한국통계학회 논문집 제1권1호
1994년 12월 pp. 27-32

A Generalized M-Estimator in Linear Regression[1]

Moon Sup Song[2], Changsoon Park[3], Ho Soo Nam[4]

Abstract

We propose a robust regression estimator which has both a high breakdown point and a bounded influence function. The main contribution of this article is to present a weight function in the generalized M (GM)-estimator. The weighting schemes which control leverage points only without considering residuals cannot be efficient, since these schemes inevitably downweight some good leverage points. In this paper we propose a weight function which depends both on design points and residuals, so as not to downweight good leverage points. Some motivating illustrations are also given.

1. Introduction

Consider the linear regression model

$$y_i = \boldsymbol{x}_i' \beta + \varepsilon_i , \quad i = 1, \cdots, n , \tag{1}$$

where $\{(\boldsymbol{x}_i, y_i) : i = 1, \cdots, n\}$ is a sequence of iid random variables with distribution function $F(\boldsymbol{x}, y)$, $\boldsymbol{x}_i$ is $p \times 1$ vector, $\boldsymbol{x}_i$ and ε_i are independent random variables with $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$, and β is a $p \times 1$ vector of unknown parameters.

The classical least squares (LS) estimator is very sensitive to influential observations. On the other hand the M-estimator is robust to outliers in y-direction, but susceptible to leverage points, especially to bad leverage points. Recently, two classes of robust procedures have emerged : the bounded influence procedures discussed by Krasker and Welsch (1982) and Simpson et al.(1992), and the high breakdown methods such as the Least Median of Squares (LMS) estimator and the Least Trimmed Squares (LTS) estimator. The former pertains to the local behavior of the estimators, whereas the latter to the global behavior.

Coakley and Hettmansperger (1993) proposed a bounded influence, high breakdown and efficient regression estimator. Their weights depend only on X, but they applied the weighting scheme to the Schweppe-type GM-estimator to avoid downweighting of good leverage points. Nevertheless the leverage points which have moderately small residuals are downweighted. To overcome this drawback, we propose a weight function which controls both residuals and leverage points simultaneously. This weighting structure downweights severely

1) The present studies were supported in part by the Basic Science Research Institute Program, Ministry of Education, 1994, Project No. BSRI-94-1415.
2) Department of Statistics, Seoul National University, Seoul, 151-742, KOREA
3) Department of Applied Statistics, Chung-Ang University, Seoul, 156-756, KOREA
4) Graduate School of Seoul National University, Seoul, 151-742, KOREA

the bad leverage points, but when the leverage points have very small residuals these are seldom downweighted. That is, we do not assent to suppress leverage points blindly.

Under some proper conditions, the proposed GM-estimator has a bounded influence, which is shown in Section 2. In Section 3 we present some illustrations which give a motivation to our proposal.

2. The Proposed GM-Estimator and It's Properties

The GM-estimator of β in model (1) is defined implicitly by the solution of the equation

$$\sum_{i=1}^{n} \eta(\boldsymbol{x}_i,\ r_i(\beta))\boldsymbol{x}_i = \mathbf{0}\ , \tag{2}$$

where $r_i(\beta)=y_i-\boldsymbol{x}_i'\beta$. Simpson et al. (1992) and Coakley and Hettmansperger (1993) proposed the Mallows-type and the Schweppe-type GM-estimators, respectively. The η -function of the Mallows-type is of the form $\eta(\boldsymbol{x},r)=w(\boldsymbol{x})\psi(r/\sigma)$, and that of the Schweppe-type $\eta(\boldsymbol{x},r)=w(\boldsymbol{x})\psi(r/\sigma w(\boldsymbol{x}))$. In this article we consider the η-function of the form

$$\eta(\boldsymbol{x},r) = w(\boldsymbol{x},r)\psi\left(\frac{r}{\sigma w(\boldsymbol{x},r)}\right).$$

The Schweppe-type estimator proposed by Coakley and Hettmansperger (1993) is a one-step estimator given by

$$\hat{\beta} = \widehat{\beta}_0 + \widehat{\sigma}_0\left[\sum_{i=1}^{n}\psi'\left(\frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0 w_i}\right)\boldsymbol{x}_i\boldsymbol{x}_i'\right]^{-1}\left[\sum_{i=1}^{n} w_i\psi\left(\frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0 w_i}\right)\boldsymbol{x}_i\right],$$

where w_i is the same weight as $v(\boldsymbol{x}_i)$ in (5). They proved that the one-step GM-estimator inherits the high breakdown property when the initial estimate has the property. The LMS, LTS and S estimators are candidates of initial estimates having high breakdown point.

Many researchers proposed weights which depend only on the design points. In this paper we propose a weight function which depends on residuals as well as the design points. This weighting scheme controls both leverage points and outliers. Therefore we consider the solution of the estimating equation

$$\sum_{i=1}^{n} w_i(\beta)\psi\left(\frac{r_i(\beta)}{\sigma w_i(\beta)}\right)\boldsymbol{x}_i = \mathbf{0}\ . \tag{3}$$

If $\widehat{\beta}_0$ is an initial estimate of β, an one-step estimator based on $\widehat{\beta}_0$ can be obtained by taking a first order Taylor series expansion of the left side of (3) about $\widehat{\beta}_0$. The resulting estimator is

$$\begin{aligned}\hat{\beta} &= \widehat{\beta}_0 - \left(\sum_{i=1}^{n}\boldsymbol{x}_i g_i'(\widehat{\beta}_0)\right)^{-1}\left(\sum_{i=1}^{n} g_i(\widehat{\beta}_0)\boldsymbol{x}_i\right)\\ &= \widehat{\beta}_0 - (X'\hat{A}X)^{-1}X'\widehat{W}\,\psi\end{aligned}$$

where X is the $n \times p$ matrix having rows $\boldsymbol{x}_i'$, $\widehat{\sigma}_0 = 1.4826\, med\{|r_i(\widehat{\beta}_0)|\}$,

$$g_i(\beta) = w_i(\beta)\psi\left(\frac{r_i(\beta)}{\sigma w_i(\beta)}\right), \qquad \widehat{A} = diag(m(\boldsymbol{x}_i,\ r_i(\widehat{\beta}_0))), \qquad \widehat{W} = diag(\widehat{w_i}),$$

Ψ denotes the $n \times 1$ vector whose ith component is $\psi(\frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0 \widehat{w_i}})$, and the $m(\cdot,\cdot)$ is a function satisfying

$$g_i'(\beta) = \frac{\partial g_i(\beta)}{\partial \beta} = m(\boldsymbol{x}_i, r_i(\beta))\boldsymbol{x}_i' . \tag{4}$$

The estimated weight $\widehat{w_i}$ is determined by the following weight function :

$$\begin{aligned} w_i &= w_i(\beta) \\ &= w(\boldsymbol{x}_i,\ r_i(\beta)) \\ &= \frac{v(\boldsymbol{x}_i)}{h(r_i(\beta)/\sigma)} \phi_c\left(\frac{h(r_i(\beta)/\sigma)}{v(\boldsymbol{x}_i)}\right) \\ &= \min\left(1,\ \frac{c \cdot v(\boldsymbol{x}_i)}{h(r_i(\beta)/\sigma)} \operatorname{sign}(r_i(\beta))\right), \end{aligned}$$

where $h(x)$ is an increasing function of x such as x or $\sqrt{|x|} \cdot \operatorname{sign}(x)$, and

$$\phi_c(x) = \begin{cases} x, & |x| < c \\ c \cdot \operatorname{sign}(x), & |x| \geq c. \end{cases}$$

According to Simpson et al.(1992),

$$v(\boldsymbol{x}_i) = \min\left[1,\ \left(\frac{b}{(\boldsymbol{z}_i - \boldsymbol{m}_z)'\, \boldsymbol{C}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{m}_z)}\right)^{\alpha/2}\right], \qquad (\alpha \geq 1), \tag{5}$$

is a possible choice of $v(\boldsymbol{x}_i)$, where $\boldsymbol{z}_i$ is $(p-1) \times 1$ vector such that $\boldsymbol{x}_i' = (1,\ \boldsymbol{z}_i')$, $\boldsymbol{m}_z$ and $\boldsymbol{C}_z$ are robust estimates of location and covariance, respectively, and b is the $(1-\gamma)$ quantile of the chi-square distribution on $p-1$ degrees of freedom ($\gamma = 0.1$ or 0.05). The proposed weights are inversely propotional to the absolute values of residuals and to the distance between the design point and the center of the design points. Thus the $\widehat{w_i}$ is given by $\widehat{w_i} = w(\boldsymbol{x}_i, r_i(\widehat{\beta}_0))$, $i = 1, \cdots, n$. In the computation of $g_i(\widehat{\beta}_0)$, $g_i'(\widehat{\beta}_0)$ and $w_i(\widehat{\beta}_0)$ we assume that σ is replaced by $\widehat{\sigma}_0$.

Now to show that the proposed estimator has a bounded influence function, consider a functional $T(F)$ satisfying

$$\int \eta(\boldsymbol{x},\ y - \boldsymbol{x}'T(F))\boldsymbol{x}\, dF(\boldsymbol{x}, y) = \boldsymbol{0}. \tag{6}$$

Let $F_{t,(u,v)} = (1-t)F + t\Delta_{(u,v)}$, and $\sigma = 1$ for simplicity. Replace F by $F_{t,(u,v)}$ in (6), and differentiate with respect to t to obtain

$$0 = \frac{\partial}{\partial t}\int \eta(\boldsymbol{x},\ y-\boldsymbol{x}'T((1-t)F+t\Delta_{(u,v)}))\boldsymbol{x}d((1-t)F+t\Delta_{(u,v)})\Big|_{t=0}$$

$$= \int \eta(\boldsymbol{x},\ y-\boldsymbol{x}'T(F))\boldsymbol{x}d(\Delta_{(u,v)}-F)$$

$$+\int \frac{\partial}{\partial \beta}[\eta(\boldsymbol{x},\ y-\boldsymbol{x}'\beta)]_{T(F)}\boldsymbol{x}dF \cdot \frac{\partial}{\partial t}T[F_{t,(u,v)}]\ |_{t=0}.$$

Thus,

$$IF(\boldsymbol{x},y;\ T(F)) = \left[-\int m(\boldsymbol{x},\ y-\boldsymbol{x}'T(F))\boldsymbol{x}\boldsymbol{x}'dF(\boldsymbol{x},y)\right]^{-1}\cdot$$

$$w(\boldsymbol{x},y-\boldsymbol{x}'T(F))\psi\left(\frac{y-\boldsymbol{x}'T(F)}{w(\boldsymbol{x},y-\boldsymbol{x}'T(F))}\right)\boldsymbol{x},$$

where $m(\cdot,\cdot)$ is implicitly defined in (4).

Suppose that $(1/n)(X'\hat{A}X) \xrightarrow{p} D$, D is positive definite $p\times p$ matrix, and only when $\|\boldsymbol{x}\|$ goes to infinity, $|y-\boldsymbol{x}'\beta| > \|\boldsymbol{x}\|^{-\alpha}, (\alpha \ge 1)$. (The latter condition does not cause any trouble in practical situation.) Then we can show that $IF(\boldsymbol{x},y;\ T(F))$ is bounded in Euclidean norm for appropriate bounded ψ-function and w.

Now, consider the breakdown point of the proposed estimator. We can prove that the one-step estimator inherits the breakdown point of the initial estimate according to Simpson et al. (1992) under the assumptions stated above.

Not knowing the distribution of the proposed one-step GM-estimator, we drive an asymptotic distribution of the estimator. From Theorem 4 of Coakley and Hettmansperger (1993), we can prove the following theorem with $m(\boldsymbol{x}_i, r_i(\hat{\beta}_0))$ in place of b_i.

Theorem Under the same conditions in Theorem 4 of Coakley and Hettmansperger (1993),

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(\mathbf{0},\ \Sigma),$$

where

$$\Sigma = D^{-1}ED^{-1},$$

$$D = \int m(\boldsymbol{x},\ r(\hat{\beta}_0))\boldsymbol{x}\boldsymbol{x}'dF(\boldsymbol{x},y),$$

$$E = \int \psi^2\left(\frac{r(\hat{\beta}_0)}{\hat{\sigma}_0 w(\boldsymbol{x},r(\hat{\beta}_0))}\right) w^2(\boldsymbol{x},r(\hat{\beta}_0))\boldsymbol{x}\boldsymbol{x}'dF(\boldsymbol{x},y).$$

Note that the covariance matrix Σ can be estimated as follows.

$$\hat{\Sigma} = S_1^{-1}\ S_2\ S_1^{-1},$$

where

$$S_1 = \frac{1}{n}X'\hat{D}_sX, \qquad S_2 = \frac{1}{n}X'\hat{E}_sX,$$

$$\hat{D}_s = diag\left(\frac{1}{n}\sum_{j=1}^{n} m(\boldsymbol{x}_i,\ r_j(\hat{\beta}_0))\right),$$

$$\widehat{E}_s = diag\left(\frac{1}{n} \sum_{j=1}^{n} \psi^2 \left(\frac{r_j(\widehat{\beta}_0)}{\widehat{\sigma}_0 \widehat{w}(\boldsymbol{x}_i, r_j(\widehat{\beta}_0))} \right) \widehat{w}^2 (\boldsymbol{x}_i, r_j(\widehat{\beta}_0)) \right).$$

The idea of this estimator can be found in Marazzi (1993, p145).

3. Numerical Examples

To compare the responses of various estimators to some extreme outliers, we consider the water flow data in Hampel et al. (1986, p310). The data indicate the water flow at two points (Libby= x, Newgate= y) on Kootenay river in January, which are represented in Table 1. We want to fit the data with a simple regression line :

$$y_i = \alpha + \beta x_i, \quad i = 1, \cdots, 13.$$

We include the LS, Huber-M, Mallows(Schweppe), LTS, and the proposed estimators in our comparative study. In every M-type estimators, the Huber's ψ-function with tuning constant 1.5, and the LTS estimator as an initial value are used. For the proposed estimator, $h(x) = x$ and $c = 3.0$ in the weight function are used. In the definition of $v(\cdot)$, we use $b = \chi^2_{1,0.9}$ and $(m_x, c_x) = (med\{x_i\}, (1.4826MAD\{x_i\})^2)$. We applied three-step methods in the computation of M- and GM-estimators. The numerical computations in this section are performed by S-PLUS (version 3.2).

Table 1. Water Flow at Two points (units of cfs)

x	27.1	20.9	33.4	77.6	37.0	21.6	17.6	35.1	32.6	26.0	27.6	38.7	27.8
y	19.7	18.0	26.1	44.9	26.1	19.9	15.7	27.6	24.9	23.4	23.1	31.3	23.8

For illustration we now move the point (77.6, 44.9) to $P_1 = (15.7, 44.9)$ in the data in Table 1. The fitted lines are displayed in Figure 1. In this case the estimators based on Schweppe-type and Mallows-type are equal.

The changed point P_1 is a bad leverage point, so it is reasonable to downweight this point in robust estimation. The weight corresponding to the leverage point is $0.3036 (= v(\boldsymbol{x}))$ in the Mallows(Schweppe)-type estimator, but 0.0293 in our estimator.

This implies that the point P_1 is severely suppressed in the proposed estimator more than in the other estimators. Therefore the proposed estimator is seldom influenced by the point P_1. On the other hand, the Mallows(or Schweppe)-type estimator is slightly influenced by the point.

Next, move the point (32.6, 24.9) to $P_2 = (76.0, 56.7)$ in addition to the first case. The fitted lines are displayed in Figure 2. Note that the new point P_2 is a good leverage point, and no downweighting is needed. The weights for P_1 and P_2 in the proposed estimator are

0.0312 and 1.0000 respectively, that is, the point P_2 is not downweighted at all, but in the Mallows-type 0.3379 and 0.3491 respectively.

According to Figure 1 and Figure 2 we can see that the proposed estimator is more robust to bad leverage point by downweighting the effect of the point, and fully uses the effect of good leverage point without downweighting.

4. Summary and Conclusions

In this paper we have proposed a Schweppe-type GM-estimator which has a high breakdown point and a bounded influence function. The proposed estimator is very resistant to bad leverage points, and it uses good leverage points without downweighting.

It is very important to choose the tuning constant in the proposed weight function. This constant determines whether a point is downweighted or not by considering the degree of leverage and residuals simultaneously. When the absolute value of studentized residual exceeds 1.5, and the measure of leverage $v(\boldsymbol{x}_i)$ is less than 0.5 some restrictions to the corresponding point will be needed. Therefore we use 3.0 as a tuning constant c in the weight function.

For the high breakdown initial estimators such as LTS estimator, one can use S-PLUS (version 3.2, 1994) or ROBETH (FORTRAN subroutine library developed by Marazzi (1993)). The S-PLUS contains the LTS, LMS, and Huber-M estimators in linear regression, and the high breakdown point regression or the bounded influence regression can be carried by ROBETH.

References

[1] Coakley, C.W., and Hettmansperger, T.P. (1993). A Bounded Influence, High Breakdown, Efficient Regression Estimator, *Journal of the American Statistical Association*, 88, 872-880.

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

[3] Krasker, W.S., and Welsch, R.E. (1982). Efficient Bounded-Influence Regression Estimation, *Journal of the American Statistical Association*, 77, 595-604.

[4] Marazzi, A. (1993). *Algorithms, Routines, and S Functions for Robust Statistics*, Wadsworth & Brooks, California.

[5] Simpson, D.G., Ruppert, D., and Carroll, R.J. (1992). On One-Step GM Estimates and Stability of Inferences in Linear Regression, *Journal of the American Statistical Association*, 87, 439-450.

한국통계학회논문집 제2권 1호
1995년 4월 pp. 266

A Generalized M-Estimator in Linear Regression

Moon Sup Song , Changsoon Park , Ho Soo Nam

아래 그림은 한국통계학회 논문집 제1권 1호에 개제된 논문 " A Generalized M-Estimator in Linear Regression"에 추가로 삽입될 그림이다.

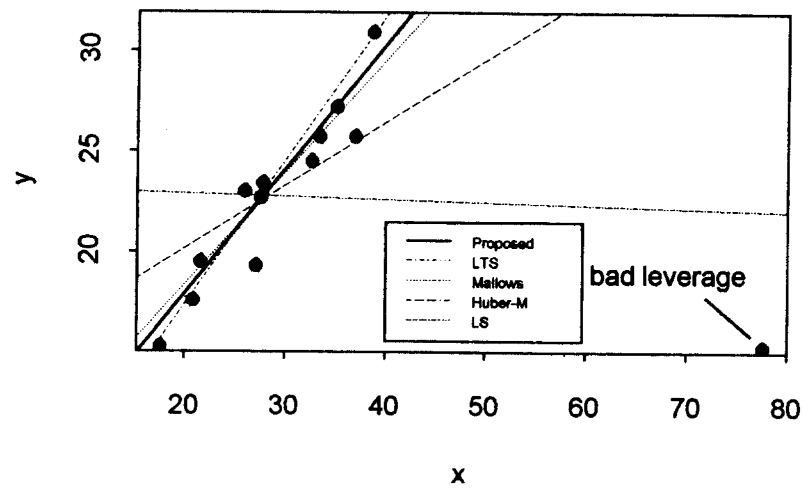

Figure 1. Fitted Lines with One Leverage Point

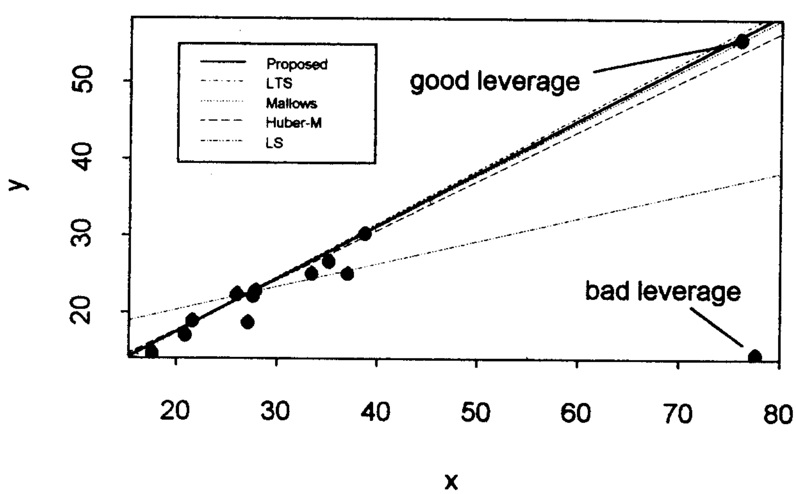

Figure 2. Fitted Lines with Two Leverage Points