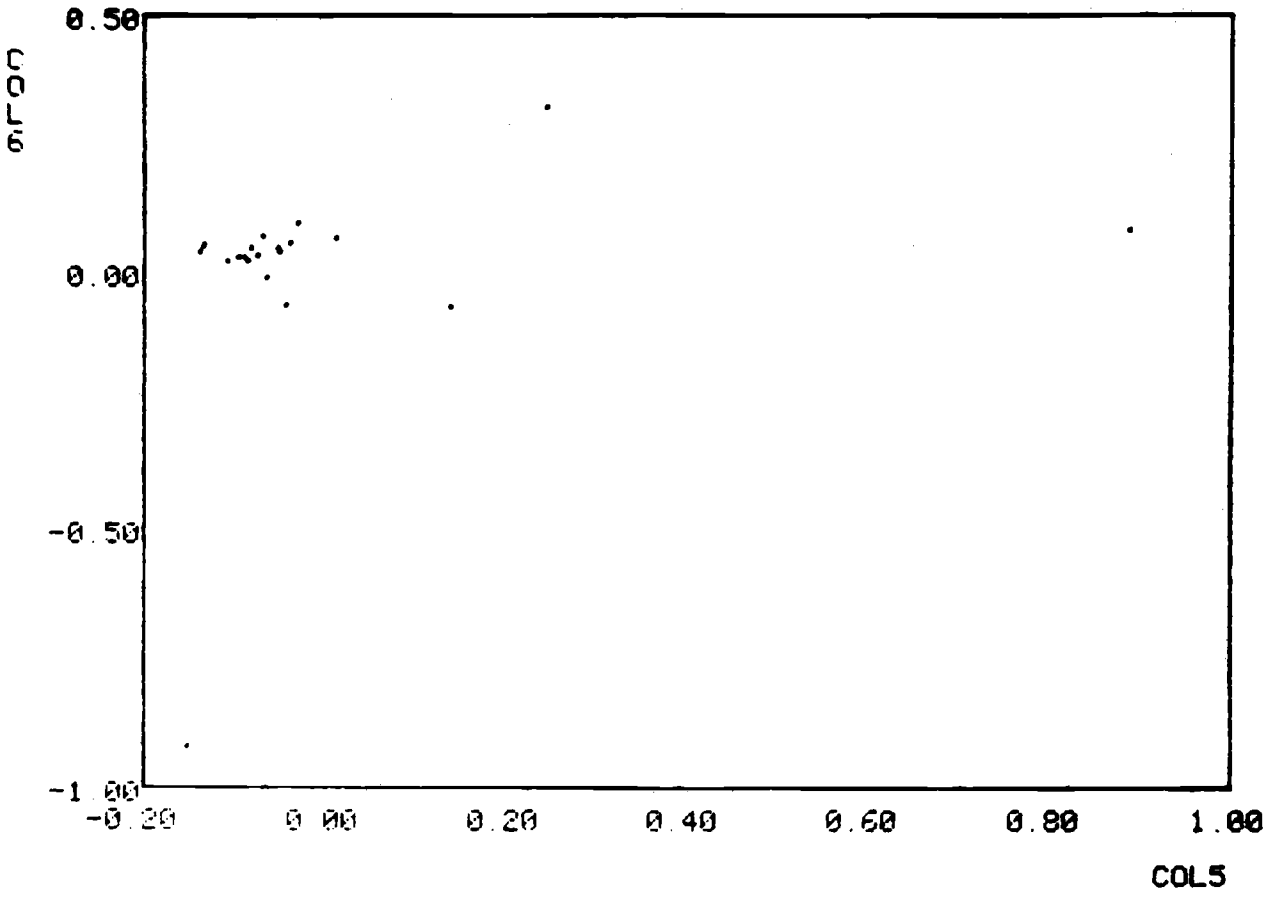


FIGURE 1-2. SCATTER PLOT OF COL5 AND COL6 FOR VDATA1

After the 6 columns of VDATA1 were selected each column was standardized to have mean zero and unit sum of squares, i.e. $(\text{VDATA1})^T\text{VDATA1}$ is a correlation matrix. In Table 1-1 we give the final, standardized columns of VDATA1. In Table 1-2 we give the correlation matrix of the columns of VDATA1. From Table 1-2 we see that the six columns of VDATA1 are all roughly orthogonal. The eigenvalues of $(\text{VDATA1})^T\text{VDATA1}$ are given in Table 1-3. The ratio of the largest to the smallest eigenvalue (the condition number of $(\text{VDAT1})^T\text{VDATA1}$ ) is 7.499. The two outliers in columns 5 and 6 occur in rows 20 and 18 respectively of Table 1-1.

Tables 1-1, 1-2 and 1-3 go about here

The various $\underset{\sim}{X}$-matrices we used were derived from VDATA1 in two basic ways. Most of our results concern N=20 and $p \leq 6$. To vary p, we formed an $\underset{\sim}{X}$-matrix from the first p columns of VDATA1. Thus as we vary the dimension of $\underset{\sim}{X}$ we are also varying the kurtosis in the columns of $\underset{\sim}{X}$ -- this is important to remember in section 3. A few results are also given for N=40. We wanted to have $40 \times p$ $\underset{\sim}{X}$-matrices that were "similar to" the $20 \times p$ ones we used. To do this we merely replicated every row of VDATA1. This has the effect of producing a new $40 \times 6$ matrix whose singular value decomposition is identical to that of VDATA1 except for the "basis" or U-matrix. The condition number and the pattern of eigenvalues are unchanged.

Error Distributions:

After a preliminary examination of various error distributions, $f(\cdot)$, we decided to use the simple 2-parameter family of a mixture of a N(0,1)

TABLE 1-1

FINAL, STANDARDIZED COLUMNS OF VDATA1

| ROW | COL1 | COL2 | COL3 | COL4 | COL5 | COL6 |
|---|---|---|---|---|---|---|
| 1 | 0.2712 | 0.2712 | -0.0453 | 0.0257 | -0.0880 | 0.0288 |
| 2 | 0.2712 | 0.1627 | 0.1092 | -0.1268 | -0.0509 | 0.0470 |
| 3 | 0.2712 | 0.0542 | 0.4513 | 0.0963 | 0.0140 | 0.0682 |
| 4 | 0.2712 | -0.0542 | -0.1605 | 0.2977 | -0.1065 | 0.0225 |
| 5 | 0.2712 | -0.1627 | 0.2242 | -0.3618 | 0.2463 | 0.3193 |
| 6 | 0.2712 | -0.2712 | 0.0107 | 0.1246 | -0.0814 | 0.0461 |
| 7 | 0.1627 | -0.2712 | 0.1937 | 0.1006 | -0.0373 | 0.0583 |
| 8 | 0.0542 | -0.2712 | -0.2435 | 0.3205 | -0.1373 | 0.0404 |
| 9 | -0.0542 | -0.2712 | -0.0094 | -0.4123 | -0.0852 | 0.0228 |
| 10 | -0.1627 | -0.2712 | 0.1382 | 0.4631 | -0.0630 | -0.0112 |
| 11 | -0.2712 | -0.2712 | 0.0956 | 0.0984 | -0.0489 | 0.0388 |
| 12 | -0.2712 | -0.1627 | 0.0597 | -0.1136 | -0.0732 | 0.0327 |
| 13 | -0.2712 | -0.0542 | -0.0613 | -0.1263 | -0.0944 | 0.0303 |
| 14 | -0.2712 | 0.0542 | 0.1282 | 0.0598 | -0.0680 | 0.0691 |
| 15 | -0.2712 | 0.1627 | -0.0966 | -0.0085 | 0.1387 | -0.0672 |
| 16 | -0.2712 | 0.2712 | -0.1060 | -0.3819 | -0.1340 | 0.0559 |
| 17 | -0.1627 | 0.2712 | 0.2013 | 0.0145 | -0.0290 | 0.0966 |
| 18 | -0.0542 | 0.2712 | -0.4324 | -0.2083 | -0.1520 | -0.9198 |
| 19 | 0.0542 | 0.2712 | 0.0914 | 0.0840 | -0.0417 | -0.0620 |
| 20 | 0.1627 | 0.2712 | -0.5486 | 0.0544 | 0.8917 | 0.0833 |

TABLE 1-2

MATRIX OF INTERCORRELATIONS FOR THE COLUMNS OF VDATA1

| ROW | COL1 | COL2 | COL3 | COL4 | COL5 | COL6 |
|---|---|---|---|---|---|---|
| 1 | 1. | 0. | 0.0573 | 0.1462 | 0.2150 | 0.1576 |
| 2 | 0. | 1. | -0.2786 | -0.2456 | 0.2382 | -0.3034 |
| 3 | 0.0573 | -0.2786 | 1. | 0.0404 | -0.3499 | 0.4818 |
| 4 | 0.1462 | -0.2456 | 0.0404 | 1. | -0.0297 | 0.0744 |
| 5 | 0.2150 | 0.2382 | -0.3499 | -0.0297 | 1. | 0.2448 |
| 6 | 0.1576 | -0.3034 | 0.4818 | 0.0744 | 0.2448 | 1. |

Matrix of intercorrelations for the columns of VDATA1 (i.e.$(\text{VDATA1})^{T}$ VDATA1).

TABLE 1-3

EIGENVALUES OF $(VDATA1)^T$ VDATA1

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0.244 | 0.624 | 0.821 | 1.068 | 1.411 | 1.832 |

density with a $N(0,k^2)$ density. This is given by

$$g(u) = (2\pi)^{-\frac{1}{2}} \left\{ (1-\alpha)\, e^{-u^2/2} + \frac{\alpha}{k} e^{-1/2 \left(\frac{u}{k}\right)^2} \right\}, \quad (1\text{-}9)$$

where $0 \leq \alpha < 1$, and $1 < k < \infty$. These are denoted CGk$\alpha$ so that CG3.01 indicates a choice of $g$ where $k=3$ and $\alpha = .01$. In all cases the scale of $f(u)$ was selected so that the errors had unit variance. Thus

$$f(u) = \frac{1}{\tau} g\left(\frac{u}{\tau}\right) \quad (1\text{-}90)$$

where

$$\tau^2 = 1-\alpha + \alpha k^2 \quad (1\text{-}11) .$$

Both of the estimators LAR and SIN1 are <u>regression invariant</u> in the sense that if the vector of observed values $\underset{\sim}{y}$ is transformed to $\underset{\sim}{y} + \underset{\sim}{X}\,\underset{\sim}{\beta}^{(0)}$ for some $\underset{\sim}{\beta}^{(0)}$ then $\hat{\underset{\sim}{\beta}}$ is transformed to $\hat{\underset{\sim}{\beta}} + \underset{\sim}{\beta}^{(0)}$. This invariance implies that

$$E_{\underset{\sim}{\beta}} ||\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}||^2 = E_{\underset{\sim}{0}} ||\hat{\underset{\sim}{\beta}}||^2 \quad (1\text{-}12)$$

for any error distribution. Hence all of our Monte Carlo results were computed with the true value of $\underset{\sim}{\beta}$ set equal to $\underset{\sim}{0}$.

2. Creeping Inefficiency in the Gaussian Case

One difference between LAR and SIN1 is the dependence of the latter on the choice of a residual scaling factor, $\sigma$ . Ignoring for the moment the particular form of $\sigma$ in (1-6) it is clear that if $\sigma$ in (1-4) is very large then the $w_i$ are all near unity and $\hat{\beta}_S$ will be near the LS estimate. On the other hand, if $\sigma$ is very small, then the $w_i$ are all zero causing $\hat{\beta}_S$ to be undefined. LAR like LS does not depend on a choice of scale for the residuals. The motivation for the choice of residual scaling given in (1-6) is its performance in the location case (p=1) where Andrews et al. [1972] found it to be an effective improvement on the scale estimate based on the interquartile range. In the location case, (1-6) gives $\sigma$ as a multiple of the median absolute deviation from the median and this is the scale analogue of the median -- see Hample [1973]. One can motivate (1-6) as follows. If the initial LAR fit is reasonably good, then the absolute residuals $\{|r_i|\}$ will behave approximately like a sample from the half-Gaussian distribution. Matching the median of this sample to that of a half-Gaussian population leads to a scale estimate that is a multiple of the one given in (1-6). The choice of the multiplying factor depends on the weighting function (1-4) and a desired level of inefficiency at the Gaussian -- in the location case Andrews used a factor of 2.1 and achieved an inefficiency of 1.07 at the Gaussian.

Some Monte Carlo Results: We estimated the covariance matrix of $\hat{\beta}_L$ and $\hat{\beta}_S$ in the Gaussian case using the Monte Carlo "swindle" described in Holland [1973], and 500 replications. In all cases,

N=20 and we varied p, the dimension of the X-matrix, over the range 1,2,3,4,6 by successively including the columns of VDATA1 as described in section 1. Table 2-1 gives the resulting estimated inefficiences for LAR and SIN1 as a function of p.

Table 2-1 goes about here

The estimated inefficiency of SIN1 for p=1 is 1.10 which compares favorably with the value of 1.07 found in the Princeton Robustness Study. When p=3 the value of 1.17 for SIN1 is close to the value 1.16 found by Andrews [1973] using a different X-matrix.

In Table 2-1 we see that SIN1 exhibits creeping inefficiency in the sense that what is a highly efficient estimator for p=1, becomes progressively less efficient as p increases. While SIN1 is still better than LAR over this range of p, the systematic increase suggests that for p=7 SIN1 would be cease to be an improvement over LAR. On the other hand, LAR does not exhibit this phenomenon. Instead, LAR becomes more inefficient as the effective sample size (N-p) increases. This is analogous to the behavior of the median which approaches its asymptotic variance from below as $N \to \infty$. (See Kendall and Stuart [1963], page 367).

It is evident that the creeping inefficiency of SIN1 can be avoided by adjusting the residual scaling factor, $\sigma$, so that it depends on p -- in particular, it should increase as p increases to make $\hat{\beta}_S$ more closely approximate $\hat{\beta}_{LS}$ in the Gaussian case. This naturally suggests a simple degree of freedom type of correction of

TABLE 2-1

INEFFICIENCY OF LAR AND SIN1 IN GAUSSIAN CASE FOR N=20, p=1,2,3,4,6

| p | 1 | 2 | 3 | 4 | 6 |
|---|---|---|---|---|---|
| $iff(\hat{\beta}_L)$ | 1.63 | 1.63 | 1.57 | 1.54 | 1.47 |
| $iff(\hat{\beta}_S)$ | 1.10 | 1.12 | 1.17 | 1.23 | 1.32 |

the form

$$\sigma_p = \sqrt{\frac{N-1}{N-p}} \ (2.1 \text{ median } \{|r_i|\}). \qquad (2\text{-}1)$$

We tried using (2-1) instead of (1-6) and it helped, but did not eliminate the problem, nor did various similar adjustments to $\sigma$ using different multiplying factors.

In rethinking the motivation for using $\sigma$ in the location case we realized that one feature of the LAR residuals is that at least p of them must be identically zero. Hence the absolute residuals from LAR do not look like a sample from the half-Gaussian distribution. This suggests eliminating p-1 of these zero residuals and then computing $\sigma$ as before, i.e.

$$\sigma_p = 2.1 \text{ median } \{\text{largest } N-p+1 \text{ of the } |r_i|\}. \qquad (2\text{-}2)$$

Table 2-2 gives the Monte Carlo estimates of the inefficiency of SIN1 using (2-2) instead of (1-6) for p=2,4,6. It is clear that this correction has been successful and the SIN1 estimator now has an inefficiency at the Gaussian that is effectively independent of the dimension of the X-matrix.

Table 2-2 goes about here

The new residual scaling given by (2-2) does not seem to have seriously altered the behavior of SIN1 in the non-Gaussian case. For the CG5.25 distribution and for p=6, our Monte Carlo estimates of $\text{iff}(\hat{\beta}_S)$ for $\sigma$ given by (1-6) was .64 while for $\sigma_p$ given by (2-2) it was .63 . While it may be useful to make a more thorough comparison than this we did not and henceforth all

TABLE 2-2
INEFFICIENCY OF SIN1 UNDER GAUSSIAN ERRORS USING (2-2) AS THE DEFINITION OF $\sigma$, FOR p = 2,4,6

| p | 2 | 4 | 6 |
|---|---|---|---|
| $iff(\hat{\beta}_S)$ | 1.08 | 1.08 | 1.09 |

results given for SIN1 use (2-2) for the residual scaling. In the next section we give more results for the contaminated Gaussian case.

## 3. Monte Carlo Results for Contaminated Gaussian Errors

After some preliminary experimentation we decided to obtain Monte Carlo estimates of the inefficiency of LAR and SIN1 under CGk$\alpha$ errors for three values of $\alpha$ (.1,.25,.5), three values of k (3,5,10) and three choices of $\underset{\sim}{X}$-matrix (the first two, the first four and all six columns of VDATA1). The Monte Carlo estimates for these two estimators and twenty-seven situations are given in Table 3-1. These Monte Carlo estimates used simple experimental sampling rather than the contaminated-Gaussian version of the swindle described in Holland [1973] and are based on 500 replications each.

Table 3-1 goes about here

The most obvious message in Table 3-1 is that, except for five cases involving LAR, both LAR and SIN1 are improvements, sometimes substantial improvements, over LS. Furthermore, except for four cases where the percent of contamination is very high (50%) SIN1 does improve upon LAR.

In addition to these effects, there are at least 3 trends in Table 3-1 that should be mentioned. First, for each level of percent contamination there is a systematic improvement in both LAR and SIN1 over LS as the variance of the contaminating distribution increases. Thus the heavier the tails of the error distribution the worse LS is and therefore the more LAR and SIN1 can improve upon it. The second trend involves varying the $\underset{\sim}{X}$-matrix for each error distribution. For SIN1 there is a systematic increase in $\text{iff}(\hat{\beta})$ as the $\underset{\sim}{X}$-matrix includes more columns of VDATA1. This is also true for LAR except for the GG3.1 errors where the trend is ambiguous. While this looks something like the creeping inefficiency

TABLE 3-1

INEFFICIENCY OF LAR AND SIN1 WITH RESPECT TO LS

| k | | 3 | | 5 | | 10 | |
|---|---|---|---|---|---|---|---|
| $\alpha$ | p | LAR | SIN1 | LAR | SIN1 | LAR | SIN1 |
| | 2 | 1.09 | .78 | .60 | .44 | .19 | .13 |
| .10 | 4 | 1.03 | .80 | .59 | .47 | .20 | .15 |
| | 6 | 1.03 | .87 | .72 | .60 | .38 | .33 |
| | 2 | .87 | .71 | .42 | .39 | .16 | .15 |
| .25 | 4 | .91 | .78 | .49 | .47 | .26 | .25 |
| | 6 | .96 | .83 | .69 | .63 | .45 | .43 |
| | 2 | .91 | .81 | .58 | .63 | .35 | .46 |
| .50 | 4 | 1.00 | .87 | .74 | .75 | .55 | .62 |
| | 6 | 1.08 | .94 | .92 | .87 | .80 | .79 |

Inefficiency of LAR and SIN1 with respect to LS. Monte Carlo estimates (500 replications).

we found in the Gaussian case, we suspect that it is more due to the changing nature of the $\underset{\sim}{X}$-matrix than the increase in its dimension. The first two columns of VDATA1 form a balanced design matrix. The next two are still well behaved but are like a bivariate Gaussian sample. The last two columns have outliers (or high leverage points) and are not well behaved at all. The kurtosis of the six columns of VDATA1 are -1.65,-1.65,0.56,-0.30,9.76,11.53, respectively. Thus as we incorporate more columns of VDATA1 into the $\underset{\sim}{X}$-matrix we increase the probability that a large error will be associated with a high leverage point. A large error associated with a high leverage point in X decreases the improvement of SIN1 and LAR over LS.

A third trend in Table 3-1 concerns the improvement of SIN1 over LAR. This is more easily seen in Table 3-2 which gives the ratio, $\text{iff}(\hat{\beta}_L)/\text{iff}(\hat{\beta}_S)$. The biggest improvement occurs in the lowest contamination case and for the first two columns of VDATA1 as the $\underset{\sim}{X}$-matrix. In the very extreme case of CG10.5, SIN1 actually degrades the performance of LAR.

Table 3-2 goes about here

Since SIN1 is a simple 1-step improvement on LAR it might be hoped that more iterations would help it out. To test this we ran a auxiliary experiment for the CG10.10 case and changed the number of steps sequentially from 1 to 5. The value of the numerator of $\text{iff}(\hat{\underset{\sim}{\beta}}_S)$ (i.e. $E||\hat{\underset{\sim}{\beta}}_S-\underset{\sim}{\beta}||^2$) is given for each number of steps in Table 3-3. An examination of the individual $\underset{\sim}{\beta}$ estimates shows that they do change with more steps but as we see in Table 3-3 on the whole they do not improve.

Table 3-3 goes about here

As a final question we asked what is the effect of N? As described

TABLE 3-2

RATIO OF $iff(\hat{\beta}_S)$ TO $iff(\hat{\beta}_L)$ FROM TABLE 3-1

| k | | 3 | 5 | 10 |
|---|---|---|---|---|
| $\alpha$ | p | | | |
| | 2 | 1.40 | 1.36 | 1.46 |
| .10 | 4 | 1.29 | 1.26 | 1.33 |
| | 6 | 1.24 | 1.20 | 1.15 |
| | 2 | 1.23 | 1.08 | 1.07 |
| .25 | 4 | 1.17 | 1.04 | 1.04 |
| | 6 | 1.16 | 1.10 | 1.05 |
| | 2 | 1.12 | .92 | .76 |
| .50 | 4 | 1.15 | .99 | .89 |
| | 6 | 1.15 | 1.08 | 1.01 |

TABLE 3-3

MONTE CARLO ESTIMATES OF $E||\hat{\beta}_S - \beta||^2$

| | Number of Steps | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| $E||\hat{\beta}_S-\beta||^2$ | 2.95 | 2.97 | 2.97 | 2.97 | 2.98 |

Monte Carlo estimates of $E||\hat{\beta}_S - \beta||^2$ for 1 to 5 steps of iteratively reweighted least squares using (1-4) (227 replications) .

in section 1, we doubled the size of our $\underset{\sim}{X}$-matrices by replicating each row. The inefficiences for LAR and SIN1 for 3 choices of $\underset{\sim}{X}$-matrix with N=40 and two error distributions (CG3.5 and CG10.1) are given in Table 3-4. These values are systematically smaller than the corresponding ones in Table 3-1 indicating that robustness is easier to achieve when the sample size is larger.

Table 3-4 goes about here

## 4. Covariance Matrices for Robust Regression Estimators

The results of Table 3-1 have an interesting implication for a class of possible estimates of the covariance matrix for robust regression estimators. Analogues to least squares as well as asymptotic theory lead to a covariance matrix for $\hat{\underset{\sim}{\beta}}$ of the form

$$h^2(\underset{\sim}{X}^T\underset{\sim}{X})^{-1} \qquad (4\text{-}1)$$

where $h^2$ is a multiplying factor that potentially depends on these factors:

(1) The estimator -- for iteratively reweighted least squares estimators like SIN1 this would include the way the weights are formed.

(2) The error distribution.

(3) The method of scaling the residuals.

(4) The desired level of inefficiency achieved at some standard distribution (e.g. the Gaussian) -- in the case of SIN1 this is about 1.08 .

(5) The $\underset{\sim}{X}$-matrix.

The original hope was that all but (5) would matter except through N and p,but Table 3-1 shows that this is not the case. If (4-1) holds then

TABLE 3-4

INEFFICIENCIES OF LAR AND SIN1 FOR N=40

| | CG3.5 | | CG10.1 | |
|---|---|---|---|---|
| P | LAR | SIN1 | LAR | SIN1 |
| 2 | .79 | .72 | .18 | .12 |
| 4 | .82 | .75 | .18 | .12 |
| 6 | 1.00 | .86 | .25 | .17 |

Inefficiencies of LAR and SIN1 for N=40. Monte Carlo estimates (500 replications).

$$E||\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}||^2 = h^2 \, tr(\underset{\sim}{X}^T\underset{\sim}{X})^{-1} \, . \qquad (4\text{-}2)$$

But

$$E||\hat{\underset{\sim}{\beta}}_{LS}-\underset{\sim}{\beta}||^2 = Var(f) \, tr(\underset{\sim}{X}^T\underset{\sim}{X})^{-1} = tr(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}$$

since in all cases the errors were scaled to have Var(f) = 1. Hence if (4-1) holds then

$$iff(\hat{\underset{\sim}{\beta}}) = h^2 \, . \qquad (4\text{-}3)$$

We observe that the results of section 2 indicate that in the Gaussian case the residuals can be scaled to remove the effect of $\underset{\sim}{X}$ on $h^2$. However, the results in Table 3-1 show that in general this is not possible. It is evident that the dependence on $\underset{\sim}{X}$ is not a simple function of N and p, and we suspect that at the very least some measure of the average kurtosis of the columns of $\underset{\sim}{X}$ is involved in $h^2$ if the form (4-1) is to give a reasonable approximation to the covariance matrix of $\hat{\underset{\sim}{\beta}}$ for either LAR or SIN1 under a variety of types of error distributions. Huber [1973 ] and Tukey [1973] propose more complex expressions for these covariance matrices which may obviate the necessity of making $h^2$ depend on X.

REFERENCES

(1) Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. [1972] Robust Estimates of Location, Princeton University Press, Princeton, N.J.

(2) Andrews, D.F. [1973] "Some Monte Carlo Results on Robust/Resistant Regression." (unpublished).

(3) Hample, F.R. [1973] "Robust Estimation: a condensed partial survey" Z. Wahrscheinlichkeitstheorie verw. Geb. 27, 87-104.

(4) Holland, P.W. [1973] "Monte Carlo for robust regression: the swindle unmasked." NBER Working Paper No. 10.

(5) Huber, P.J. [1973] "Robust regression: Asymptotics, conjectures and Monte Carlo" Annals of Statistics 1, 799-821.

(6) Kendall, M.G. and Stuart, A. [1963] The Advanced Theory of Statistics Vol. I 2nd edition. Hafner, New York.

(7) Tukey, J.W. [1973] "A way forward for robust regression" (unpublished).