NBER WORKING PAPER SERIES

CONFIDENCE REGIONS FOR ROBUST REGRESSION

Roy E. Welsch*

Working Paper No. 111

National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

November, 1975

Preliminary: Not for Quotation

*National Bureau of Economic Research, Inc., and Massachusetts Institute of Technology. Research supported in part by National Science Foundation Grant No. GJ1154X-3.

Abstract

This paper describes the results of a Monte Carlo study of certain aspects of robust regression confidence region estimation for linear models with one, five, and seven parameters. One-step sine estimators (c = 1.42) were used with design matrices consisting of short-tailed, Gaussian, and long-tailed columns. The samples were generated from a variety of contaminated Gaussian distributions.

A number of proposals for covariance matrices were tried, including forms derived from asymptotic considerations and from weighted-least squares with data dependent weights. Comparisons with the Monte Carlo "truth" were made using generalized eigenvalues. In order to measure efficiency and compute approximate t-values, linear combinations of parameters corresponding to the largest eigenvalues of the "truth" were examined.

For design matrices with columns of modest kurtosis, the covariance estimators all give reasonable results and, after adjusting for asymptotic bias, some useful approximate t-values can be obtained. This implies that the standard weighted least-squares output using data-dependent weights need only be modified slightly to give useful robust confidence intervals.

When design matrix kurtosis is high and severe contamination is present in the data, these simple approximations are not adequate.

TABLE OF CONTENTS

1. INTRODUCTION

In the past few years, a number of ways have been proposed to perform robust regression. Perhaps the simplest to implement is called iteratively reweighted least-squares. Initial values for the coefficients are found, a scale for the residuals from this start is computed, and then a set of weights is determined by using a weight function applied to the scaled residuals. The weight function usually gives a value near one for small residuals, and near zero (or equal to zero) for large residuals. The weights are then used as if they were the weights for a weighted least-squares (WLS) regression. The above process can, of course, be iterated.

Therefore, all one needs is a device to compute the start, scale, and weights and the rest of the computation (the hard part) can be done with a standard WLS routine or by multiplying each observation by the square root of the corresponding weight and using an LS routine.

Naturally, the question arises - can we use the output from the WLS routine to find confidence regions for the regression parameters? This would make life simpler and make robust regression more readily usable. One purpose of our study of robust regression was to see if it would be feasible to use WLS output in a simple way.

This paper is organized as follows: section two discusses the robust estimator (weight function) we used, section three covers the covariance formulas, section four examines the design [X] matrix we used, and section five summarizes the parameters and distributions used in the Monte Carlo. In section six, we discuss our results for the location case. The seventh section covers our work on the regression problem and in the last section, we try to give some advice and indicate where we think we stand at this point.

2. THE REGRESSION ESTIMATOR

We want to estimate the parameter vector $\underline{\beta}$ in the model

$$\underline{y} = \underline{X}\,\underline{\beta} + \underline{e}$$

where $\underline{y}$ is the n x 1 response vector, $\underline{X}$ is an n x p design matrix, and $\underline{e}$ is an n x 1 random vector whose coordinates are independent and identically distributed symmetric random variables.

Only one family of estimators is used, the one-step sine M-estimators, which have been found to have good robustness properties in a number of previous studies. As starting values for this estimate, we used the least absolute residual estimator (called LAR):

$$\hat{\beta}_{LAR} = \min_{\underline{\beta}}{}^{-1} \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{p} x_{ij}\,\beta_j \right| .$$

Several efficient algorithms exist for finding this estimate. (See [3] for an example.) If only least-squares is available as a start, then it may not be wise to use a one-step estimator.

For an initial scale, we form $\underline{r} = \underline{y} - \underline{X}\,\underline{\beta}_{LAR}$ and compute

$$s = \frac{1}{.6745} \text{ median } \{\text{largest n-p+1 elements of } |\underline{r}|\}.$$

This modified median absolute deviations scale [HMAD] is discussed in [7] and is especially designed for an LAR start.

The weights are found from the function

$$w(t) = \begin{cases} 1 & t = 0 \\ \dfrac{\sin(t/c)}{t/c} & |t| \le \pi c \\ 0 & |t| > \pi c \end{cases} \qquad (2.1)$$

where t is replaced by the scaled residual, r_i/s. Often w (t) is approximated by $[1 - (\frac{t}{b})^2]^2$ which is the bisquare weight function. In this study we used c = 1.42 (or 2.) if the .6745 in s is omitted). This choice of c corresponds to about 96 percent asymptotic efficiency for the Gaussian error model.

Once we have found the weights, our estimate is:

$$\hat{\beta}_{SIN1} = (\underline{X}^T\underline{W}\underline{X})^{-1}\,\underline{X}^T\underline{W}\underline{y} \qquad (2.2)$$

where W is an n x n diagonal matrix of the weights. For reference purposes, we also computed the standard least-squares estimate $\hat{\underline{\beta}}_{LS}$. The weight function for least-squares is, of course, w (t) ≡ 1. Some properties of SIN1 are explored in [7].

3. COVARIANCE ESTIMATORS

In order to construct confidence regions, we would like to estimate the covariance matrix for $\hat{\underline{\beta}}_{SIN1}$. A WLS program automatically gives us

$$(W) \qquad \frac{\sum_{i=1}^{n} w_i \hat{r}_i^{\,2}}{n-p}\,(\underline{X}^T\underline{W}\underline{X})^{-1} \qquad (3.1)$$

where $\hat{\underline{r}} = \underline{y} - \underline{X}\,\hat{\underline{\beta}}_{SIN1}$. Viewing the sum of weights as a measure of "degrees of freedom" (note that the

weight function has been standardized to unity at the origin), led us to propose

$$\text{(WW)} \qquad \frac{\sum_{i=1}^{n} w_i \hat{r}_i^2}{\sum_{i=1}^{n} w_i - p} \; (\underline{X}^T \underline{W} \underline{X})^{-1} \qquad (3.2)$$

as an easily computed alternative.

Asymptotic considerations (see [9]) imply that a good choice might be

$$\text{(A)} \qquad \frac{(ns)^2}{n-p} \frac{\sum_{i=1}^{n} \psi^2\left(\frac{\hat{r}_i}{s}\right)}{\left[\sum_{i=1}^{n} \psi'\left(\frac{\hat{r}_i}{s}\right)\right]^2} (\underline{X}^T \underline{X})^{-1} \qquad (3.3)$$

where $\psi(r) = r \cdot w(r)$. Another alternative that we began to use after a part of the study was completed combined part of (W) with part of (A) to get

$$\text{(AW)} \qquad \frac{(ns)^2}{n-p} \frac{\sum_{i=1}^{n} \psi^2\left(\frac{\hat{r}_i}{s}\right)}{\left[\sum_{i=1}^{n} \psi'\left(\frac{\hat{r}_i}{s}\right)\right]^2} (\underline{X}^T \underline{W} \underline{X})^{-1} \qquad (3.4)$$

Many other forms have been proposed in [9] and [13].

What is the truth to which we should compare these forms? There is no agreed upon answer, but we chose the unconditional covariance matrix

$$A = \frac{1}{R} \sum_{j=1}^{R} \hat{\underline{\beta}}_j \hat{\underline{\beta}}_j^T \qquad (3.5)$$

where R is the number of samples used in the Monte Carlo. We emphasize the word unconditional because a covariance matrix proportional to $(\underline{X}^T \underline{W} \underline{X})^{-1}$ involves the weights and is clearly conditional on the data.

How can we measure the difference between, say, (3.1) and (3.5)? Basically, we are seeking confidence intervals for linear combinations of the parameters, $\underline{\ell}^T \underline{\beta}$. If the true covariance matrix is A, then $\underline{\ell}^T A \underline{\ell}$ is the variance of this linear combination. Since we did not want to consider, at this point, specific linear combinations we chose to examine for an estimated covariance matrix, B,

$$\max_{\underline{\ell}} \frac{\underline{\ell}^T A \underline{\ell}}{\underline{\ell}^T B \underline{\ell}} = \max_{\underline{u}} \frac{\underline{u}^T B^{-T/2} A B^{-1/2} \underline{u}}{\underline{u}^T \underline{u}} \qquad (3.6)$$

where $\underline{u} = B^{1/2} \underline{\ell}$ and $B^{1/2}$ denotes one of the standard ways to find a square root of a positive definite matrix. The right-hand side of (3.6) is equivalent to finding the largest value of λ in

$$\det (\underline{A} - \lambda \underline{B}) = 0. \qquad (3.7)$$

The solutions of (3.7) are known as generalized eigenvalues. Instead of considering just λ_{max} and λ_{min}, we followed a suggestion of Tukey [13] and computed

$$H^2 = \frac{1}{p} \sum_{i=1}^{p} \left(\frac{1}{\lambda_i} - \lambda_i\right)^2 \qquad (3.8)$$

The quantity, H, is not scale invariant. In order to measure difference in shape we followed a suggestion of Paul Holland and put A and B in correlation form. Since correlation matrices are not invariant (i.e., do not remain correlation matrices) under changes in the basis of the X-space, this approach is very X dependent. H does, however, use more than one parameter to remove "scale".

A one parameter scale-free approach would be to consider the minimum over d of $H^2(d)$ where

$$H^2(d) = \frac{1}{p} \sum_{i=1}^{p} \left(\frac{d}{\lambda_i} - \frac{\lambda_i}{d}\right)^2 . \qquad (3.9)$$

We did not do this for the full covariance matrix but we did use it for the diagonal elements (the variances). We computed

$$v_i = a_{ii}/b_{ii}$$

$$d = \left(\frac{\sum_{i=1}^{p} v_i^2}{\sum_{i=1}^{p} \frac{1}{v_i^2}}\right)^{1/4} \qquad (3.10)$$

$$D^2 = \frac{1}{p} \sum_{i=1}^{p} \left(\frac{d^*}{v_i} - \frac{v_i}{d^*}\right)^2 \qquad (3.11)$$

D^2 is, of course, invariant under a change of basis.

H was tried on the full covariance matrix after some adjustment factors, suggested by Colin Mallows [11] were used to correct for asymptotic bias.

The above measures could be termed "diagnostic" and are useful in finding which classes of covariance estimate seem reasonable. We would still need, however, to consider the "scale" of the covariance matrix and find something like t-statistics. There are two kinds of scalar multipliers of $(\underline{X}^T \underline{X})^{-1}$ or $(\underline{X}^T \underline{W} \underline{X})^{-1}$, those that depend on the data and those that do not. The later kind are very difficult to separate from t-statistics and can make it hard to develop useful approximate t tables.

We can choose among data dependent scalars by looking at confidence region "size". In the

TABLE 4-1

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

location case this can be accomplished by computing average confidence interval length once a t-statistic has been determined.

In the regression case measuring "size" is more complicated. Here we chose to consider specific linear combinations of the parameters so that the problem could be reduced to considering t-statistics and average confidence interval lengths. The linear combination we chose was the eigenvector of the largest eigenvalue of $\underline{A}$, the Monte Carlo "truth". This corresponds to the linear combination giving the largest variance, $\underline{\ell}^T\underline{A}\underline{\ell}$. This means, of course, that $\underline{\ell}$ changes with the sampling situation.

4. THE X-MATRIX

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

The 6 columns of VDATA1 were divided 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.

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 were reduced in magnitude until they contributed 80% and 95% of the total sum of squares of their columns, respectively.

After the 6 columns of VDATA1 were selected each column was standardized to have mean zero and unit sum of squares. In Table 4-1 we give the final, standardized columns of VDATA1. Further details about VDATA1 can be found in [7].

The two X-matrices we used were formed by appending a column of ones to the first 4 columns of VDATA1 to get V5 (where 5 refers to the number of fitted parameters) and to all of VDATA1 to get V7. Thus V5 could be considered to be well-behaved while V7 contains outliers.

5. THE MONTE CARLO

The Monte Carlo work was done in two parts. The first dealt with the location problem for samples of size 10, 20, and 60 using SIN1. Swindling techniques, such as those described in [12] and [1], were used throughout. The Gaussian and the slash (see [1]) were the error distributions. The slash distribution has the very long tails of a Cauchy distribution, but with a Gaussian-like center.

The second part using V5 and V7 relied on samples generated by Richard Hill and Paul Holland for their related study [7] where it was felt that Monte Carlo swindling would not be worth the cost. Unfortunately, this meant we could not use swindling for out study, except in the Gaussian case. The regression results are therefore less precise than we would have liked. For details on the swindles used in the Gaussian case see [8].

The error distributions, $f(\cdot)$, for the regression case are a simple 2-parameter family of a mixture of a $N(0,1)$ density with a $N(0,k^2)$ density. This is given by

$$g(u) = \frac{1}{\sqrt{2\pi}}\left[(1-\alpha)e^{-u^2/2} + \frac{\alpha}{k}e^{-\frac{1}{2}(\frac{u}{k})^2}\right]$$

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

$$f(u) = \frac{1}{\tau}\ g(\frac{u}{\tau})$$

where $\tau^2 = 1 - \alpha + \alpha k^2$.

Both of the estimators LAR and SIN1 are regression invariant in the sense that if the vector of observed values $\underline{y}$ is transformed to $\underline{y} = \underline{X}\ \underline{\beta}^*$ for some $\underline{\beta}^*$ then $\hat{\underline{\beta}}$ is transformed to $\hat{\underline{\beta}} + \underline{\beta}^*$. Here all of the Monte Carlo results were computed with the true values of $\underline{\beta}$ set equal to 0.

In all cases the number of Monte Carlo samples was 500. Standard errors were computed directly except in the case of the t-statistics where 5 blocks of size 100 were used to obtain an estimate of standard error.

6. LOCATION

The location problem with $\underline{X}$ equal to a column of ones provided a basic starting point for our study. Location has been extensively studied by Gross [5] who found that an estimator similar to SIN1 performed very well when used with the (A) variance to construct confidence intervals. Of course, we were mainly interested in seeing if (W) would also work, since it comes naturally from weighted least-squares.

For each sampling situation and variance estimator we computed a value, t*, corresponding to the .975 point of the empirical distribution of the t-statistics, and a measure of sampling error. We will study t* in detail later.

As a measure of efficiency, we computed the average length of the intervals [ACIL] that resulted when t* was used to form confidence intervals. These numbers (times 1000) are their standard errors are displayed in Exhibit 6-1 as tables that have been analyzed by using medians to decompose each column into:

$$\text{ACIL} = \text{residual+column fit (CF)}. \qquad (6.1)$$

The standard errors (SE) displayed at the bottom of the table are the median of the standard errors of the ACILs in the respective column.

TABLE 6-1

ACIL (x1000)

	n=10 G	n=10 S		n=20 G	n=20 S
W	57	5	W	10	2
WW	-5	-60	WW	0	-9
A	0	0	A	-4	0
CF	1673	3944	CF	991	2343
SE	23	90	SE	8	34

n=60

	G	S
W	1	0
WW	0	-2
A	-1	0
CF	538	1231
CE	2	10

We see that in all cases, (W) is worse than (A), but probably not enough to warrant writing special programs to compute (A). Except for two cases (WW) is better than (A), but the gain is within a single standard error.

To make these confidence procedures work, we need to find a t-like value that is independent of the underlying distribution, since we do not know what that distribution is. Our t* values are listed in Exhibit 6-2. The standard errors listed are the median of the column standard errors.

In this case, the t* value at the Gaussian is always larger than that for slash, so we will focus our attention on the Gaussian values.

TABLE 6-2

t*

	n=10 G	n=10 S		n=20 G	n=20 S
W	3.21	2.88	W	2.53	2.49
WW	2.93	2.53	WW	2.39	2.23
A	2.71	2.43	A	2.21	2.13
SE	.09	.06	SE	.03	.04

n=60

	G	S
W	2.30	2.27
WW	2.20	2.05
A	2.05	1.95
SE	.01	.03

This leaves us with a table of t* numbers that we could use with formula (W) and feel reasonably sure of getting 95% confidence intervals of moderately high efficiency for distributions in the "neighborhood" of the Gaussian.

We had originally hoped to base the degrees of freedom on $\sum_{i=1}^{n} w_i - 1$, i.e., to have a conditional degrees of freedom formula. This has not worked well, giving t-numbers too small at the Gaussian and too large at the slash. A close look at our tables of t* for (W) shows that it is not easily related to the standard t-tables. However the t* tables for (A) seem to be approximable by the standard t on (n-1)/2 degrees of freedom. (Gross [5] also noticed this.)

Mallows [11] had proposed a form of asymptotic adjustment for covariance formulas that would remove any asymptotic bias relative to the correct asymptotic formula, (A). For example, the adjustment factor for (W) is:

$$AF_W = \frac{E(\psi^2(Z))}{E^2(\psi'(Z))} \frac{E(w(Z))}{E(Z\psi(Z))},$$

where Z is N(0,1) since we want to do the adjustment for the Gaussian case. To adjust t* we use $(AF)^{-1/2}$. Table 6-3 lists these t* adjustment factors which have been computed by using the bisquare approximation to the sine weight function with b = 4.72 (or 1.5π).

TABLE 6-3

	Adjustment Factor
W	.89
WW	.93
AW	1.04
A	1.00

Exhibit 6-4 contains the adjusted Gaussian t* values. We see that the values are much more comparable and that $[\frac{n-1}{2}]$ is a useful, but not perfect degrees of freedom approximation.

TABLE 6-4

	Adjusted t*		
n	10	20	60
W	2.86	2.25	2.05
WW	2.73	2.22	2.05
A	2.71	2.21	2.05
$t_{[\frac{n-1}{2}]}$	2.78	2.26	2.05

7. REGRESSION RESULTS

For the regression problem we only used one sample size, 20, and the matrices V5 and V7, giving 15 and 13 degrees of freedom or 3 degrees of freedom per parameter for V5 and aobut 2 degrees of freedom per parameter for V7. For location we had 9, 19 and 59 degrees of freedom, so we were not expecting and did not get the pleasant results obtained in the location case. On the other hand, we see many regressions run with 2 and 3 degrees of freedom per parameter and we felt it was necessary to gather some information about these cases.

We first computed H [see (3.8)], comparing the Monte Carlo truth to $(\underline{X}^T\underline{X})^{-1}$ and $(\underline{X}^T\underline{W}\underline{X})^{-1}$ where all matrices have been put in correlation form. The results are contained in Exhibit 7-1 where we have included the midspread (interquartile range) of the H values for $(\underline{X}^T\underline{W}\underline{X})^{-1}$ and also the values of H obtained when the standard least-squares Monte Carlo truth is compared to $(\underline{X}^T\underline{X})^{-1}$ in correlation form (LS).

EXHIBIT 7-1

H(x100) for $(\underline{X}^T\underline{W}\underline{X})^{-1}$

	G	3.1	3.25	10.1	10.25
			V5 (LS=10)		
$(X^TX)^{-1}$	8	12	17	21	42
MED	12	18	23	32	53
MS	8	10	11	17	19
			V7 (LS=23)		
$(X^TX)^{-1}$	28	28	32	108	84
MED	27	31	36	119	100
MS	2	7	8	16	17

These results seem to indicate that $(\underline{X}^T\underline{X})^{-1}$ is a better choice than $(\underline{X}^T\underline{W}\underline{X})^{-1}$ most of the time. In other words, for overall shape, it may not pay to use a form conditional on the data. On the other hand, we have been comparing both forms to the unconditional Monte Carlo truth. We do not yet see how to do these comparisons in a conditional way.

Next we looked at D [see (3.11)] and found a similar story (Exhibit 7-2) although less pronounced. The worst cases occurred, as we might expect, for V7 and for CG10.1 and CG10.25. We are still puzzled by the fact that for V7, CG10.1 is worse than CG10.25. Similar results show up in later tables.

At this point we decided to look at six covariance formulas - the four discussed in section three and (W') which is (W) but using $(\underline{X}^T\underline{X})^{-1}$ instead of $(\underline{X}^T\underline{W}\underline{X})^{-1}$ and (WW') which modifies (WW) in a similar way. The asymptotic adjustment factors developed in section six were applied (the factors for W' and WW' are .85 and .89) but this time we used AF and not $AF^{-1/2}$ since we were looking at variances and not t-statistics.

H, as defined in (3.8), was then computed for all these cases; the results form Exhibit 7-3. The

EXHIBIT 7-2

D(x100) for $(X^TWX)^{-1}$

V5 (LS=5)

	G	3.1	3.25	10.1	10.25
$(X^TX)^{-1}$	7	9	12	16	21
MED	9	11	14	19	28
MS	2	5	5	7	11

V7 (LS=8)

$(X^TX)^{-1}$	9	12	12	91	57
MED	9	15	16	98	66
MS	2	5	6	8	9

EXHIBIT 7-3

Median H(x100) for Adjusted Covariances

V5 (LS=57)

	G	3.1	3.25	10.1	10.25
W	0	5	1	5	1
WW	1	0	-3	-3	-30
AW	5	1	6	1	-2
W'	0	7	-1	2	17
WW'	0	0	-3	-8	-20
A	2	-1	3	-2	19
CF	60	78	100	114	294
SE	88	130	119	152	381

V7 (LS=64)

W	0	1	-1	20	12
WW	-3	-2	1	-17	-76
AW	4	8	15	51	44
W'	0	-3	-5	-17	-12
WW'	-4	6	-7	-49	-100
A	4	3	9	18	14
CF	75	102	113	456	573
SE	155	214	227	600	1052

medians of the standard errors for each column, SE, show that these numbers are quite variable and comparisons will be difficult. Generally speaking (WW) and (WW') perform well, confirming some of our earlier results. It also appears to be the case that forms involving $(X^TX)^{-1}$ perform a little better, which agrees with Exhibits 7-1 and 7-2. We decided not to continue to look at (W') and (WW') because $(X^TX)^{-1}$ is not available when computing SIN1 using weighted least-squares The form (A) was carried along as a benchmark.

We now tried to measure efficiency by using the eigenvectors discussed in section three. As in the location case we computed a t* for each situation and then the average confidence interval length. The ACIL results are listed in Exhibit 7-4.

EXHIBIT 7-4

ACIL (x100) for Regression

V5 (LS=536)

	G	3.1	3.25	10.1	10.25
W	4	-1	4	2	5
WW	-3	1	-1	1	1
AW	-1	-6	0	-5	-1
A	1	3	-4	-1	0
CF	566	481	472	217	220

V7 (LS=864)

W	21	6	26	14	-9
WW	-4	-2	5	3	0
AW	-5	2	-34	-4	7
A	5	-12	-6	-2	0
CF	984	826	812	464	522

The average standard error in this table is about 10 so that it is hard to make any real distinctions. Generally, (AW) seems to perform best. As far as efficiency goes we see no reason to go beyond (W) or perhaps (WW).

Finally we computed adjusted t*-values, as in the location case. These numbers are listed in Exhibit 7-5 where we have added a new column for GS, the values obtained by swindling rather than simple sampling. Since the maximum swindle gain occurs at the Gaussian, we may not have lost too

EXHIBIT 7-5

Adjusted t*

V5 (t_7=2.37)

	G	GS	3.1	3.25	10.1	10.25
W	2.32	2.35	2.37	2.38	2.53	2.49
WW	2.31	2.30	2.30	2.36	2.41	2.33
AW	2.26	2.38	2.25	2.33	2.47	2.46
A	2.21	2.41	2.22	2.25	2.35	2.23
SE	.10	.04	.06	.10	.14	.10

V7 (t_6=2.45)

W	2.62	2.44	2.55	2.58	3.64	3.39
WW	2.61	2.37	2.59	2.48	3.53	3.31
AW	2.53	2.53	2.46	2.53	3.52	3.13
A	2.44	2.55	2.39	2.41	3.28	2.97
SE	.13	.04	.03	.13	.28	.08

much by our inability to swindle all of these numbers (see section three).

No longer can we say that the values at the Gaussian are the largest, so perhaps adjustment at the Gaussian is suspect. However, for V5 we would feel reasonably happy with $t_7 = 2.37$ as an approximate t-value for SIN1 regression, especially for (WW).

Clearly this breaks down for V7 and CG10.1 and Cg10.25 where $t_6 = 2.45$ is not adequate. Severe contamination coupled with a distorted X-matrix (high kurtosis) has diminished our hopes for simple approximations.

8. CONCLUSIONS

Strictly interpreted our results only apply to the specific situations and X-matrices examined in this study. However, we would like to generalize somewhat. Hill and Holland [7] have shown that SIN1 is a reasonably good robust regression estimator. We feel that the usual output from the one-step of weighted least-squares can be used for inference providing the t-statistic is found using $[\frac{n-p}{2}]$ degrees of freedom and then divided by the adjustment factor (.89). If the X-matrix is really unusual then caution is advised.

We have most often used the above results in a diagnostic way. The least-squares and SIN1 regression results are both obtained and then compared in various ways. If the confidence intervals are radically different (or test results reversed) we explore further, attempting to diagnose the cause of this instability.

We have left many questions unanswered. We are examining what happens to our regression results when n=40. Why $[\frac{n-p}{2}]$ for degrees of freedom? Why doesn't some conditional degrees of freedom formula $\sum_{i=1}^{n} w_i$ or $\sum_{i=1}^{n} |\psi'_i|$ work? For the abnormal X-matrix problem we may have to examine the proposal of Mallows [10] which attempts to smooth the X-matrix and reduce its kurtosis.

Finally, there is the question of how these results generalize to the F-statistics associated with more complex simultaneous confidence regions.

ACKNOWLEDGEMENTS

The author has benefitted greatly from the advice and assistance of Paul Holland, David Hoaglin, John Tukey, Alan Gross, Richard Hill, Stanley Wasserman, and Sheila Howard.

This work has been supported by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research.

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. (1974). A Robust Method for Multiple Linear Regression. *Technometrics* 16, 523-531.

[3] Barrodale, I. and F.D.K. Roberts (1973). An Improved Algorithm for Discrete L_1 Approximation. *SIAM Journal of Numerical Analysis* 10, 839-848.

[4] Denby, L. and Larsen, W.A. (1974). Robust Regression Estimators Compared Via Monte Carlo. Bell Laboratories, TM-74-1215-26/1214-18.

[5] Gross, A.M. (1974). Confidence Interval Robustness with Long-Tailed, Symmetric Distributions. Submitted to *JASA*.

[6] Gross, A.M. (1975). Confidence Intervals for Bisquare Regression Estimates. Bell Laboratories Technical Memorandum.

[7] Hill, R.W. and Holland, P.W. (1974). A Monte Carlo Study of Two Robust Alternatives to Least Squares Regression Estimation. NBER Computer Research Center Working Paper No. 58, National Bureau of Economic Research, Inc., Cambridge, Massachusetts.

[8] Holland, P.W. (1975). A Variance Reduction Technique for Monte Carlo Studies of Robust Regression Confidence Intervals. Unpublished Memorandum.

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

[10] Mallows, C.L. (1973). Influence Functions. Talk at NBER Conference on Robust Regression, 21 June 1973.

[11] Mallows, C.L. (1973). On Some Topics in Robustness. Paper delivered at the Eastern Regional ASA Meeting, University of Rochester.

[12] Relles, D.A. (1970). Variance Reduction Techniques for Monte Carlo Sampling from Student Distributions, *Technometrics*, 12 499-515.

[13] Tukey, J.W. (1973). A Way Forward for Robust Regression. Unpublished Memorandum.