ON THE BOUNDED-INFLUENCE REGRESSION ESTIMATOR OF KRASKER AND WELSCH

David Ruppert[1]

[1]David Ruppert is Associate Professor, Department of Statistics, University of North Carolina, Chapel Hill, NC 27514. This research was supported by National Science Foundation Grant MCS-8100748.

ABSTRACT

Recently, Krasker and Welsch (1982) considered a class of bounded-influence regression estimators. They showed that within this class the so-called Krasker-Welsch estimator is the only solution to a first-order necessary condition for strong optimality, i.e., for minimizing, in the sense of positive definiteness, the asymptotic covariance matrix. However, whether any strongly optimal estimator in fact exists remained an open question. In this note, an example is given where no strongly optimal estimator exists.

1. INTRODUCTION

In a recent article, Krasker and Welsch (1982) consider robust estimators for the linear regression model

$$y_i = x_i\beta + \epsilon_i$$

where x_i is a p-dimensional row vector, (y_i, x_i), $i=1,\ldots,n$, are independent and identically distributed, and ϵ_i is distributed $N(0,\sigma^2)$ independently of x_i. Their interest is in bounding the influence of outliers in both x_i and y_i. They consider M-estimators of the form

$$(1.1) \qquad 0 = \sum_{i=1}^{n} \omega(y_i, x_i; \beta_n)(y_i - x_i\beta_n)x_i^t \quad ,$$

where ω is a scalar-valued weighting function. If ω depends on σ, then we substitute an estimator σ_n for σ. The influence function (Hampel, 1974) for this estimator is

$$\Omega(y,x;\omega) = B_\omega^{-1}\,\omega(y,x;\beta)(y-x\beta)x^t$$

where

$$(1.2) \qquad B_\omega = E_{y,x}[\omega(y,x;\beta)(\frac{y-x\beta}{\sigma})^2 x^t x] .$$

The asymptotic covariance matrix is $V_\omega = B_\omega^{-1}A_\omega B_\omega^{-1}$

where

$$(1.3) \qquad A_\omega = E_{y,x}[\omega^2(y,x;\beta)(\frac{y-x\beta}{\sigma})^2 x^t x] .$$

Krasker-Welsch discuss several measures of sensitivity. They choose to use an invariant (and quite reasonable) measure γ_ω defined to be

$$\gamma_\omega = \sup_{y,x} \sup_{\lambda} \frac{|\lambda^t \Omega(y,x;\omega)|}{(\lambda^t V_\omega \lambda)^{\frac{1}{2}}}$$

$$= \sup_{y,x} [\Omega^t(y,x;\omega) V_\omega^{-1} \Omega(y,x;\omega)]^{\frac{1}{2}}$$

$$= \sup_{y,x} [(x\ A_\omega^{-1} x^t)^{\frac{1}{2}} |y-x\beta| \omega(y,x;\beta)]$$

Because an estimator's influence function is normed by the asymptotic covariance matrix of the estimator itself, Stahel (1981) calls γ_ω the self-standardized sensitivity.

Now let a bound $a > 0$ be given, and consider the class of all weighting functions ω such that

$$\gamma_\omega \le a$$

and ω depends on y only through $|y-x\beta|$. We say that w is strongly optimal within this class of $(V_\omega - V_w)$ is positive semidefinite for all ω in the class. Krasker and Welsch show that if a strongly optimal w exists, then (up to a scalar multiple) it must be (letting $\epsilon = y-x\beta$)

$$w(y,x;\beta)$$

$$= \min\{1, a/[|\epsilon|(xA_w x^t)^{\frac{1}{2}}]\} .$$

(This is an implicit definition since A_w appears on the right hand side.) They conjecture that this w is strongly optimal. In this note we show by example that in general no strongly optimal estimators exists.

The Krasker-Welsch estimator is invariant to nonsingular reparametrization. Invariance is not necessarily desirable when the parameters have physical meanings. To show the practical significance of the lack of a strongly optimal estimator, we discuss how one might choose alternatives to the Krasker-Welsch estimator when there are nuisance parameters.

2. AN EXAMPLE

We will take $\sigma = 1$. Suppose $p = 3$ and

$$x_i = (1 \quad (1-U_i)Z_i \quad U_iZ_i)$$

where U_i is distributed Bernoulli($\frac{1}{2}$), Z_i is distributed N(0,1), and ϵ_i, U_i, and Z_i are mutually independent.

For $\Delta > 0$ let

$$\begin{aligned} \omega(y,x;\beta,\Delta) &= \Delta \min\{1, a/(\Delta|\epsilon|(x A_\Delta^{-1}x^t)^{\frac{1}{2}})\} \\ &\qquad \text{if } x^{(2)} \neq 0 \\ &= \min\{1, a/(|\epsilon|(xA_\Delta^{-1}x^t)^{\frac{1}{2}})\} \\ &\qquad \text{if } x^{(2)} = 0 , \end{aligned}$$

where $x^{(2)}$ is the second coordinate of x and $A_\Delta = A_{\omega_\Delta}$ is a solution to equation (1.3). By theorem 1 of Maronna (1976) A_Δ is unique. This also proves that A_Δ is diagonal, since the symmetry of Z implies that the matrix obtained by multiplying the off-diagonal elements of A_Δ by -1 is also a solution to (1.3). Say $A_\Delta = \text{diag}(A_{1,\Delta}, A_{2,\Delta}, A_{3,\Delta})$. Now, equation (1.3) can be rewritten as

$$A_{1,\Delta} = \tfrac{1}{2} E \min\{(\Delta\epsilon)^2, a^2(A_{1,\Delta}^{-1} + A_{2,\Delta}^{-1}Z^2)^{-1}\} + \tfrac{1}{2} E \min\{\epsilon^2, a^2(A_{1,\Delta}^{-1} + A_{3,\Delta}^{-1}Z^2)^{-1}\} \tag{2.1}$$

$$A_{2,\Delta} = \tfrac{1}{2} E \min\{(\Delta\epsilon)^2, a^2(A_{1,\Delta}^{-1} + A_{2,\Delta}^{-1}Z^2)^{-1}\}Z^2 \tag{2.2}$$

$$A_{3,\Delta} = \tfrac{1}{2} E \min\{\epsilon^2, a^2(A_{1,\Delta}^{-1} + A_{3,\Delta}^{-1}Z^2)^{-1}\}Z^2 \tag{2.3}$$

We will take $a > 2$. If we divide (2.1) by $A_{1,\Delta}$, then divide (2.2) by $A_{2,\Delta}$, and add the resulting equations, we obtain

$$(2.4) \qquad 2 = \tfrac{1}{2}\, E \min\{(\Delta\epsilon)^2,\ a^2(A_{1,\Delta}^{-1} + A_{2,\Delta}^{-1}Z^2)^{-1}\}(A_{1,\Delta}^{-1} + A_{2,\Delta}^{-1}Z^2)$$

$$+ \tfrac{1}{2}\, E \min\{\epsilon^2,\ a^2(A_{1,\Delta}^{-1} + A_{3,\Delta}^{-1}Z^2)^{-1}\}A_{1,\Delta}^{-1}$$

Now, choose a sequence $\Delta_m \to \infty$ such that for $i = 1,2,3$ the limits $A_{i,\infty} = \lim_{\Delta_m \to \infty} A_{i,\Delta_m}$ exist (but are possibly $+\infty$). If $A_{1,\infty} < \infty$ and $A_{2,\infty} < \infty$, then by (2.4) we reach a contradiction,

$$2 = \tfrac{1}{2}\, a^2 + \tfrac{1}{2}\, E\{\epsilon^2,\ a^2(A_{1,\infty}^{-1} + A_{3,\infty}^{-1}\, Z^2)^{-1}\}A_{1,\infty}^{-1} > 2$$

Similarly, (2.2) shows that $A_{1,\infty} = \infty$ and $A_{2,\infty} < \infty$ is impossible, and (2.3) shows that $A_{1,\infty} < \infty$ and $A_{2,\infty} = \infty$ is impossible. Thus, $A_{1,\infty} = A_{2,\infty} = \infty$ and $A_{3,\infty}$ solves

$$A_{3,\Delta} = \tfrac{1}{2}\, E \min\{\epsilon^2,\ a^2\, A_{3,\Delta}Z^{-2}\}Z^2\ .$$

Notice that $B_\Delta = B_{\omega_\Delta}$ is diagonal (as well as A_Δ). Thus $\beta_n^{(1)}$, $\beta_n^{(2)}$, and $\beta_n^{(3)}$ are asymptotically uncorrelated, and we can find the asymptotic variance of $\beta_n^{(3)}$ by examining its estimating equation alone. This is

$$(2.5) \qquad 0 = \sum_{i=1}^{n} e_i \min\{1,\ a(A_{1,\Delta}^{-1} + A_{2,\Delta}^{-1}Z_i^2)^{-\frac{1}{2}}|e_i|^{-1}\}\ U_iZ_i$$

where $e_i = y_i - x_i\beta_n$. The asymptotic variance of $\beta_n^{(3)}$ is the same as the asymptotic variance of β_n in the model

$$y_i = U_iZ_i\ \beta + \epsilon_i \qquad\qquad (p = 1)$$

with estimating equation given by (2.5). For this new model, $\gamma_{\omega_\Delta} = a$ for any Δ. Since the new model has a univariate parameter we can use a result of Hampel (1968, lemma 5 or 1974, page 391) to show that

$\Lambda_{1,\Delta}^{-1} = 0$ (i.e., $\Delta = \infty$) is optimal, and the $\Delta = 1$ is strictly suboptimal.

Now $\Delta = \infty$ does not give us an estimator of type (1.1), but it can be approximated arbitrarily closely by taking Δ sufficiently large. Thus, for large Δ, ω_Δ is more efficient for estimating $\beta^{(3)}$ than for the Krasker-Welsch estimator. This fact is a practical importance if $\beta^{(1)}$ and $\beta^{(2)}$ are nuisance parameters, or at least are of only secondary importance compared with $\beta^{(3)}$.

3. DISCUSSION

The "psi-functions" of efficient bounded-influence M-estimators are constructed by downweighting the maximum likelihood estimator's ψ-function wherever the latter is too large. The downweighting can be equal for all coordinates, as when one considers only estimators of form (1.1) with ω scalar. Alternatively, the coordinates can be downweighting differently, e.g., by letting ω in (1.1) be a $p \times p$ diagonal matrix. If some parameters are nuisance parameters, then it might be advantageous to downweight their coordinates severely. This allows one to maintain a given bound on the sensitivity while estimating the non-nuisance parameters more efficiently.

Carroll (1983) mentions the example of two-group analysis of covariance with a balanced covariate and only the treatment difference of interest. The Krasker-Welsch estimator will cause a loss of efficiency for the treatment difference in order to bound the sensitivity to outlying values of the co-variate.

General classes of M-estimators which allow unequal downweighting of the components of the score function have been discussed by Hampel (1978), Krasker (1980), and Stahel (1981). Within such a general class, it is generally impossible to find an estimator which is strongly optimal subject to a bound on some measure of sensitivity.

The conjecture that the Krasker-Welsch estimator was optimal within the class of estimators with ω scalar seemed reasonable to the present author, since having ω scalar greatly restricts flexibility. It did not seem possible, for example, to treat nuisance parameters differently than other parameters. After repeated attempts to prove strong optimality, I stopped work on the problem. Recently, Roy Welsch (oral communication) mentioned that the question of strong optimality was still open, but that unpublished work of Peter Bickel suggested that the Krasker-Welsch estimator might not be strongly optimal.

Shortly after this, I realized that nuisance parameters could be treated differently than the other parameters, if some observations contain information only about the nuisance parameters. If the example, observations with $U_i = 0$ are upweighted, so in the limit as $\Delta \to \infty$ the observations with $U_i = 1$ are used only for estimating $\beta^{(3)}$.

ACKNOWLEDGMENT

I would like to thank R.J. Carroll for encouragement and useful discussions about this research.

REFERENCES

CARROLL, R.J. (1983). Comment on "Minimax Aspects of Bounded-Influence Regression" by Peter J. Huber. Journal of the American Statistical Association, 78. 78-79.

HAMPEL, F.R. (1968). Contributions to the Theory of Robust Estimation. Ph.D. Thesis. University of California, Berkeley.

HAMPEL, F.R. (1974). The Influence Curve and Its Role in Robust Estimation. Journal of the American Statistical Association, 62. 1179-1186.

HAMPEL, F.R. (1978). Optimally Bounding the Gross-error-sensitivity and the Influence of Position in Factor Space. 1978 Proceedings of the ASA Statistical Computing Section. ASA, Washington, DC 59-64.

KRASKER, W.S. (1980). Estimation in Linear Regression Models with Disparate Data Points. Econometrica, 48. 1333-1346.

KRASKER, W.S. and WELSCH, R.E. (1982). Efficient Bounded-Influence Regression Estimation. Journal of the American Statistical Association, 77 . 595-604.

MARONNA, R.A. (1976). Robust M-estimators of Multivariate Location and Scatter. Annals of Statistics, 4. 51-67.

STAHEL, W.A. (1981). Robuste Schaetzungen: Infinitesimale Optimalitaet und Schaetzungen von Kovarianzmatrizen. Ph.D. Dissertation. Swiss Federal Institute of Technology, Zurich.