Technical Report 89-25

ON DELETING INFLUENTIAL RESPONSES UNDER A MODEL OF ASYMMETRIC ERRORS IN REGRESSION

BY

M. Hlynka, J.N. Sheahan and D.P. Wiens

ON DELETING INFLUENTIAL RESPONSES UNDER A MODEL OF ASYMMETRIC ERRORS IN REGRESSION

by

M.Hlynka,* J.N. Sheahan,† and D.P. Wiens**

1. INTRODUCTION

This report consists of some heuristic proposals for choosing the support of re-descending ψ-functions, used in M-estimation of regression parameters. It is meant to be read in conjunction with Hlynka, Sheahan and Wiens (1990), which will be referred to as HSW.

This report comprised a section of an early version of HSW. All definitions necessary for an understanding of this report are contained in Section 1 of HSW. Some of the references are detailed in the bibliography of HSW, others in the bibliography of this report. Equation numbers refer to equations in HSW.

2. ESTIMATION OF a_0 IN (1.5).

In this section, we assume that in the linear model (1.1), F is known to belong to the class $\mathcal{F}$ of (1.5) and that ϵ and σ are known. Now if a_0 were also known, the optimal M-estimator of θ would be the solution of (1.2) with ψ given by (1.3), $\widehat{\sigma} = \sigma, y_0$ and y_1 given in (1.6) and ω as near zero as desired - see Lind, Mehra and Sheahan (1989). When a_0 is unknown an asymptotically optimal procedure is as follows: Find a consistent estimator $\widehat{a}_0$ of a_0; then solve (1.2) with $\widehat{a}_0$ replacing the unknown a_0 in (1.3). The aim of this section is to present some heuristic methods for estimating a_0. We remark that if σ is unknown one can obtain scale equivariance either by solving (1.2) with $\widehat{\sigma}$ as given there,

* Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario N9B 3P4

† ** Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta T6G 2G1

** †* Research supported in part by the Natural Sciences and Engineering Research Council of Canada

or by simultaneous estimation of θ and σ, as in Sheahan (1988).

To simplify the discussion we assume $\epsilon = 0$ and without loss of generality we assume $\sigma = 1$. Then the distribution of the i.i.d. errors is standard normal in an unknown interval $(-a_0, a_0)$ and completely unknown outside $(-a_0, a_0)$. We may assume that the error distribution is non-normal in every interval containing (a_0, a_0), for otherwise we can re-define a_0 by $a_0 = \sup\{a_0^* |$ the distribution is of standard normal form in $(-a_0^*, a_0^*)\}$. Our aim is then to estimate the point a_0 where outside $(-a_0, a_0)$ the distribution is not of standard normal form.

2.1. One heuristic procedure for estimating a_0 is to make a normal probability plot of the ordered residuals $\widehat{U}_1 \leq \ldots \leq \widehat{U}_n$, where

$$\widehat{U}_i = X_i - c_i'\tilde{\theta}, \quad i = 1, \ldots, n \tag{2.1}$$

and $\tilde{\theta}$ is an initial consistent and shift equivariant estimator of θ. Such an estimator can be constructed as follows, in the case where F has a positive and continuous density f. Let $\tilde{\theta}_L$ be the least absolute deviation estimator, defined in the Introduction to HSW. It was shown by Koenker and Bassett (1978) - see also Bassett and Koenker (1978) and Ruppert and Carroll (1980) - that

$$\tilde{\theta}_L \quad \longrightarrow_{\mathcal{P}} \theta + (F^{-1}(\frac{1}{2}), \ldots, F^{-1}(\frac{1}{2}))'.$$

If $F^{-1}(1/2) = 0$ we define $\tilde{\theta} = \tilde{\theta}_L$. If $F^{-1}(1/2) \neq 0$ but the column space of $C^{(n)}$ contains the vector $(1, \ldots, 1)'$, one can, by transforming X, arrange that $F^{-1}(1/2) = 0$. If finally the median of the error distribution is not zero and the column space of $C^{(n)}$ does not contain the unit vector, we obtain a consistent estimator by defining $\tilde{\theta}$ to be the Newton's method solution of (1.2) with $\tilde{\theta}_L$ as initial value and a ψ-function with "artificially-trimmed" support - see Lind, Mehra and Sheahan (1989) sec. 3.2, where $\tilde{\theta}$ is denoted $\tilde{\theta}(2)$.

If $(-\widehat{a}_0, \widehat{a}_0)$ is an interval within which the normal probability plot is "approximately" linear and outside of which the plot becomes non-linear, or linear with a different slope, we propose $\widehat{a}_0$ as an estimator of a_0. This procedure, which can be somewhat rigorized by a regression analysis, is rather easy to implement and provides a reasonable "starting" value for the tests that follow.

2.2 The idea here is to perform a goodness of fit test for various values of a_0: Partition $\mathbf{R}^1$ into m intervals $(-\infty, b_1), (b_1, b_2), \ldots, (b_1, b_{m-1}), (b_{m-1}, \infty)$. For any fixed a_0, let j and l be the indices for which $b_j \leq -a_0 \leq b_{j+1}$ and $b_l \leq a_0 \leq b_{l+1}$. With $\widehat{U}_i$ as defined in (2.1), let $N_1, \ldots, N_K$ be the numbers of the $\widehat{U}_i$ that will fall in the intervals $(-a_0, b_{j+1}), (b_{j+1}, b_{j+2}) \ldots, (b_l, a_0)$ respectively, and let $E(N_i)$ be the corresponding i^{th} expected frequency under normality of the error distribution in $(-a_0, a_0), i = 1, \ldots, K$. Now define

$$\chi^2(a_0) = \sum_{i=1}^{K} \frac{(N_i - E(N_i))^2}{E(N_i)}$$

and reject the hypothesis that F is normal in $(-a_0, a_0)$ if the observed value of $\chi^2(a_0)$ exceeds $\chi^2_{\alpha[K-p-1]}$, the upper $\alpha\%$ criticial point of a χ^2 distribution with $K-p-1$ degrees of freedom.

An estimator of a_0 is then

$$\widehat{a}_1 = \max\{a_0 | \chi^2(a_0) \leq \chi^2_{\alpha[K-p-1]}\}.$$

A reasonable starting value for the above tests is the $\widehat{a}_0$ of sec. 2.1. (We remark that as a variation of this method one may employ a "minimum χ^2" procedure, simultaneously estimating θ along with the testing procedure).

2.3. Since the procedure of sec. 2.2 requires an arbitrary partitioning of $\mathbf{R}^1$ into intervals, one may prefer to employ a Kolmogorov-Smirnov type of procedure instead. For fixed a_0, define

$$T(a_0) = \sup_{u \in (-a_0, a_0)} |F_n(u) - \Phi(u)|$$

where $F_n(u)$ is the empirical distribution function of the residuals (2.1). Note that the observed value $t(a_0)$ of $T(a_0)$ is often found graphically. We reject the hypothesis of normality in (a_0, a_0) if $t(a_0)$ exceeds a critical point which may be obtained in tables in non-parametric texts (e.g. the critical point for Lilliefors test). The procedure could start with the $\widehat{a}_0$ of sec. 2.1. If $t(\widehat{a}_0)$ is significant (-indicating non-normality in $(-\widehat{a}_0, \widehat{a}_0)$), we compute $t(a_0')$ where $a_0' < a_0$, then $t(a_0^2)$ with $a_0^2 < a_0'$ etc., continuing until an a_0^m is obtained for which $t(a_0^m)$ is not significant. On the other hand, if $t(\widehat{a}_0)$ is not significant, we compute $t(a_0')$ for an $a_0 > \widehat{a}_0$, and continue moving to the right of $\widehat{a}_0$ until significance is obtained. Finally we will reach points a_0^* and a_0^{**}, $a_0^* < a_0^{**}$,with $t(a_0^*)$ not significant and $t(a_0^{**})$ significant. We can then repeat the whole procedure for selected points in (a_0^*, a_0^{**}). Ultimately we will obtain an approximation to a reasonable estimate

$$\widehat{a}_2 = \max\{a_0 | t(a_0) \text{ is not significant}\}$$

of the true a_0.

2.4. The procedures of secs. 2.2 and 2.3 are sensitive to the following reality: Their power may be low when the statistics are computed at values of a_0 that exceed but are close to the true a_0. For example, suppose we conduct the χ^2 test to see if there is evidence against normality of the errors in $(-a_0^*, a_0^*)$. Suppose it happens that $a_0^* = a_0 + \delta$ where a_0 is the true parameter value and $\delta > 0$ is small. Only the few residuals that fall in $(-a_0^*, -a_0)$ and (a_0, a_0^*) will aid us in detecting non-normality, and their ability to do this with the χ^2 test, which uses all of the residuals, will be masked by the contributions, towards the value of $\chi^2(a_0^*)$, provided by the vast majority of residuals in $(-a_0, a_0)$, these being values of genuine approximately normal variables.

This observation suggests a possible improvement: For a range of values of a_0, conduct the tests of secs. 2.2 and 2.3 using only residuals in neighbourhoods of $-a_0$ and a_0. Taking Lilliefors test for example, we define

$$S(a_0) = \sup_{u \epsilon A_n} |F_n(u) - \Phi(u)|$$

where $A_n = (-a_0 - \delta_n, -a_0 + \delta_n) \cup (a_0 - \delta_n, a_0 + \delta_n)$, and δ_n depends only on n and satisfies $\delta_n \to 0$ as $n \to \infty$. An estimator of the true a_0 is then

$$\widehat{a}_3 = \max\{a_0 | \text{ the observed value of } S(a_0) \text{ is not significant}\}.$$

2.5 At the cost of quite extensive analytical and computational inconvenience, we can estimate a_0 using a likelihood ratio approach. (Recall that, for simplicity of presentation, we are supposing that $\epsilon = 0$ in (1.5)). The density of U_i in the model (1.1) can, from (1.5), be written as

$$\phi(u)I(|u| \leq a_0) + g(u)I(|u| > a_0)$$

where g is a (assumed to exist) density of the unknown tail portion of F and I(A) denotes the indicator function of a set A. The likelihood function of the sample $x = (x_1, \ldots, x_n)'$ is then

$$L(\theta, a_0, g|x) = \prod_{i=1}^{n} \{\phi(x_i - c_i'\theta)I(|x_i - c_i'\theta| \leq a_0) + g(x_i - c_i'\theta)I(|x_i - c_i'\theta| > a_0)\}.$$

In this, we propose replacing θ by the estimate $\tilde{\theta}$ of sec. 2.1. In order to use L to perform a likelihood ratio test of the alternatives $H_0 : a_0 = a_0^*, H_1 : a_0 > a_0^*$, we have to contend with the complication that g is unknown. One possibility, which we have not explored theoretically, is to replace g by a certain least favourable density g_0 for the likelihood ratio test of H_0 versus H_1. Specifically, if we define

$$\lambda(X; \tilde{\theta}, a_0^*, g) = \sup_{a_0} L(\tilde{\theta}, a_0, g|X)/L(\tilde{\theta}, a_0^*, g|X),$$

we let g_0 minimize the power of the test based on λ, of H_0 versus H_1, at some specified alternative.

A rather complicated estimate of a_0 is then defined as

$$\widehat{a}_4 = \max\{a_0^* | \lambda(x; \tilde{\theta}, a_0^*, g_0) \text{ is not significant}\}.$$

As with the preceding procedures, an initial choice for the null hypothesis value a_0^* of a_0 would be the $\widehat{a}_0$ of sec. 2.1.

2.6. Finally we note that, instead of estimating merely a_0, we could consider estimating the tail portions of F inself. This procedure is adaptive and seems to rely on a very large amount of data. Since it is not in the spirit of robust procedures, which are semi-parametric and designed to work for moderate sample sizes, we will not elaborate on this procedure.

ADDITIONAL REFERENCES

Bassett, G., Jr., and Koenker, R. (1978). Asymptotic Theory of Least Absolute Error Regression. Journal of the American Statistical Association, 73, 618-622.

Hlynka, M., Sheahan, J.N., Wiens, D.P. (1990) On the choice of support of re-descending ψ-function in linear models with asymmetric error distributions. Metrika, to appear

Koenker, R. and Bassett, G., Jr. (1978). Regression Quantiles.Econometrika, 46, 33-50.

Ruppert, D. and Carroll, R.J. (1980). Trimmed Least Squares Estimation in the Linear Model. Journal of the American Statistical Association, 75, 828-138.