A NOTE ON COMPUTING ROBUST REGRESSION ESTIMATES VIA ITERATIVELY REWEIGHTED LEAST SQUARES

Raymond J. Carroll

David Ruppert

James O. Street

Raymond J. Carroll and David Ruppert are with the Department of Statistics, the University of North Carolina at Chapel Hill. James O. Street is with Boehringer Ingelheim Pharmaceuticals in Ridgefield, Connecticut. Carroll's research was supported by the Air Force Office of Scientific Research Contract AFOSR F 49620 85 C 0144. Ruppert's research was supported by the National Science Foundation Grant MCS 8100748.

ABSTRACT

The 1985 SAS User's Guide: Statistics provides a method for computing robust regression estimates using iterative reweighted least squares and the nonlinear regression procedure NLIN. We show that, while the estimates are asymptotically correct, the resulting standard errors are not. We also discuss computation of the estimates.

Section 1 : Introduction

Parameter estimates for generalized linear models (McCullagh & Nelder, 1983) and robust regression (Hampel, et al., 1986) can be computed by iteratively reweighted least squares techniques using the SAS nonlinear regression procedure NLIN. Examples of these computations are given in the 1985 SAS User's Guide: Statistics (pages 597 - 605). Despite the fact that the User's Guide makes no mention of standard errors, casual readers may assume standard errors from such a fitting algorithm are correct. In fact this is the case only for generalized linear regression models. That such an algorithm works for generalized linear regression models is shown by McCullagh & Nelder (1983); see also Carroll & Ruppert (1987) for similar results. In this section, we show that the standard errors in the SAS user's guide are inconsistent for robust regression. In section 2, we discuss computation of the estimates. In section 3, we present an example to show that the use of these standard errors can give results noticably at variance with the usual formula.

Consider an ordinary robust regression. Here the model is written as

$$y_i = x_i^t\beta + \sigma\epsilon_i,$$

where the errors $\{\epsilon_i\}$ are independent and identically distributed. For a given estimate $\hat{\sigma}$ of the scale parameter σ, the classic M-estimate of the regression parameter β solves the equation

(1) $$0 = \sum_{i=1}^{N} x_i \, \psi\left[(y_i - x_i{}^t\hat{\beta})/\hat{\sigma} \right] .$$

The M-estimators defined by equation (1) are not robust against the effects of leverage, i.e., unusual design points. For discussion of M-estimates which control for leverage, see Hampel, et al. (1986). In equation (1) the function ψ is usually bounded. Typical choices include $\psi(u) = \max(-k, \min(u,k))$ (Huber's function), Tukey's biweight as in the SAS User's Guide and the Hampel function

$$\psi(u) = -\psi(-u) = \begin{cases} u & 0 \le u < a_H \\ a_H & a_H \le u < b_H \\ a_H(c_H-u)/(c_H-b_H) & b_H \le u < c_H \\ 0 & u \ge c_H \end{cases}$$

The solution $\hat{\beta}$ is usually computed by the following algorithm, see Holland & Welsch (1977) :

(a) Begin with an initial estimate $\hat{\beta}$ of β.

(b) Form the residuals $r_i = (y_i - x_i{}^t\hat{\beta})/\hat{\sigma}$.

(c) Define weights $w_i = \psi(r_i)/r_i$.

(d) Update the estimate $\hat{\beta}$ by performing a weighted least squares regression with the weights w_i.

(e) Iterate until convergence.

Choices of $\hat{\sigma}$, which may also involve iteration, are discussed in section 2.

Let $\dot{\psi}$ be the derivative of ψ. It is well known (Bickel (1976) and Schrader & Hettmansperger (1981)) that the final estimate $\hat{\beta}$ is asymptotically normally distributed with mean β and covariance Λ_R, where

$$(2) \qquad \Lambda_R = \sigma^2 \, E\{\psi^2(\epsilon)\} \left[[E \, \dot{\psi}(\epsilon)]^2 \sum_{i=1}^{N} x_i \, x_i^t \right]^{-1} .$$

Estimated standard errors are formed by making the following substitutions, see Bickel (1976) and Schrader & Hettmansperger (1981).

$$(3.i) \qquad a = N^{-1} \sum_{i=1}^{N} \dot{\psi}(r_i) \longrightarrow E\{\dot{\psi}(\epsilon)\}; \quad \hat{\sigma} \longrightarrow \sigma ;$$

$$(3.ii) \qquad \lambda \, b \longrightarrow E\psi^2(\epsilon) \;, \text{ where}$$

$$(3.iii) \qquad b = (N-p)^{-1} \sum_{i=1}^{N} \psi^2(r_i); \qquad \lambda = 1 + (p/N)(1-a)/a .$$

The procedure NLIN treats the weights $\{w_i\}$ as if they were fixed and known a priori. This is the crux of the problem, because robust regression is one instance where the randomness of the weights is crucial. As shown in the appendix, NLIN pretends that $\hat{\beta}$ is asymptotically normally distributed with mean β and covariance Λ_{NLIN}, where

$$(4) \qquad \Lambda_{NLIN} = d \, \Lambda_R, \text{ and}$$

If one runs a linear regression replacing the responses $\{y_i\}$ by the pseudo values $\{\tilde{y}_i\}$, the estimated covariance is asymptotically correct, being

$$(5) \qquad \hat{\Lambda}_R = b \lambda (\hat{\sigma} / a)^2 \left[\sum_{i=1}^{N} x_i \, x_i^t \right]^{-1} .$$

Tests and confidence intervals using the pseudo values are also asymptotically correct. Auxiliary quantitites such as R^2 would not be meaningful when computed using pseudo values. An alternative approach to hypothesis testing is discussed by Schrader & Hettmansperger (1980).

It thus remains to consider numerical calculation of $\hat{\sigma}$ and $\hat{\beta}$. For a given value of $\hat{\sigma}$, the algorithm discussed in the first section can be employed. For a given $\hat{\beta}$, there are two common estimates of σ. The first is based on the median absolute deviation (MAD). The resulting estimate of σ is defined by

$$(6) \qquad \hat{\sigma} = \text{MAD} / .6745 = \text{median}\{ \, |y_i - x_i^t \hat{\beta}| \, \} / .6745 .$$

The division by .6745 is made so that for normally distributed data $\hat{\sigma}$ is an estimate of the standard deviation. Hill & Holland (1977) suggest that for smaller sample sizes the MAD in (6) be replaced by the modified estimator

$$\text{(Normalized) MAD} = \text{median}\{ \text{ largest } N-p+1 \text{ of the } |y_i - x_i^t \hat{\beta}| \, \} .$$

The MAD and normalized MAD are easily calculated.

An alternative estimate of σ is Huber's Proposal 2, the usual form of which is the solution to the equation

(7) $$(N-p)^{-1} \sum_{i=1}^{N} \psi^2\left[(y_i - x_i^t \hat{\beta})/\hat{\sigma} \right] = E_Z\psi^2(\epsilon) \ ,$$

where $E_Z\psi^2(\epsilon)$ is the expected value of $\psi^2(\epsilon)$ when ϵ has a standard normal distribution. The right hand side of (7) is again chosen so that for normally distributed data, $\hat{\sigma}$ estimates the standard deviation. Solving (7) requires iteration. If $\hat{\sigma}_o$ is the present estimate of σ, then the next step in the iteration is defined by

(8) $$\hat{\sigma}^2 = (N-p)^{-1} \sum_{i=1}^{N} w_i^2 \ (y_i - x_i^t \hat{\beta})^2 \ / \ E_Z\psi^2(\epsilon),$$

where as before $w_i = \psi(r_i^{(o)})/r_i^{(o)}$, with $r_i^{(o)} = (y_i - x_i^t \hat{\beta})/\hat{\sigma}_o$.

In practice, one would calculate σ and β for a fixed number of iterations or until convergence. The calculation is easily programmed in any matrix language such as SAS/IML, GAUSS or APL.

SECTION 3 : An Example

To illustrate the foregoing remarks, we computed estimates and standard errors for the data used in the SAS/NLIN example, namely a regression of the US population against time. The model is

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \sigma \epsilon_i ,$$

where $\{y_i\}$ is the US population in millions at year $t_i = 1780 + 10i$, $i = 1,\ldots,19$. Rather than use the actual year, we centered and standardized so that

$$x_i = (t_i - 1880)/90 .$$

As with any polynomial model with equally spaced time points, there is a bit of a problem with leverage here, since the highest leverage value is 0.38. However, we will procede with the usual analyses. We computed parameter estimates and standard errors using least squares, the Huber method with $\psi(x) = \max(-1.25, \min(x,1.25))$ and the Hampel method with $a=1.25$, $b=3.5$ and $c=8.0$. The robust methods computed estimates of σ by Proposal 2, see equations (7) and (8). The results of the calculations are given in Table 1. For purposes of comparison, we also reproduce the results using the Tukey biweight as in the SAS/NLIN manual, where $\hat{\sigma} = 2$. The SAS/NLIN standard errors are about 30% larger than our estimates when using the Huber method, while they are about 20% smaller for the Hampel method.

APPENDIX :

If we pretend that the weights are fixed, then the estimated covariance matrix is Λ, where

$$\Lambda = (N-p)^{-1} \sum_{i=1}^{N} w_i \, [y_i - x_i{}^t\hat{\beta}]^2 \left[\sum_{i=1}^{N} x_i \, x_i{}^t \, w_i \right]^{-1} .$$

By standard asymptotic theory,

$$(N-p)^{-1} \sum_{i=1}^{N} w_i \, [y_i - x_i{}^t\hat{\beta}]^2 = \hat{\sigma}^2 \, (N-p)^{-1} \sum_{i=1}^{N} \psi(r_i) \, r_i \xrightarrow{p} \sigma^2 \, E\{\epsilon\psi(\epsilon)\} \; ;$$

and

$$N^{-1} \sum_{i=1}^{N} x_i \, x_i{}^t \, w_i \; - \; N^{-1} \sum_{i=1}^{N} x_i \, x_i{}^t \, E\{\psi(\epsilon)/\epsilon\} \xrightarrow{p} 0 \; .$$

This verifies (4).

REFERENCES :

Bickel, P. J. (1976). Another look at robustness: A review of reviews and some new developments. Scandinavian Journal of Statistics 3, 145-168.

Carroll, R. J. & Ruppert, D. (1987). Transformations and Weighting In Regression. Chapman & Hall, New York and London.

Gross, A. M. (1977). Confidence intervals for bisquare regression estimates. Journal of the American Statistical Association 72, 341-351.

Hampel, F. R., Ronchetti, E., Rousseeuw, P. J. & Stahel, W. (1986). Robust Statistics: The Approach Based on Influence Functions. John Wiley & Sons, New York.

Hill, R. W. & Holland, P. W. (1977). Two robust alternatives to robust regression. Journal of the American Statistical Association 72, 828-833.

Holland, P. W. & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. Communications in Statistics - Theory and Methods 6, 813-827.

McCullagh, P. & Nelder, J. (1983). Generalized Linear Models. Chapman & Hall, New York and London.

SAS User's Guide: Statistics (1985). Version 5 edition published by the SAS Institute, Inc., Cary, NC.

Schrader, R. M. & Hettmansperger, T. P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. Biometrika 67, 93-101.

TABLE 1

Parameter Estimates and Standard Errors for the Example

	β_0	β_1	β_2
Least Squares			
Estimates	50.73	97.09	51.40
Standard Errors	0.96	1.05	1.93
Huber Method			
Estimates	50.98	98.37	52.44
Std. Err. as in (5)	0.45	0.49	0.90
Std. Err. via SAS/NLIN	0.56	0.64	1.12
Hampel Method			
Estimates	51.08	98.85	52.83
Std. Err. as in (5)	0.36	0.39	0.73
Std. Err. via SAS/NLIN	0.30	0.35	0.60
Tukey Biweight with σ = 2			
Estimates	51.14	98.82	52.68
Std. Err. as in (5)	0.39	0.43	0.79
Std. Err. via SAS/NLIN	0.35	0.41	0.71