Technical Report 90.02

A NOTE ON THE COMPUTATION OF ROBUST, BOUNDED INFLUENCE ESTIMATES AND TEST STATISTICS IN REGRESSION

by

Douglas P. Wiens

Page	Line	Change
1	-4	omit σ from $\sigma\varepsilon_i$
3	-8	"Here" to "Here, $\sigma = \operatorname{plim}_{n \to \infty} \hat{\sigma}$ and"
	-6	$\eta^{\ell}(\mathbf{x},\varepsilon)$ to $\eta^{\ell}(\mathbf{x},\frac{\varepsilon}{\sigma})$ (once only)
	-5	$\eta^2(\mathbf{x},\varepsilon)$ to $\eta^2(\mathbf{x},\frac{\varepsilon}{\sigma})$
6	8	$(n - p)^2$ to $(n - p)$
8	8, 10	$\eta(\mathbf{x}_0)$ to $\nu(\mathbf{x}_0)$.

A NOTE ON THE COMPUTATION OF ROBUST, BOUNDED INFLUENCE ESTIMATES AND TEST STATISTICS IN REGRESSION

Douglas P. Wiens[1]

University of Alberta

ABSTRACT

We give a method of computing bounded influence M-estimates of regression coefficients. The method has the advantage that the accompanying printout will be asymptotically correct, in that the standard errors and p-values have the correct asymptotic values. The method is easily implemented on any package which can perform weighted least squares regression, and Choleski decompositions.

The p-values are those which result from substituting robust estimates into the usual F-statistic for testing a general linear hypothesis. The influence function of this test is obtained, and shown to be bounded with respect both to the influence of residuals, and to the influence of the position of the carriers. A numerical example is given, comparing several robust estimators, and the least squares estimator.

1. INTRODUCTION

A barrier to the widespread adoption of robust regression procedures within the statistical community appears to be the perception that such procedures are difficult to compute. Street, Carroll and Ruppert (1988) addressed this problem in the case of Huber-type M-estimation. One of their main points may be summarized as follows.

Consider the linear model

$$y_i = \mathbf{x}_i^T\boldsymbol{\theta} + \sigma\varepsilon_i,\ 1 \le i \le n;$$

with i.i.d. errors ε_i. Suppose that one has obtained a Huber M-estimate $\boldsymbol{\theta}_*$, through solving the defining equations

$$n^{-1}\sum_{i=1}^{n}\psi\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\theta}_*}{\widehat{\sigma}}\right)\mathbf{x}_i = \mathbf{0}.$$

Key words and phrases: Robust regression, bounded influence estimation

[1] Research supported by the Natural Sciences and Engineering Research Council of Canada

One can then perform a least squares regression of appropriately defined *pseudovalues* on the independent variables, obtaining an asymptotically equivalent estimate $\widehat{\boldsymbol{\theta}}$. If this is carried out on any of the usual software packages, then the accompanying printout will be asymptotically correct. That is:

a) The printed estimated covariance matrix of $\widehat{\boldsymbol{\theta}}$ will be a consistent estimate of the true covariance matrix.

b) If $\boldsymbol{\theta}$ is partitioned as

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}, \tag{1}$$

then the classical test of the hypothesis $H_0 : \ \boldsymbol{\theta}_2 = \mathbf{0}$ consists of rejecting H_0 if $F > F(1-\alpha; p-q, n-p)$, where

$$F = \begin{matrix} \{\text{Regression sum of squares due to the} \\ \text{last } (p-q) \text{ regressors }\} \div (p-q)S^2. \end{matrix} \tag{2}$$

These regression sums of squares are typically printed out, so that F is easily calculated. By virtue of the regression on pseudovalues, the mean residual sum of squares S^2 estimates the "right" quantity. This F-test is then asymptotically of the correct size α, since the distributions of both F and an F^{p-q}_{n-p} random variable tend weakly to the same $\chi^2_{p-q}/(p-q)$ distribution.

In this article we discuss a similar computational approach, valid for *bounded influence estimators*. In contrast to Huber M-estimators, bounded influence estimators are robust against high leverage points. The estimate may be defined as a solution to

$$n^{-1} \sum_{i=1}^{n} \eta\left(\mathbf{x}_i, \ \frac{y_i - \mathbf{x}_i^T \boldsymbol{\theta}}{\sigma}\right)\mathbf{x}_i = \mathbf{0}. \tag{3}$$

A common class of functions η is that suggested by Schweppe (see Hampel, Ronchetti, Rousseeuw and Stahel (1986)):

$$\eta(\mathbf{x}, r) = \nu(\mathbf{x})\psi(r/\nu(\mathbf{x})) \tag{4}$$

for an appropriate positive function $\nu(\mathbf{x})$ and odd function $\psi(r)$. Specific choices of ν and ψ are investigated in Sections 3 and 4 below. See Hampel, Ronchetti, *et al* (1986), Krasker and Welsh (1982), Huber (1983) for further discussions related to the choice of the function η.

A commmon, easily programmed algorithm for solving (3) is as follows:

i) From $\boldsymbol{\theta}_{(k)}$ ($k = 0, 1, ...$) compute a scale estimate $\sigma_{(k)}$. Put $e_{i,k} = y_i - \mathbf{x}_i^T \boldsymbol{\theta}_{(k)}$, $r_{i,k} = e_{i,k}/\sigma_{(k)}$.

ii) Form weights $w_{i,k} = \eta(\mathbf{x}_i, r_{i,k})/r_{i,k}$. With $\boldsymbol{\theta} = \boldsymbol{\theta}_{(k)}$, (3) now becomes the weighted least squares problem

$$\sum_{i=1}^{n} w_{i,k} y_i \mathbf{x}_i = (\sum_{i=1}^{n} w_{i,k} \mathbf{x}_i \mathbf{x}_i^T)\boldsymbol{\theta} \tag{5}$$

with weights dependent upon $\boldsymbol{\theta}_{(k)}$.

iii) "Solve" (5), thus obtaining $\theta_{(k+1)}$, by performing a weighted least squares regression of the y's on the $\mathbf{x}$'s, using weights $w_{i,k}$.

iv) Iterate to convergence.

A frequent choice of $\sigma_{(k)}$, suggested by Hill and Holland (1977), is

$$\sigma_{(k)} = 1.48 \times \{ \text{ Median of the largest } n - p + 1 \text{ of the } |e_{i,k}|\}. \tag{6}$$

The factor $1.48 (= 1/\Phi^{-1}(.75))$ is for consistency at the normal distribution.

The final values $(\boldsymbol{\theta}_*, \sigma_*)$ of the iterates are, under mild conditions (see Maronna and Yohai (1981)), consistent and asymptotically normally distributed, and

$$\sqrt{n}(\boldsymbol{\theta}_* - \boldsymbol{\theta}) \xrightarrow{w} N(0, \sigma^2 C^{-1}). \tag{7}$$

Here,

$$C = MQ^{-1}M,$$
$$M = E[\eta'(\mathbf{x}, \varepsilon)\mathbf{x}\mathbf{x}^T], \ (\eta'(\mathbf{x}, \varepsilon) = \frac{\partial}{\partial \varepsilon}\eta(\mathbf{x}, \varepsilon))$$
$$Q = E[\eta^2(\mathbf{x}, \varepsilon)\mathbf{x}\mathbf{x}^T].$$

The regressors $\mathbf{x}$ may be random, in which case they are assumed to be distributed independently of ε. Note that if (4) is used, then $\eta'(\mathbf{x}, \varepsilon) = \psi'(\varepsilon/\nu(\mathbf{x}))$.

A further problem is the estimation of the matrix C, and the computation of test procedures for which estimates of C are required. Street, Carroll and Ruppert (1988) pointed out that the estimated standard errors normally appearing on the weighted least squares printout are inconsistent, in the case of Huber M-estimation. For bounded influence estimation such estimates are not merely inconsistent but meaningless, due to the relatively complex structure of C.

In Section 2 below we give an easily implemented method, involving a further least squares regression on pseudovalues, which has both properties a) and b) above. We give as well the non-centrality parameter in the limiting χ^2-distribution of the test statistic, when H_0 is false. In Section 3 the influence function of the test is presented. It is shown that, provided $\psi(r)$ and $\|\mathbf{x}\|\nu(\mathbf{x})$ are bounded, the influence function is bounded with respect to both the influence of residuals, and to the influence of the position of the carriers. Two appropriate choices of $\nu(\mathbf{x})$ are given. Some numerical comparisons are made in the example of Section 4.

The test statistic determined by (2), following the regression on pseudovalues, is in fact identical to $R_n^2/(p-q)$, in the notation of Hampel et al (1986, p. 364). Hampel et al discuss $R_n^2/(p-q)$ as well as another statistic for testing H_0 - their $W_n^2/\widehat{\sigma}^2$ - and several asymptotically equivalent versions. In two of the cases in which they are able to explicitly evaluate the statistics and their asymptotic distributions - Huber M-estimation, and the case $p-q=1$ - it turns out that the F of (2), $R_n^2/(p-q)$, and $W_n^2/\widehat{\sigma}^2$ are in fact all identical. In general, however, the F of (2) does not agree with $W_n^2/\widehat{\sigma}^2$. In such cases, the F-based test enjoys the advantage that, as described in Section 2, both the test statistic and its asymptotic distribution are easier to compute than are those for $W_n^2/\widehat{\sigma}^2$.

We note that there are several sophisticated, main-frame based computer packages available for the computation of bounded influence estimates. See Marazzi (1987) for one such package. The procedures outlined here have the advantage of being easily implemented by the casual user, or by students. Indeed, a program which runs on MINITAB has been used successfully in classes, and is available from the author.

2. COMPUTATION OF THE ESTIMATES

Suppose that $(\boldsymbol{\theta}_*, \sigma_*)$ satisfy (3), with σ_* determined from $\boldsymbol{\theta}_*$ as at (6), or in any other manner which ensures its consistency for σ. Let M_n, Q_n be consistent estimates of M and

Q, e.g.

$$M_n = n^{-1} \sum_{i=1}^{n} \eta'(\mathbf{x}_i, r_i)\mathbf{x}_i\mathbf{x}_i^T,$$

$$Q_n = n^{-1} \sum_{i=1}^{n} \eta^2(\mathbf{x}_i, r_i)\mathbf{x}_i\mathbf{x}_i^T,$$

where $r_i = (y_i - \mathbf{x}_i^T\boldsymbol{\theta}_*)/\sigma_*$. Assume that M_n, Q_n and the design matrix X are of full rank p. Let

$$C_n = M_n Q_n^{-1} M_n,$$

and let A_n be an upper triangular matrix satisfying

$$A_n^T A_n = nC_n. \tag{8}$$

Decompose the design matrix as

$$X = \Gamma U \tag{9}$$

where $\Gamma: \; n \times p$ has orthonormal columns and U is upper triangular. Note that (8), (9) involve only Choleski deompositions. Define

$$V_n = \Gamma A_n = XU^{-1}A_n, \tag{10}$$

$$\boldsymbol{\eta}(\mathbf{x}, r) = (\eta(\mathbf{x}_1, r_1), \ldots, \eta(\mathbf{x}_n, r_n))^T.$$

Then by (3) and (10),

$$X^T\boldsymbol{\eta}(\mathbf{x}, r) = V_n^T\boldsymbol{\eta}(\mathbf{x}, r) = \mathbf{0}. \tag{11}$$

Compute a vector of pseudovalues

$$\mathbf{y}_* = V_n\boldsymbol{\theta}_* + k_n\boldsymbol{\eta}(\mathbf{x}, r),$$

where

$$k_n = \sqrt{n-p}\,\sigma_*/||\boldsymbol{\eta}(\mathbf{x}, r)||.$$

Regress $\mathbf{y}_*$ on the columns of V_n-remembering to fit a no-intercept model since V_n does not have a column of ones - to obtain a final estimate $\widehat{\boldsymbol{\theta}}$. By virtue of (11),

$$\widehat{\boldsymbol{\theta}} = (V_n^TV_n)^{-1}V_n^T\mathbf{y}_* = \boldsymbol{\theta}_*.$$

Define $\widehat{\sigma}$ to be σ_*.

On typical regression packages, the printout for this final regression will include standard errors of estimates, p-values, etc. determined from the estimated covariance matrix of $\widehat{\boldsymbol{\theta}}$. This matrix is calculated as

$$\text{“}est.cov.(\widehat{\theta})\text{”} = s^2(V_n^T V_n)^{-1} = \frac{s^2}{n} C_n^{-1},$$

where

$$\begin{aligned} s^2 &= ||(I - V_n(V_n^T V_n)^{-1} V_n^T)\mathbf{y}_*||^2/(n-p) \\ &= k_n^2 ||\eta(\mathbf{x}, r)||^2/(n-p)^2 \\ &= \widehat{\sigma}^2. \end{aligned}$$

Since C_n is consistent for C and $\widehat{\sigma}$ for σ, objective a) of Section 1 is met. To see that objective b) is met as well, first decompose $\widehat{\boldsymbol{\theta}}$ as $(\widehat{\boldsymbol{\theta}}_1^T, \widehat{\boldsymbol{\theta}}_2^T)^T$, compatibly with (1). Standard algebraic manipulations show that the F statistic, calculated from a printout as at (2), is in fact given by

$$F = \left(n\widehat{\boldsymbol{\theta}}_2^T \, \frac{C_{n;22.1}}{\widehat{\sigma}^2} \, \widehat{\boldsymbol{\theta}}_2 \right) / (p-q), \tag{12}$$

where

$$\widehat{\sigma}^2 C_{n;22.1}^{-1} = \widehat{\sigma}^2 (C_n^{-1})_{22}$$

is the estimated covariance matrix of $\sqrt{n}\widehat{\boldsymbol{\theta}}_2$. Assuming that (7) holds, the limiting distribution of F, under alternatives

$$H_a^{(n)} : \ \boldsymbol{\theta}_2 = \boldsymbol{\Delta}/\sqrt{n},$$

is that of a $\chi^2_{p-q}(\delta^2)/(p-q)$ random variable. The non-centrality parameter is given by

$$\delta^2 = \boldsymbol{\Delta}^T C_{22.1} \boldsymbol{\Delta}/\sigma^2,$$

with $C_{22.1} = ((C^{-1})_{22})^{-1}$. Thus, objective b) is met.

We note that it is not necessary that an exact zero be attained, at (3), by $(\boldsymbol{\theta}_*, \sigma_*)$. It suffices if

$$n^{-1/2} V_n^T \eta(\mathbf{x}, r) \xrightarrow{pr} \mathbf{0}.$$

3. INFLUENCE FUNCTION

Recall (12), and put

$$T^2 = \widehat{\boldsymbol{\theta}}_2^T C_{n;22.1}\boldsymbol{\theta}_2/(p-q).$$

The influence function of T is defined by

$$IF(\mathbf{z}_0; T, H_{\boldsymbol{\theta}}) = \lim_{t\to 0} \frac{T((1-t)H_{\boldsymbol{\theta}} + t\Delta_{\mathbf{z}_0}) - T(H_{\boldsymbol{\theta}})}{t}.$$

Here, $\Delta_{\mathbf{z}_0}$ is the distribution function which places all mass at $(\mathbf{z}_0^T, y_0)^T$, and $H_{\boldsymbol{\theta}}$ is the true distribution function, under the model.

The influence function represents the limiting influence of an observation at $\mathbf{z}_0$ on the test statistic, normalized by the amount of mass at $\mathbf{z}_0$. For robustness against outlying y-values, and for bounded influence, we require that the gross error sensitivity, defined by

$$GES(T, H_{\boldsymbol{\theta}}) = \sup_{\mathbf{z}_0} |IF(\mathbf{z}_0; T, H_{\boldsymbol{\theta}})|$$

be finite.

For the calculations, we follow the procedures in chapter 6 of Hampel *et al*(1986). We take $\sigma^2 = 1$, and assume that the null hypothesis $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$ is true.

Partition M^{-1} as

$$M^{-1} = \begin{pmatrix} M_{(1)} \\ M_{(2)} \end{pmatrix} \begin{matrix} q \times p \\ (p-q) \times p \end{matrix}.$$

Define

$$\tilde{\boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \mathbf{0} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}, A_{p\times p} = M_{(2)}^T C_{22.1} M_{(2)}.$$

Under the regularity conditions of Maronna and Yohai (1981), we then find that

$$IF(\mathbf{z}_0; T, H_{\tilde{\boldsymbol{\theta}}}) = |\eta(\mathbf{x}_0, y_0 - \mathbf{x}_0^T\tilde{\boldsymbol{\theta}})|\{\mathbf{x}_0^T A\mathbf{x}_0/(p-q)\}^{1/2}. \tag{13}$$

For $\eta(\cdot)$ as at (4), (13) becomes

$$IF(\mathbf{z}_0; T, H_{\tilde{\boldsymbol{\theta}}}) = \|\mathbf{x}\|\nu(\mathbf{x}_0) \cdot |\psi(\frac{y_0 - \mathbf{x}_0^T\tilde{\boldsymbol{\theta}}}{\nu(\mathbf{x}_0)})| \cdot \{\frac{\mathbf{x}_0^T A\mathbf{x}_0}{(p-q)\mathbf{x}_0^T\mathbf{x}_0}\}^{1/2}. \tag{14}$$

Since A does not depend upon $\mathbf{z}_0$, (14) then gives

$$GES(T, H_{\tilde{\boldsymbol{\theta}}}) \le \sup_{\mathbf{x}_0} \|\mathbf{x}_0\|\nu(\mathbf{x}_0) \cdot \sup_r |\psi(r)| \cdot \{\frac{ch_{\max}(A)}{p-q}\}^{1/2}, \tag{15}$$

APPENDIX I: VERIFICATION OF EQUATION (12)

Partition X, Γ, U as

$$X = \underset{q}{(X_1} \;\vdots\; \underset{(p-q)}{X_2\;)} = (\;\underset{q}{\Gamma_1} \;\vdots\; \underset{(p-q)}{\Gamma_2\;)} \begin{pmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix} = \Gamma U.$$

Since $s^2 = \widehat{\sigma}^2$, the statistic F at (2) has

$$\begin{aligned}(p-q)\widehat{\sigma}^2 F &= \text{"SSE in reduced model"} - \text{" SSE in full model"} \\ &= \|(I - H_{X_1})\mathbf{y}_*\|^2 - \|(I - H_X)y_*\|^2,\end{aligned}$$

where

$$\begin{aligned}H_{X_1} &= X_1(X_1^T X_1)^{-1} X_1^T = \Gamma_1 \Gamma_1^T, \\ H_X &= X(X^T X)^{-1} X^T = \Gamma\Gamma^T.\end{aligned}$$

Thus

$$\begin{aligned}(p-q)\widehat{\sigma}^2 F &= \mathbf{y}_* \Gamma_2 \Gamma_2^T \mathbf{y}_* \\ &= \|\Gamma_2^T (V_n \boldsymbol{\theta}_* + k_n \boldsymbol{\eta}(\mathbf{x}, r))\|^2 \\ &= \|\Gamma_2^T V_n \widehat{\boldsymbol{\theta}}\|^2 \text{ (by (7))} \\ &= \|(0 \vdots I_{p-q}) A_n \widehat{\boldsymbol{\theta}}\|^2 \text{ (by (10))} \\ &= \|A_{n;22} \boldsymbol{\theta}_2\|^2, \end{aligned} \qquad (A.1)$$

where

$$A_n = \begin{pmatrix} A_{n;11} & A_{n;12} \\ 0 & A_{n;22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix}.$$

Now

$$((nC_n)^{-1})_{22} = (A_n^{-1} A_n^{-T})_{22} = (A_{n;22}^T A_{n;22})^{-1},$$

so that

$$A_{n;22}^T A_{n;22} = (nC_n)_{22.1}. \qquad (A.2)$$

Now (12) follows from (A.1) and (A.2).

APPENDIX II: MINITAB CODE FOR BOUNDED INFLUENCE REGRESSION

SUBROUTINE 'BIDRIVE'

```
noecho
outfile '-temp';
noterm.
#   This program drives a bounded influence regression package,
using
# iteratively reweighted least squares.  A final regression then
uses
# pseudo-values.  A Schweppe's type eta function is used: eta(x,r)=
# w(x)*psi(r/w(x)).  Here, r is the residual/sigmahat and psi is
# Huber's psi-c, with c=2*sqrt(p/n).  The weight function is given
by
# w(x)=sqrt(1-h(i), where h(i) is the ith diagonal element of
# the hat matrix.
#  The  program  calls  BIREG1,  which  performs  the  iteratively
reweighted
# least squares regressions.  Then it calls BIREG2 which calculates
#  pseudo-values  (using  subroutines  CHOLESKI  and  CHOLITER)  and
performs the
# final regression, printing, plotting (using subroutine BIPLOTS)
# etc.   Robust  estimates  of  scale  are  calculated  on  subroutine
SIGMAHAT.
# It is required that the data file have X1 in C1,...,Xk1 in Ck1,
# Y in Ck2, where k2=p.  The next statements get the data.
retr 'stack'
let k1=3 #k1 is number of columns of regressors in data file.
let k2=4 #k2 is p, the number of parameters, and column containing
y.
let k4=count(ck2) # k4 is n, the number of cases.
#  Now  set  the  initial  scale  estimate,  and  initial  coefficient
vector.
let k3= 2
name c101 'coeffs' c102 'ones'
set 'coeffs'
 -40 .7 1.3 -.15
end
#  Set number of iterations, for computation of M-estimates
let k5=10
# Empty BIPRINT
#  Now  create the X-matrix.  For regression through the origin,
omit 'ones'.
set 'ones'
k4(1)
end
copy 'ones', c1-ck1 into m1
# Now get initial vector of (unstandardized) errors
name c103 'fits' c104 'errors'
mult m1 'coeffs' 'fits'
```

```
let 'errors'=ck2-'fits'
# Set iteration number
let k11=0
name c105 'studtres' c106 'psi(r/w)' c107 'pseudos' c199 'scales'
c198 'iter.no.' c109 'obj' c110 'hatdiags'
name c116 'weights' c111 'index'  c114 'r/w' c115 'eta(x,r)' c117
'w(x)'
# Collect iter.no., scale estimate, coefficients
let 'iter.no.'=k11
let 'scales'=k3
let c200='coeffs'
# Calculate studentized residuals (=errors/sigmahat)
let 'studtres'='errors'/k3
# Calculate w(x)
brief 1
regr ck2 k1 c1-ck1;
hi 'hatdiags'.
let 'w(x)'=sqrt(1-'hatdiags')
# Set turning point k6 of the Huber psi-function.
let k6=2*sqrt(k2/k4)
let k7=-k6
# Initialize objective=norm of X-transpose eta, over root n.
let 'obj'(1)='*'
exec 'bireg1' k5 times
exec 'bireg2'
echo
nooutfile
end

SUBROUTINE 'BIREG1'

# Calculate r/w(x)
let 'r/w'='studtres'/'w(x)'
# Calculate psi=Huber's psi(k6)
rmax 'r/w' k7 'psi(r/w)'
rmin 'psi(r/w)' k6 'psi(r/w)'
# Calculate eta(x,r).
let 'eta(x,r)'='psi(r/w)'*'w(x)'
# Calculate weights=eta(x,r)/r.
let 'weights'='eta(x,r)'/'studtres'
# Do weighted regression
brief 1
regr ck2 k1 c1-ck1 c108 'fits';
residuals in 'errors';
coeff 'coeffs';
weights are in 'weights'.
erase c108
# Calculate scale estimate
exec 'sigmahat'
# Up-date the studentized residuals
let 'studtres'='errors'/k3
# Collect iteration number, scale estimate, coefficients
let k11=k11+1
```

```
stack 'iter.no.' k11 'iter.no.'
stack 'scales' k3 'scales'
let k12=200+k11
let ck12='coeffs'
# Calculate objective=norm of X-transpose eta, over root n.
trans m1 m2
mult m2 'eta(x,r)' c108
let k13=sqrt(ssq(c108)/k4)
erase c108
stack 'obj' k13 'obj'
end
```

SUBROUTINE 'SIGMAHAT'

```
#   This subroutine calculates a robust estimate of scale, for
regression.
# The estimate is 1.4826*median(largest n-p+1 absolute residuals)
# Input: Residuals in 'errors', p in k2
let c108=abso('errors')
sort c108 c108
let k17=k2-1
let k18=k4-k17
set c160
k17(0) k18(1)
end
copy c108 c108;
omit c160=0.
let k3=1.4826*medi(c108)
erase c108
end
```

SUBROUTINE 'BIREG2'

```
# Calculate r/w(x)
let 'r/w'='studtres'/'w(x)'
# Calculate Huber's psi(k6)
rmax 'r/w' k7 'psi(r/w)'
rmin 'psi(r/w)' k6 'psi(r/w)'
# Calculate eta(x,r).
let 'eta(x,r)'='psi(r/w)'*'w(x)'
# Calculate etaprime(x,r)=psiprime(r/w(x)), and etasqrd.
# Then calculate nC, the inverse cov. matrix of thetahat/sigma.
let c160=abso('r/w')
let c160=k6-c160
signs c160 c160
let c160=(c160+1)/2 #c160 is etaprime
diag c160 m5
let c160='eta(x,r)'**2 #c160 is etasqrd
diag c160 m6
erase c160
```

```
# Now diag of m5 is etaprime, diag of m6 is etrasqrd.
mult m2 m5 m9
mult m9 m1 m7
mult m2 m6 m9
mult m9 m1 m8
# Now nM is M7, nQ is m8.
invert m8 m9
mult m7 m9 m9
mult m9 m7 m10
# M10 is nC
exec 'choleski'
# M15 is A, A'A=nC, A upper triangular.
brief 1
regr ck2 k1 c1-ck1;
rmatrix m16.
# M16 is U, U'U=X'X, U upper triangular.
invert m16 m9
mult m1 m9 m9
mult m9 m15 m4
# M4 is V, V'V=nC, V in column space of X
# Calculate pseudovalues
mult m4 'coeffs' c108
let'pseudos'=c108+k3*sqrt(k4-k2)*'eta(x,r)'/sqrt(ssq('eta(x,r)'))
erase c108
outfile 'biprint';
noterm.
# Print data
print ck2 c1-ck1
# Collect iteration results for printing
copy c200-ck12 into m3
trans m3 m3
erase c200-ck12
let k14=299+k2
copy m3 into c300-ck14
name c300 'theta0' c301 'theta1' c302 'theta2' c303 'theta3' c304
'theta4' c305 'theta5'
name c306 'theta6' c307 'theta7' c308 'theta8' c309 'theta9' c310
'theta10'
print 'iter.no.' 'obj' 'scales' c300-ck14
# Read columns of V into original data columns, for final
regression
let k10=152+k2
copy 'ones' c1-ck1 into c153-ck10
copy m4 into 'ones' c1-ck1
# Regress pseudo-values on V
brief 2
regress 'pseudos' k2 'ones'  c1-ck1;
noconstant;
vif;
xpxinv m3;
coeff 'coeffs'.
mult m1 'coeffs' 'fits'
copy c153-ck10 into 'ones' c1-ck1
# Calculate final scale estimate
```

```
let 'errors'=ck2-'fits'
exec 'sigmahat'
# Calculate k13 = final value of objective function
mult m2 'eta(x,r)' c108
let k13=sqrt(ssq(c108)/k4)
erase c108
Note: Print final objective, final scale estimate, turning point in
psi-function
print k13 k3 k6
# Calculate studentized residuals (=errors/sigmahat)
let 'studtres'='errors'/k3
# Calculate est. cov. matrix of estimates
invert m10 m11
trans m11 m12
add m11 m12 m11
mult m11 .5 m11
let k21=k3*k3
mult m11 k21 m12
Note: Est. cov. matrix of thetahat:
print m12
# Calculate est. corr. matrix of estimates
diag m12 c108
let c108=1/sqrt(c108)
diag c108 m13
mult m13 m12 m9
mult m9 m13 m9
Note: Est. corr. matrix of thetahat:
print m9
Note:  The est. cov matrix of thetahat is (s**2)*(nC-inverse).
Note:  nC-inverse and the upper triangular root of nC are:
print m11 m15
Note:  The root V, of nC, which lies in the column space of X, is:
print m4
Note:  X'X-Inv and Upper triangular R-matrix:
print m3 m16
Print 'hatdiags' 'fits' 'errors' 'studtres' 'w(x)' 'psi(r/w)'
set 'index'
 1:k4
end
let k16=1
exec 'biplots' k1 times
plot 'studtres' vs 'fits'
plot 'studtres' vs 'index'
info
end


SUBROUTINE 'BIPLOTS'


plot 'studtres' vs ck16
let k16=k16+1
end
```

SUBROUTINE 'CHOLESKI'

```
# This program computes a Choleski decomposition - a p by p, upper
# triangular matrix R, in M15, such that R'R=S, where the positive
definite input
# matrix S is in M10.  The value of p is in k2.
# The matrix S is first symmetrized, in case numerical errors have
# made it slightly asymmetric.
# The iterative steps are carried out by CHOLITER.
trans m10 m9
add m10 m9 m9
mult .5 m9 m9
let k10=150+k2
copy m9 into c151-ck10
let k14=sqrt(c151(1))
define k14 into a 1 by 1 matrix m15
let k15=k2-1
let k16=1
# Now m15 is a root of the upper left corner of S, and is a k16 by
k16
# matrix.  Next p-1=k15 iterations will be performed, enlarging m15
to
# a p by p matrix.
exec 'choliter' k15 times
end
```

SUBROUTINE 'CHOLITER'

```
let k17=151+k16
let k18=k16+1
let k19=ck17(k18)
copy ck17 into ck17;
use 1:k16.
# Now ck17 is the first i(=k16) elements of the (i+1)th column of
S,
# and k19 is the (i+1)th element.
trans m15 m16
invert m16 m17
mult m17 ck17 ck17
let k19=sqrt(k19-ssq(ck17))
stack ck17 k19 ck17
# Now ck17 is the (i+1)th column of the enlarged root m15: i+1 by
i+1.
let k20=k17-1
copy m16 into c151-ck20
set c108
k16(0)
end
copy c151-ck20 c108 into m16
trans m16 m16
```

```
# Now m16 is the old m15, augmented by a row of 0's.
copy m16 into c151-ck20
copy c151-ck20 ck17 into m15
# Now m15 is the enlarged root
let k16=k16+1
end
```

APPENDIX III: MINITAB CODE FOR HUBER-TYPE REGRESSION

SUBROUTINE 'REGDRIVE'

```
noecho
outfile '-temp';
noterm.
#   This program drives a robust regression package, using
# pseudo-values.  It requires the data file to have X1 in C1,...,
# Xk1 in Ck1, Y in Ck2, where k2=p.  The next statements get the
data.
# Set k4=n.
retr 'stack'
let k1=3 #k1 is number of columns of regressors in data file.
let k2=4 #k2 is p, the number of parameters, and column containing
y.
let k4=21 # k4 is n, the number of cases.
# Now set the initial scale estimate, and initial coefficient
vector.
let k3=1.26134
name c101 'coeffs' c102 'ones'
set 'coeffs'
-34.5 .71429 .35714 .0000
end
#  Now  create the X-matrix.  For regression through the origin,
omit 'ones'.
set 'ones'
k4(1)
end
copy 'ones', c1-ck1 into m1
# Now get initial vector of (unstandardized) errors
name c103 'fits' c104 'errors'
mult m1 'coeffs' 'fits'
let 'errors'=ck2-'fits'
# Set iteration number
let k11=0
name c105 'studtres' c106 'psi(r)' c107 'pseudos' c199 'scales'
c198 'iter.no.' c109 'obj' c110 'hatdiags' c111 'index'
# Collect iter.no., scale estimate, coefficients
let 'iter.no.'=k11
let 'scales'=k3
let c200='coeffs'
# Calculate studentized residuals (=errors/sigmahat)
let 'studtres'='errors'/k3
# Set k6, and calculate psi=Huber's psi(k6)
let k6=1.5
let k7=-k6
rmax 'studtres' k7 'psi(r)'
rmin 'psi(r)' k6 'psi(r)'
 # Calculate objective=norm of X-transpose psi, over root n.
```

```
 trans m1 m2
 mult m2 'psi(r)' c108
 let 'obj'=sqrt(ssq(c108)/k4)
 erase c108
#  Set number of iterations, for computation of M-estimates
let k5=10
exec 'robreg1' k5 times
exec 'robreg2'
echo
nooutfile
end


SUBROUTINE 'ROBREG1'


# Calculate k8=sum(psi**2), k9=sum(psiprime) k10=lambda
let k8=ssq('psi(r)')
copy 'psi(r)' into c108;
omit 'psi(r)'=k6,k7.
let k9=count(c108)
let k10=1+(k2/k9)-(k2/k4)
erase c108
# Calculate pseudo-values
let 'pseudos'='fits'+(k10*k3*k4/k9)*'psi(r)'
# Regress pseudo-values on X
brief 1
regress 'pseudos' on k1 in c1-ck1 c108 'fits';
coeff 'coeffs'.
erase c108
# Calculate scale estimate
let 'errors'=ck2-'fits'
let c108=abso('errors')
sort c108 c108
let k17=k2-1
let k18=k4-k17
set c160
k17(0) k18(1)
end
copy c108 c108;
omit c160=0.
let k3=1.4826*medi(c108)
erase c108
#  Collect iteration number, scale estimate, coefficients
let k11=k11+1
stack 'iter.no.' k11 'iter.no.'
stack 'scales' k3 'scales'
# Calculate studentized residuals (=errors/sigmahat)
let 'studtres'='errors'/k3
# Set k6, and calculate psi=Huber's psi(k6)
let k6=1.5
let k7=-k6
rmax 'studtres' k7 'psi(r)'
rmin 'psi(r)' k6 'psi(r)'
```

```
# Calculate objective=norm of X-transpose psi, over root n.
trans m1 m2
mult m2 'psi(r)' c108
let k13=sqrt(ssq(c108)/k4)
erase c108
stack 'obj' k13 'obj'
let k12=200+k11
let ck12='coeffs'
end
let 'studtres'='errors'/k3
end
```

SUBROUTINE 'ROBREG2'

```
outfile 'robprint';
noterm.
# Print data
print ck2 c1-ck1
# Collect iteration results for printing
copy c200-ck12 into m3
trans m3 m3
erase c200-ck12
let k14=299+k2
copy m3 into c300-ck14
name c300 'theta0' c301 'theta1' c302 'theta2' c303 'theta3'
print 'iter.no.' 'obj' 'scales' c300-ck14
# Set k6, and calculate psi=Huber's psi(k6)
let k6=1.5
let k7=-k6
rmax 'studtres' k7 'psi(r)'
rmin 'psi(r)' k6 'psi(r)'
# Calculate k8=sum(psi**2), k9=sum(psiprime) k10=lambda
let k8=ssq('psi(r)')
copy 'psi(r)' into c108;
omit 'psi(r)'=k6,k7.
let k9=count(c108)
let k10=1+(k2/k9)-(k2/k4)
erase c108
# Calculate pseudo-values
let 'pseudos'='fits'+(k10*k3*k4/k9)*'psi(r)'
# Regress pseudo-values on X
brief 2
regress 'pseudos' on k1 in c1-ck1 c108 'fits';
vif;
xpxinv m3;
rmatrix m4;
hi 'hatdiags';
coeff 'coeffs'.
erase c108
# Calculate final scale estimate
let 'errors'=ck2-'fits'
let c108=abso('errors')
```

```
sort c108 c108
let k17=k2-1
let k18=k4-k17
set c160
k17(0) k18(1)
end
copy c108 c108;
omit c160=0.
let k3=1.4826*medi(c108)
erase c108
# Calculate k13 = final value of objective function
mult m2 'psi(r)' c108
let k13=sqrt(ssq(c108)/k4)
erase c108
Note: Print final objective, final scale estimate, turning point in
psi-funcion
print k13 k3 k6
Note: Print estimates of E(psiprime), E(psi**2), V(psi,F), lambda
let k16=k9/k4 # k16 estimates E(psiprime)
let k17=k8/(k4-k2) # k17 estimates E(psi**2)
let k18=k17/(k16*k16) # k18 estimates V(psi,F)
print k16 k17 k18 k10
# Calculate studentized residuals (=errors/sigmahat)
let 'studtres'='errors'/k3
Note: Print X'X-Inv and Upper triangular R-matrix
print m3 m4
Print 'hatdiags' 'fits' 'errors' 'studtres'
set 'index'
1:k4
end
plot 'studtres' vs c1
plot 'studtres' vs c2
plot 'studtres' vs c3
plot 'studtres' vs 'fits'
plot 'studtres' vs 'index'
info
end
```

APPENDIX IV: MINITAB OUTPUT, 'STACKLOSS' DATA

RUN 1: LEAST SQUARES, FULL DATA

```
MTB > retr 'stack'
MTB > info

COLUMN     NAME      COUNT
C1         rate         21
C2         temp         21
C3         acid         21
C4         loss         21

CONSTANTS USED: NONE

MTB > regr c4 3 c1-c3

The regression equation is
loss = - 39.9 + 0.716 rate + 1.30 temp - 0.152 acid

Predictor       Coef          Stdev      t-ratio
Constant      -39.92          11.90        -3.36
rate          0.7156         0.1349         5.31
temp          1.2953         0.3680         3.52
acid         -0.1521         0.1563        -0.97

s = 3.243       R-sq = 91.4%      R-sq(adj) = 89.8%

Analysis of Variance

SOURCE        DF          SS           MS
Regression     3     1890.41       630.14
Error         17      178.83        10.52
Total         20     2069.24

SOURCE        DF      SEQ SS
rate           1     1750.12
temp           1      130.32
acid           1        9.97

Unusual Observations
Obs.     rate      loss        Fit Stdev.Fit  Residual    St.Resid
 21      70.0    15.000     22.238     1.730    -7.238      -2.64R

R denotes an obs. with a large st. resid.
```

RUN 2: LEAST SQUARES, OBSERVATIONS 1,3,4,21 OMITTED

```
MTB > retr 'small'
MTB > regr c4 3 c1-c3

The regression equation is
loss = - 37.7 + 0.798 rate + 0.577 temp - 0.0671 acid

Predictor        Coef          Stdev     t-ratio
Constant      -37.652          4.732       -7.96
rate          0.79769        0.06744       11.83
temp           0.5773         0.1660        3.48
acid         -0.06706        0.06160       -1.09

s = 1.253       R-sq = 97.5%     R-sq(adj) = 96.9%

Analysis of Variance

SOURCE        DF          SS           MS
Regression     3      795.83       265.28
Error         13       20.40         1.57
Total         16      816.24

SOURCE        DF      SEQ SS
rate           1      775.48
temp           1       18.49
acid           1        1.86

Unusual Observations
Obs.     rate       loss       Fit Stdev.Fit   Residual    St.Resid
 10      58.0     11.000    13.506     0.552     -2.506      -2.23R

R denotes an obs. with a large st. resid.
```

RUN 3: HUBER'S PSI(C=2*SQRT(P/N))

ROW	loss	rate	temp	acid
1	42	80	27	89
2	37	80	27	88
3	37	75	25	90
4	28	62	24	87
5	18	62	22	87
6	18	62	23	87
7	19	62	24	93
8	20	62	24	93
9	15	58	23	87
10	14	58	18	80
11	14	58	18	89
12	13	58	17	88
13	11	58	18	82
14	12	58	19	93
15	8	50	18	89
16	7	50	18	86
17	8	50	19	72
18	8	50	19	79
19	9	50	20	80
20	15	56	20	82
21	15	70	20	91

ROW	iter.no.	obj	scales	theta0	theta1	theta2	theta3
1	0	114.157	1.26134	-34.5000	0.714290	0.357140	0.000000
2	1	39.540	1.88936	-35.9831	0.682571	0.735645	-0.046733
3	2	46.999	2.25600	-37.4706	0.761180	0.772216	-0.091118
4	3	13.549	2.33615	-38.7372	0.788412	0.804009	-0.100049
5	4	7.936	2.29488	-39.2482	0.810345	0.794780	-0.106317
6	5	2.665	2.22089	-39.3788	0.821040	0.778932	-0.107814
7	6	0.757	2.18417	-39.3387	0.826226	0.765743	-0.108485
8	7	0.187	2.18341	-39.3167	0.828139	0.759567	-0.108514
9	8	0.143	2.18419	-39.3214	0.828593	0.759017	-0.108629
10	9	0.045	2.18461	-39.3257	0.828701	0.759004	-0.108642

```
   11        10      0.029   2.18479  -39.3271   0.828764   0.759009
-0.108668


 The regression equation is
 pseudos = - 39.3 + 0.829 rate + 0.759 temp - 0.109 acid

 Predictor       Coef          Stdev     t-ratio        VIF
 Constant     -39.328          8.447       -4.66
 rate         0.82879        0.09576        8.65        2.9
 temp          0.7590         0.2613        2.90        2.6
 acid         -0.1087         0.1110       -0.98        1.3

 s = 2.303        R-sq = 95.1%     R-sq(adj) = 94.2%

 Analysis of Variance

 SOURCE        DF          SS          MS
 Regression     3     1736.39      578.79
 Error         17       90.17        5.30
 Total         20     1826.55


 SOURCE        DF      SEQ SS
 rate           1     1686.55
 temp           1       44.75
 acid           1        5.09


 MTB > Note: Print final objective, final scale estimate, turning
point in psi-f
 MTB > uncion
 K13      0.0287331
 K3       2.18489
 K6       0.872872
 MTB > Note: Print estimates of E(psiprime), E(psi**2), V(psi,F),
lambda
 K16      0.714286
 K17      0.489492
 K18      0.959404
 K10      1.07619
 MTB > Note: Print X'X-Inv and Upper triangular R-matrix
  MATRIX M3

   13.4527    0.0273   -0.0620   -0.1594
    0.0273    0.0017   -0.0035   -0.0007
   -0.0620   -0.0035    0.0129    0.0000
   -0.1594   -0.0007    0.0000    0.0023
```

MATRIX M4

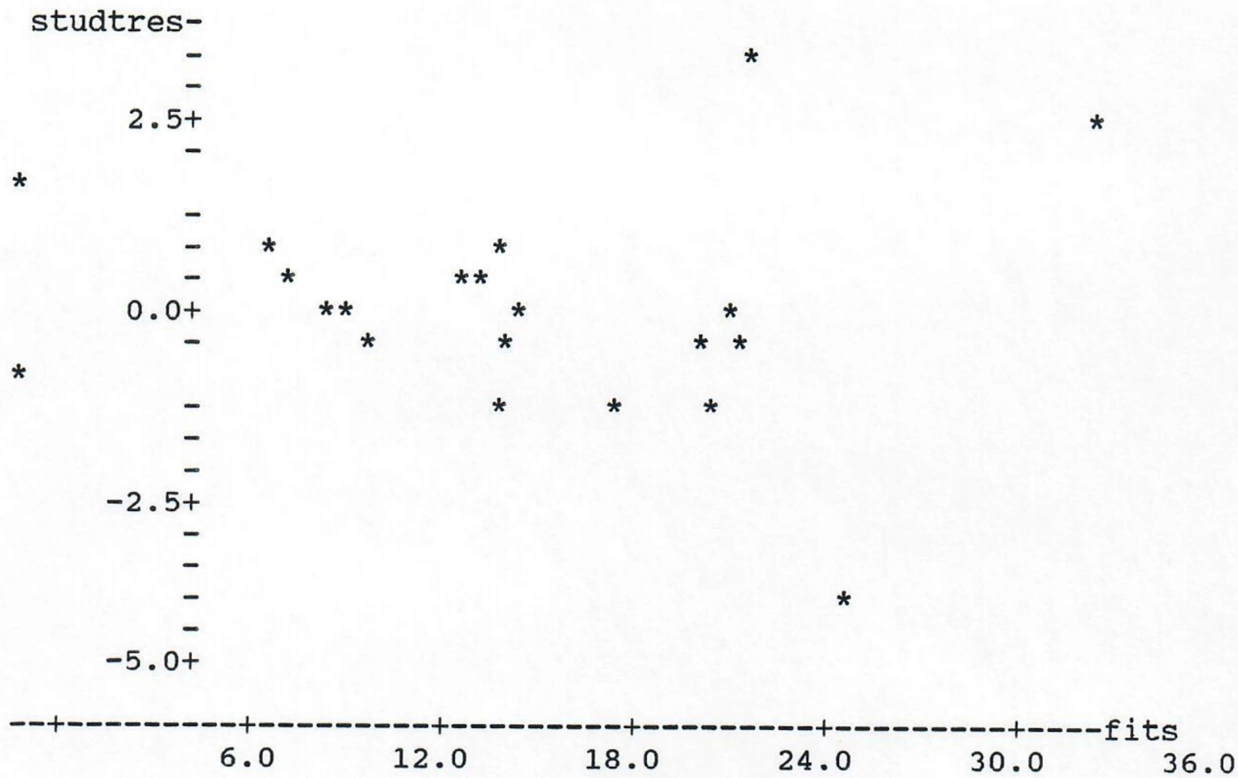
4.583	276.918	96.671	395.411
0.000	-41.002	-11.052	-11.986
0.000	0.000	-8.813	0.004
0.000	0.000	0.000	-20.752

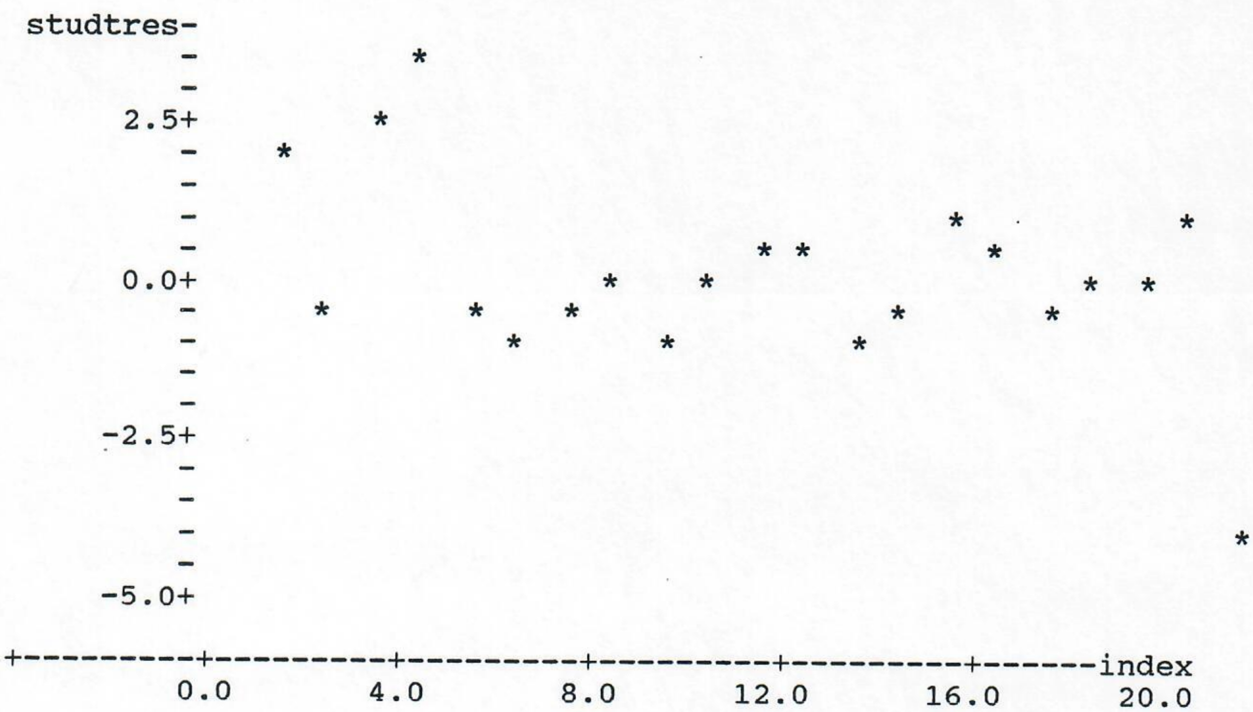
ROW	hatdiags	fits	errors	studtres
1	0.301555	37.7964	4.20360	1.92394
2	0.317841	37.9051	-0.90506	-0.41424
3	0.174615	32.0258	4.97423	2.27665
4	0.128505	20.8186	7.18144	3.28686
5	0.052220	19.3005	-1.30054	-0.59524
6	0.077487	20.0596	-2.05956	-0.94263
7	0.219237	20.1665	-1.16655	-0.53392
8	0.219237	20.1665	-0.16655	-0.07623
9	0.140184	16.7444	-1.74442	-0.79840
10	0.200044	13.7100	0.28996	0.13271
11	0.155033	12.7320	1.26799	0.58035
12	0.217176	12.0817	0.91834	0.42031
13	0.157531	13.4927	-2.49270	-1.14088
14	0.205823	13.0563	-1.05634	-0.48348
15	0.190465	6.1017	1.89827	0.86882
16	0.131074	6.4277	0.57226	0.26192
17	0.412123	8.7081	-0.70813	-0.32410
18	0.160593	7.9474	0.05256	0.02406
19	0.174537	8.5978	0.40222	0.18409
20	0.080186	13.3532	1.64684	0.75374
21	0.284533	23.9781	-8.97810	-4.10917

```
 studtres-
         -                        *
         -
      2.5+                                              *
         -                                                      *
         -
         -    *         *
         -    *             2
      0.0+    2             *     *
         -    *             *     2                             *
         -                  2     *
         -
         -
     -2.5+
         -
         -
         -                                      *
         -
     -5.0+

+---------+---------+---------+---------+---------+------rate
        48.0      54.0      60.0      66.0      72.0      78.0
```

```
 studtres-
         -                                       *
         -
      2.5+                                            *
         -                                                      *
         -
         -         *         *
         -    *    2
      0.0+         *    *    *                   *
         -              2              *         *              *
         -         *                        2
         -
         -
     -2.5+
         -
         -
         -                   *
         -
     -5.0+

--------+---------+---------+---------+---------+--------temp
                18.0      20.0      22.0      24.0      26.0
```

```
 studtres-
         -                                         *
         -
      2.5+                                                *
         -                                              *
         -
         -                            *                 *
         -                                      *    *  *
      0.0+                     * 2
*
         -   *                                      * *
2
         -                            *            2
         -
         -
     -2.5+
         -
         -
         -                                                   *
         -
     -5.0+

--+---------+---------+---------+---------+---------+----acid
          72.0      76.0      80.0      84.0      88.0      92.0
```


studtres
2.5
0.0
-2.5
-5.0
fits
6.0
12.0
18.0
24.0
30.0
36.0


studtres
2.5
0.0
-2.5
-5.0
index
0.0
4.0
8.0
12.0
16.0
20.0

RUN 4: HUBER'S PSI(C=1.5)

ROW	iter.no.	obj	scales	theta0	theta1	theta2	theta3
1	0	161.873	1.26134	-34.5000	0.714290	0.35714	0.000000
2	1	78.084	2.41800	-36.6511	0.675739	0.88134	-0.068882
3	2	47.089	2.78667	-38.7861	0.752035	1.01225	-0.124487
4	3	7.679	2.93967	-40.7194	0.794273	1.05278	-0.137937
5	4	0.256	3.00462	-41.0816	0.798360	1.04942	-0.135215
6	5	0.094	3.00742	-41.0715	0.796693	1.05469	-0.135430
7	6	0.022	3.00648	-41.0665	0.796257	1.05603	-0.135516
8	7	0.002	3.00626	-41.0652	0.796201	1.05618	-0.135530
9	8	0.001	3.00622	-41.0651	0.796198	1.05619	-0.135532
10	9	0.001	3.00623	-41.0651	0.796199	1.05618	-0.135532
11	10	0.001	3.00623	-41.0651	0.796199	1.05618	-0.135531

The regression equation is
pseudos = - 41.1 + 0.796 rate + 1.06 temp - 0.136 acid

Predictor	Coef	Stdev	t-ratio	VIF
Constant	-41.07	10.79	-3.81	
rate	0.7962	0.1223	6.51	2.9
temp	1.0562	0.3338	3.16	2.6
acid	-0.1355	0.1418	-0.96	1.3

s = 2.942 R-sq = 92.9% R-sq(adj) = 91.6%

Analysis of Variance

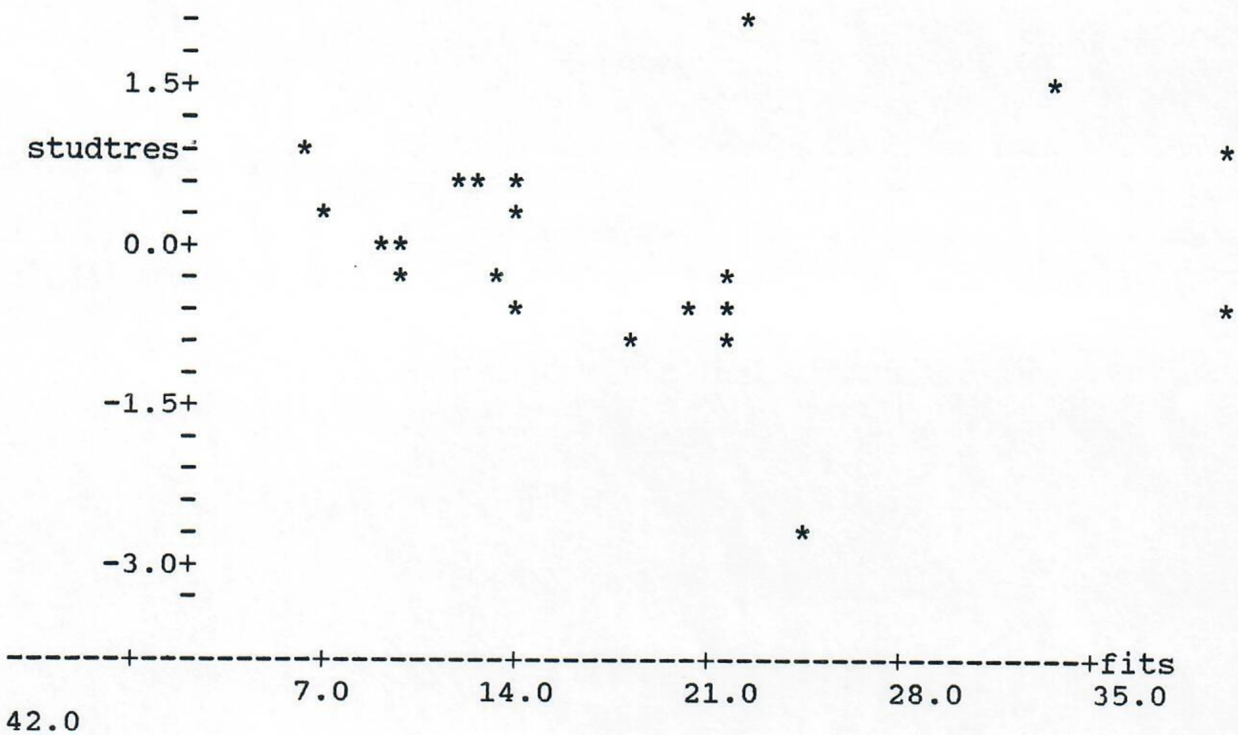
SOURCE	DF	SS	MS
Regression	3	1917.32	639.11
Error	17	147.13	8.65
Total	20	2064.45	

```
SOURCE       DF       SEQ SS
rate          1      1822.76
temp          1        86.65
acid          1         7.91

Unusual Observations
Obs.     rate    pseudos        Fit  Stdev.Fit  Residual    St.Resid
 21      70.0     18.375     23.459      1.569    -5.084      -2.04R

R denotes an obs. with a large st. resid.

MTB > Note: Print final objective, final scale estimate, turning
point in psi-f
MTB > uncion
K13       0.000907157
K3        3.00623
K6        1.50000
MTB > Note: Print estimates of E(psiprime), E(psi**2), V(psi,F),
lambda
K16       0.904762
K17       0.753406
K18       0.920366
K10       1.02005
MTB > Note: Print X'X-Inv and Upper triangular R-matrix
 MATRIX M3

  13.4527     0.0273    -0.0620    -0.1594
   0.0273     0.0017    -0.0035    -0.0007
  -0.0620    -0.0035     0.0129     0.0000
  -0.1594    -0.0007     0.0000     0.0023


 MATRIX M4

    4.583    276.918     96.671    395.411
    0.000    -41.002    -11.052    -11.986
    0.000      0.000     -8.813      0.004
    0.000      0.000      0.000    -20.752


 ROW   hatdiags       fits     errors   studtres

   1   0.301555    39.0854    2.91455    0.96950
   2   0.317841    39.2210   -2.22099   -0.73880
   3   0.174615    32.8566    4.14343    1.37828
   4   0.128505    21.8564    6.14362    2.04363
   5   0.052220    19.7440   -1.74402   -0.58013
   6   0.077487    20.8002   -2.80020   -0.93147
   7   0.219237    21.0432   -2.04320   -0.67965
   8   0.219237    21.0432   -1.04320   -0.34701
   9   0.140184    17.6154   -2.61540   -0.86999
  10   0.200044    13.2832    0.71678    0.23843
  11   0.155033    12.0634    1.93657    0.64418
  12   0.217176    11.1428    1.85722    0.61779
```

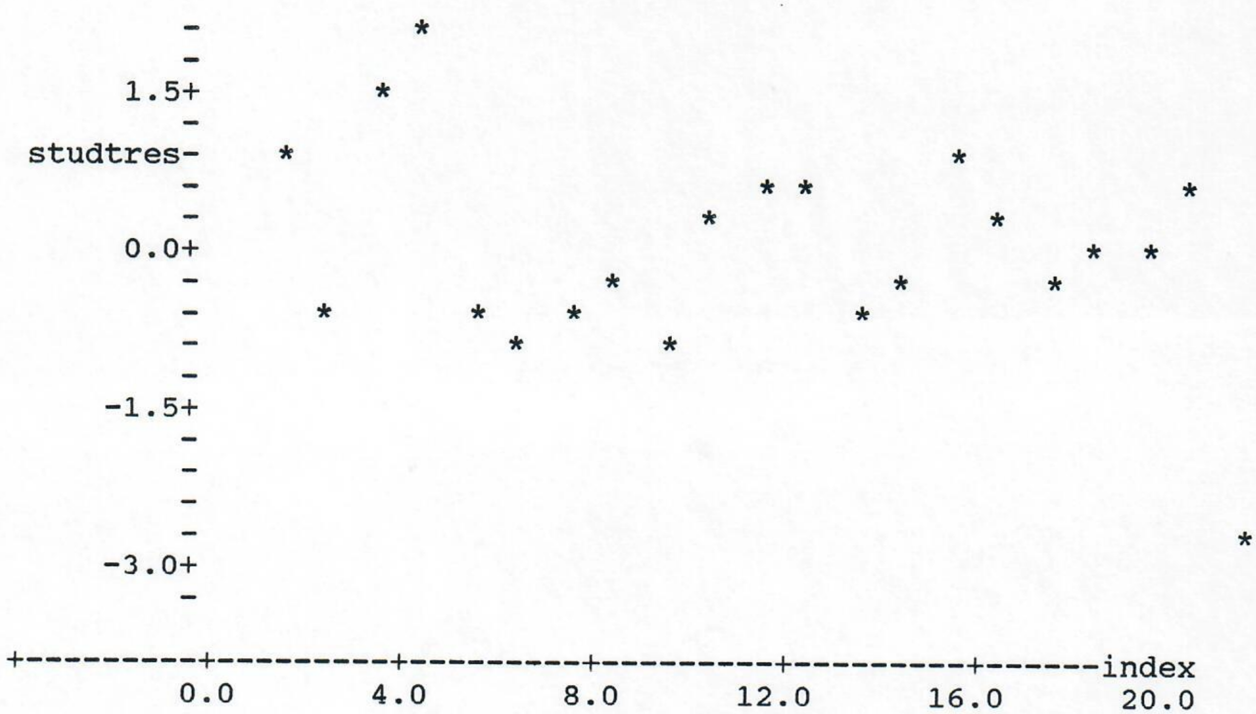
13	0.157531	13.0122	-2.01215	-0.66933
14	0.205823	12.5775	-0.57749	-0.19210
15	0.190465	5.6938	2.30616	0.76713
16	0.131074	6.1004	0.89957	0.29923
17	0.412123	9.0541	-1.05406	-0.35062
18	0.160593	8.1053	-0.10534	-0.03504
19	0.174537	9.0260	-0.02599	-0.00865
20	0.080186	13.5321	1.46788	0.48828
21	0.284533	23.4591	-8.45912	-2.81386

```
                   -                        *
                   -
                1.5+                                              *
                   -
           studtres-    *                                                 *
                   -              *   2
                   -    *             *
                0.0+    2
                   -    *             *     *
                   -                  *     2                             *
                   -                  *     *
                   -
               -1.5+
                   -
                   -
                   -
                   -                                      *
               -3.0+
                   -

         +---------+---------+---------+---------+---------+------rate
                 48.0      54.0      60.0      66.0      72.0      78.0
```

```
         -                                       *
         -
      1.5+                                            *
         -
 studtres-         *                                            *
         -    *    *         *
         -         2
      0.0+              *    *
         -              2                        *
         -         *                   *         *              *
         -                                  2
         -
     -1.5+
         -
         -
         -
         -                   *
     -3.0+
         -

--------+---------+---------+---------+---------+--------temp
                18.0      20.0      22.0      24.0      26.0

         -                                        *
         -
      1.5+                                               *
         -
 studtres-                                             2
         -                            *             *  *
         -                       *             *
      0.0+                     * *
         -   *
2
         -                            *            * *
*
         -                                        2
         -
     -1.5+
         -
         -
         -
         -                                                  *
     -3.0+
         -

--+---------+---------+---------+---------+---------+----acid
          72.0      76.0      80.0      84.0      88.0      92.0
```



studtres
1.5+
0.0+
-1.5+
-3.0+
fits
7.0
14.0
21.0
28.0
35.0
42.0



studtres
1.5+
0.0+
-1.5+
-3.0+
index
0.0
4.0
8.0
12.0
16.0
20.0

RUN 5: BOUNDED INFLUENCE, V(X)=1-H(I), PSI(C=2*SQRT(P/N))

ROW	iter.no.	obj	scales	theta0	theta1	theta2	theta3
1	0	*	2.00000	-40.0000	0.700000	1.30000	-0.150000
2	1	50.8550	2.72727	-39.0431	0.776172	0.93665	-0.121544
3	2	31.8595	2.31058	-39.6708	0.811030	0.81432	-0.106297
4	3	2.4277	2.17312	-39.2700	0.826304	0.75293	-0.106434
5	4	0.3852	2.15099	-38.9723	0.831358	0.72844	-0.107452
6	5	0.3299	2.13412	-38.8911	0.832276	0.72207	-0.107500
7	6	0.2833	2.12592	-38.8527	0.832531	0.71944	-0.107501
8	7	0.1733	2.12183	-38.8342	0.832603	0.71825	-0.107488
9	8	0.0953	2.11978	-38.8251	0.832624	0.71769	-0.107477
10	9	0.0502	2.11873	-38.8205	0.832630	0.71741	-0.107470
11	10	0.0259	2.11819	-38.8181	0.832632	0.71727	-0.107466

* NOTE * ones is highly correlated with other predictor variables
* NOTE * rate is highly correlated with other predictor variables
* NOTE * temp is highly correlated with other predictor variables
* NOTE * acid is highly correlated with other predictor variables

The regression equation is
pseudos = - 38.8 ones + 0.833 rate + 0.717 temp - 0.107 acid

Predictor	Coef	Stdev	t-ratio
Noconstant			
ones	-38.819	3.883	-10.00
rate	0.8326	0.1106	7.53
temp	0.7174	0.2258	3.18
acid	-0.10747	0.06145	-1.75

s = 2.118
Analysis of Variance

```
SOURCE          DF          SS           MS
Regression       4      6453.3       1613.3
Error           17        76.3          4.5
Total           21      6529.6

SOURCE          DF      SEQ SS
ones             1      5251.6
rate             1      1142.0
temp             1        46.0
acid             1        13.7


 MTB > Note: Print final objective, final scale estimate, turning
point in psi-f
 MTB > unction
 K13        0.0131210
 K3         2.11841
 K6         0.872872
 MTB > Note: Est. cov. matrix of thetahat:
  MATRIX M12

   15.0773   -0.0497   -0.1172   -0.1205
   -0.0497    0.0122   -0.0165   -0.0036
   -0.1172   -0.0165    0.0510    0.0002
   -0.1205   -0.0036    0.0002    0.0038


 MTB > Note: Est. corr. matrix of thetahat:
  MATRIX M9

   1.00000  -0.11576  -0.13368  -0.50478
  -0.11576   1.00000  -0.66087  -0.52466
  -0.13368  -0.66087   1.00000   0.01414
  -0.50478  -0.52466   0.01414   1.00000


 MTB   >   Note:     The   est.   cov   matrix   of   thetahat   is
(s**2)*(nC-inverse).
 MTB > Note:  nC-inverse and the upper triangular root of nC are:
  MATRIX M11

   3.35973  -0.01108  -0.02612  -0.02685
  -0.01108   0.00273  -0.00368  -0.00079
  -0.02612  -0.00368   0.01137   0.00004
  -0.02685  -0.00079   0.00004   0.00084

  MATRIX M15

    6.0766  332.0554  119.4993  501.0176
    0.0000   35.1088   11.2354   32.5553
    0.0000    0.0000    9.3815   -0.4868
    0.0000    0.0000    0.0000   34.4683
```

MTB > Note: The root V, of nC, which lies in the column space of X, is:

MATRIX M4

1.326	55.702	20.044	98.820
1.326	55.702	20.044	100.481
1.326	59.983	22.108	98.665
1.326	71.115	23.005	107.795
1.326	71.115	25.134	107.686
1.326	71.115	24.070	107.740
1.326	71.115	23.005	97.829
1.326	71.115	23.005	97.829
1.326	74.540	24.018	109.033
1.326	74.540	29.340	120.387
1.326	74.540	29.340	105.439
1.326	74.540	30.405	107.045
1.326	74.540	29.340	117.065
1.326	74.540	28.276	98.849
1.326	81.390	29.237	108.024
1.326	81.390	29.237	113.007
1.326	81.390	28.173	136.315
1.326	81.390	28.173	124.688
1.326	81.390	27.108	123.082
1.326	76.252	27.186	117.821
1.326	64.265	27.366	98.348

MTB > Note: X'X-Inv and Upper triangular R-matrix:
MATRIX M3

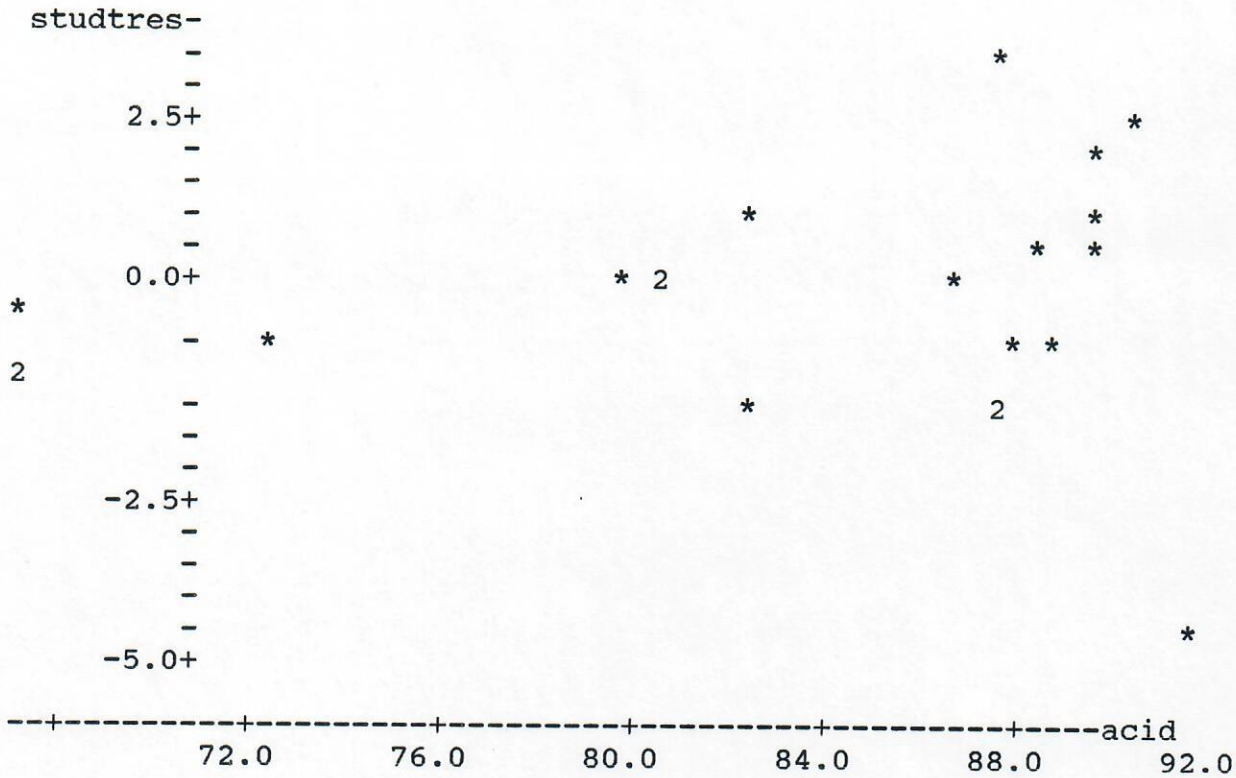
3.35965	-0.01107	-0.02614	-0.02684
-0.01107	0.00272	-0.00368	-0.00079
-0.02614	-0.00368	0.01136	0.00004
-0.02684	-0.00079	0.00004	0.00084

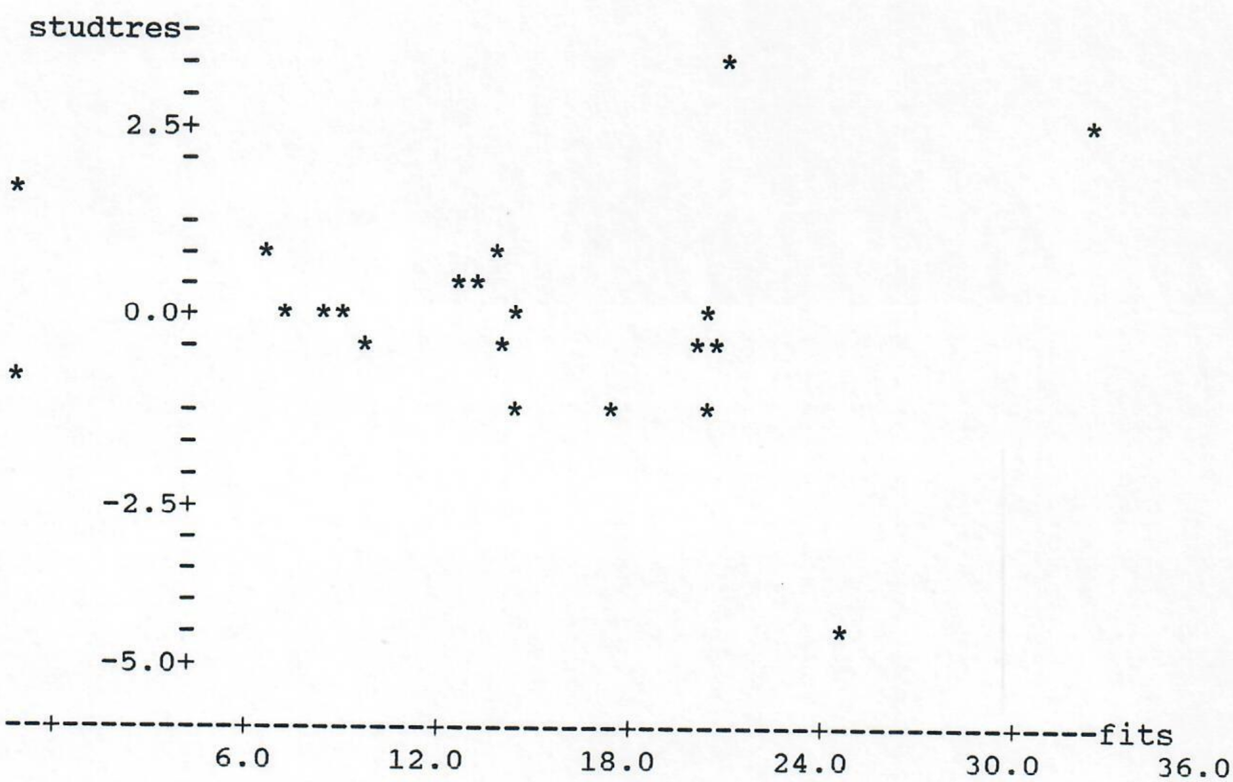
MATRIX M16

4.583	276.918	96.671	395.411
0.000	-41.002	-11.052	-11.986
0.000	0.000	-8.813	0.004
0.000	0.000	0.000	-20.752

ROW	hatdiags	fits	errors	studtres	w(x)	psi(r/w)
1	0.301555	37.5945	4.40550	2.07963	0.835730	0.872872
2	0.317841	37.7020	-0.70197	-0.33136	0.825929	-0.401016
3	0.174615	31.8892	5.11079	2.41256	0.908507	0.872872
4	0.128505	20.6701	7.32988	3.46009	0.933539	0.872872
5	0.052220	19.2354	-1.23543	-0.58319	0.973540	-0.599054
6	0.077487	19.9528	-1.95277	-0.92181	0.960475	-0.872872
7	0.219237	20.0253	-1.02533	-0.48401	0.883608	-0.547682
8	0.219237	20.0253	-0.02533	-0.01196	0.883608	-0.013395
9	0.140184	16.6223	-1.62227	-0.76580	0.927263	-0.825850
10	0.200044	13.7878	0.21221	0.10017	0.894402	0.111901
11	0.155033	12.8206	1.17941	0.55674	0.919221	0.605620
12	0.217176	12.2107	0.78930	0.37259	0.884773	0.420995
13	0.157531	13.5729	-2.57285	-1.21452	0.917861	-0.872872
14	0.205823	13.1081	-1.10808	-0.52307	0.891166	-0.587079
15	0.190465	6.1596	1.84041	0.86877	0.899742	0.872872
16	0.131074	6.4820	0.51801	0.24453	0.932162	0.262268
17	0.412123	8.7039	-0.70387	-0.33226	0.766731	-0.433442
18	0.160593	7.9516	0.04840	0.02285	0.916192	0.024897
19	0.174537	8.5615	0.43851	0.20700	0.908550	0.227861
20	0.080186	13.3423	1.65769	0.78252	0.959069	0.815979
21	0.284533	24.0319	-9.03186	-4.26351	0.845853	-0.872872

```
  studtres-
          -                       *
          -
       2.5+                                            *
          -                                                    *
          -
          -   *         *
          -                 2
       0.0+   3             *     *
          -   *             *     2                             *
          -                 2     *
          -
          -
      -2.5+
          -
          -
          -
          -                                     *
      -5.0+

+---------+---------+---------+---------+---------+------rate
        48.0      54.0      60.0      66.0      72.0      78.0

 studtres-
          -                                       *
          -
       2.5+                                            *
          -                                                      *
          -
          -         *         *
          -    *    *
       0.0+         2    *    *                   *
          -              2              *         *              *
          -         *                        2
          -
          -
      -2.5+
          -
          -
          -
          -                   *
      -5.0+

--------+---------+---------+---------+---------+--------temp
               18.0      20.0      22.0      24.0      26.0
```


studtres
2.5
0.0
-2.5
-5.0
72.0 76.0 80.0 84.0 88.0 92.0
acid


studtres
2.5
0.0
-2.5
-5.0
6.0 12.0 18.0 24.0 30.0 36.0
fits

```
 studtres-
         -           *
         -
      2.5+         *
         -    *
         -
         -                                       *           *
         -                             * *
      0.0+                     *    *              *    *  *
         -      *       *    *                *       *
         -                *       *         *
         -
         -
     -2.5+
         -
         -
         -
         -                                                      *
     -5.0+

+---------+---------+---------+---------+---------+------index
         0.0       4.0       8.0      12.0      16.0      20.0
```

RUN 6: BOUNDED INFLUENCE, V(X)=(1-H(I))/SQRT(H(I)), PSI(C=2*SQRT(P/N))

ROW	iter.no.	obj	scales	theta0	theta1	theta2	theta3
1	0	*	2.00000	-40.0000	0.700000	1.30000	-0.150000
2	1	149.216	2.59346	-40.3710	0.785249	0.97072	-0.116698
3	2	38.805	3.10400	-41.5933	0.810048	0.99303	-0.122877
4	3	11.344	3.20935	-41.8258	0.803013	1.05419	-0.129466
5	4	1.309	3.19461	-41.7521	0.799634	1.06351	-0.130342
6	5	0.055	3.19395	-41.7489	0.799491	1.06391	-0.130379
7	6	0.002	3.19393	-41.7488	0.799486	1.06392	-0.130381
8	7	0.001	3.19393	-41.7488	0.799486	1.06392	-0.130381
9	8	0.001	3.19393	-41.7488	0.799486	1.06392	-0.130381
10	9	0.001	3.19393	-41.7488	0.799486	1.06392	-0.130381
11	10	0.001	3.19393	-41.7488	0.799486	1.06392	-0.130381

* NOTE * ones is highly correlated with other predictor variables
* NOTE * rate is highly correlated with other predictor variables
* NOTE * temp is highly correlated with other predictor variables
* NOTE * acid is highly correlated with other predictor variables

The regression equation is
pseudos = - 41.7 ones + 0.799 rate + 1.06 temp - 0.130 acid

Predictor	Coef	Stdev	t-ratio
Noconstant			
ones	-41.749	5.426	-7.69
rate	0.7995	0.1442	5.55
temp	1.0639	0.3945	2.70
acid	-0.13038	0.07336	-1.78

s = 3.194
Analysis of Variance

```
SOURCE        DF        SS          MS
Regression     4    10857.5      2714.4
Error         17      173.4        10.2
Total         21    11030.9

SOURCE        DF     SEQ SS
ones           1     7862.1
rate           1     2899.3
temp           1       63.8
acid           1       32.2

Unusual Observations
Obs.    ones  pseudos       Fit Stdev.Fit  Residual   St.Resid
  4     1.66   22.157    15.107     1.145     7.051      2.36R

R denotes an obs. with a large st. resid.

MTB > Note: Print final objective, final scale estimate, turning
point in psi-f
MTB > unction
K13       0.000573914
K3        3.19393
K6        0.872872
MTB > Note: Est. cov. matrix of thetahat:
 MATRIX M12

  29.4545   -0.0778   -0.0278   -0.2990
  -0.0778    0.0208   -0.0491   -0.0014
  -0.0278   -0.0491    0.1557   -0.0035
  -0.2990   -0.0014   -0.0035    0.0054


MTB > Note: Est. corr. matrix of thetahat:
 MATRIX M9

  1.00000  -0.09947  -0.01296  -0.75074
 -0.09947   1.00000  -0.86322  -0.12943
 -0.01296  -0.86322   1.00000  -0.12094
 -0.75074  -0.12943  -0.12094   1.00000


MTB > Note: The est. cov matrix of thetahat is
(s**2)*(nC-inverse).
MTB > Note:  nC-inverse and the upper triangular root of nC are:
 MATRIX M11

  2.88735  -0.00763  -0.00272  -0.02931
 -0.00763   0.00204  -0.00482  -0.00013
 -0.00272  -0.00482   0.01527  -0.00034
 -0.02931  -0.00013  -0.00034   0.00053
```

MATRIX M15

7.58722	414.20239	145.98689	621.57544
0.00000	49.60910	16.16757	23.14246
0.00000	0.00000	8.15595	5.29031
0.00000	0.00000	0.00000	43.54021

MTB > Note: The root V, of nC, which lies in the column space of X, is:

MATRIX M4

1.656	66.706	23.557	130.522
1.656	66.706	23.557	132.621
1.656	72.756	26.132	128.572
1.656	88.485	28.941	132.725
1.656	88.485	30.792	133.928
1.656	88.485	29.867	133.326
1.656	88.485	28.941	120.136
1.656	88.485	28.941	120.136
1.656	93.325	30.446	132.483
1.656	93.325	35.073	150.176
1.656	93.325	35.073	131.292
1.656	93.325	35.999	133.992
1.656	93.325	35.073	145.979
1.656	93.325	34.148	122.298
1.656	103.004	36.232	129.605
1.656	103.004	36.232	135.899
1.656	103.004	35.307	164.672
1.656	103.004	35.307	149.985
1.656	103.004	34.381	147.286
1.656	95.745	33.512	144.355
1.656	78.806	31.484	128.425

MTB > Note: X'X-Inv and Upper triangular R-matrix:
MATRIX M3

2.88574	-0.00761	-0.00280	-0.02928
-0.00761	0.00204	-0.00481	-0.00013
-0.00280	-0.00481	0.01526	-0.00034
-0.02928	-0.00013	-0.00034	0.00053

MATRIX M16

4.583	276.918	96.671	395.411
0.000	-41.002	-11.052	-11.986
0.000	0.000	-8.813	0.004
0.000	0.000	0.000	-20.752

ROW	hatdiags	fits	errors	studtres	w(x)	psi(r/w)
1	0.301555	39.3320	2.66801	0.83534	1.27189	0.656757
2	0.317841	39.4624	-2.46237	-0.77095	1.20999	-0.637171
3	0.174615	33.0763	3.92366	1.22847	1.97522	0.621935
4	0.128505	22.0103	5.98975	1.87535	2.43111	0.771394
5	0.052220	19.8824	-1.88242	-0.58937	4.14751	-0.142105
6	0.077487	20.9463	-2.94633	-0.92248	3.31403	-0.278358
7	0.219237	21.2280	-2.22797	-0.69756	1.66749	-0.418338
8	0.219237	21.2280	-1.22797	-0.38447	1.66749	-0.230574
9	0.140184	17.7484	-2.74838	-0.86050	2.29645	-0.374716
10	0.200044	13.3415	0.65853	0.20618	1.78856	0.115276
11	0.155033	12.1680	1.83197	0.57358	2.14599	0.267276
12	0.217176	11.2345	1.76550	0.55277	1.67980	0.329065
13	0.157531	13.0807	-2.08070	-0.65146	2.12261	-0.306915
14	0.205823	12.7104	-0.71043	-0.22243	1.75053	-0.127069
15	0.190465	5.7721	2.22785	0.69753	1.85493	0.376036
16	0.131074	6.1633	0.83671	0.26197	2.40008	0.109148
17	0.412123	9.0525	-1.05255	-0.32955	0.91574	-0.359874
18	0.160593	8.1399	-0.13988	-0.04380	2.09464	-0.020911
19	0.174537	9.0734	-0.07342	-0.02299	1.97585	-0.011637
20	0.080186	13.6096	1.39043	0.43534	3.24826	0.134019
21	0.284533	23.6289	-8.62894	-2.70167	1.34129	-0.872872

```
 studres-
         -
         -                        *
     1.5+
         -                                              *
         -                                                      *
         -    *             2
         -    *         *   *
     0.0+    2
         -    *             *     *
         -                  *     2
         -                  *     *                             *
         -
    -1.5+
         -
         -
         -
         -                                      *
    -3.0+

+---------+---------+---------+---------+---------+------rate
        48.0      54.0      60.0      66.0      72.0      78.0
```

```
 studtres-
         -
         -                                       *
      1.5+
         -                                            *
         -                                                      *
         -    *    2
         -         2         *
      0.0+              *    *
         -              2                        *
         -         *                   *         *
         -                                  2                   *
         -
     -1.5+
         -
         -
         -
         -                   *
     -3.0+

--------+---------+---------+---------+---------+--------temp
                18.0      20.0      22.0      24.0      26.0


 studtres-
         -
         -                                         *
      1.5+
         -                                                *
         -                                              *
         -                                           *  2
         -                       *    *         *
      0.0+                     * *
         -   *
2
         -                             *            *
*
         -                                         2 *
         -
     -1.5+
         -
         -
         -
         -                                                   *
     -3.0+

--+---------+---------+---------+---------+---------+----acid
          72.0      76.0      80.0      84.0      88.0      92.0
```

```
 studtres-
         -
         -                            *
      1.5+
         -                                            *
         -                                                     *
         -     *       **
         -      *         2
      0.0+         **
         -          *    *           *
         -                *        * *
         -                      *    *                         *
         -
     -1.5+
         -
         -
         -
         -                               *
     -3.0+

------+---------+---------+---------+---------+---------+fits
              7.0      14.0      21.0      28.0      35.0
42.0


  studtres-
          -
          -          *
       1.5+
          -        *
          -   *
          -
          -                            * *       *
          -                         *              *         *
       0.0+                                             *  *
          -                    *              *       *
          -             *    *              *
          -     *         *       *
          -
      -1.5+
          -
          -
          -
          -                                                    *
      -3.0+

+---------+---------+---------+---------+---------+------index
         0.0       4.0       8.0      12.0      16.0      20.0
```