Journal of the Korean
Statistical Society
Vol. 25, No. 4, 1996

A High Breakdown and Efficient GM-Estimator in Linear Models†

Moon Sup Song [1], Changsoon Park[2] and Ho Soo Nam[3]

Abstract

In this paper we propose an efficient scoring type one-step GM-estimator, which has a bounded influence function and a high breakdown point. The main point of the estimator is in the weighting scheme of the GM-estimator. The weight function we used depends on both leverage points and residuals. So we construct an estimator which does not downweight good leverage points. Under some regularity conditions, we compute the finite-sample breakdown point and prove asymptotic normality. Some simulation results are also presented.

Key Words : Influence function; One-step GM-estimator; Breakdown point; LMS and LTS estimators; MVE estimator.

†The present studies were supported in part by the Basic Science Research Institute Program, Ministry of Education, Korea, 1995, Project BSRI-95-1415.

[1]Department of Statistics, Seoul National University, Seoul, 151-742, Korea.
[2]Department of Applied Statistics, Chung-Ang University, Seoul, 156-756, Korea.
[3]Department of Industrial Engineering, Dongseo University, Pusan, 617-716, Korea.

1. MODEL AND MOTIVATION

We consider the linear regression model

$$y_i = \mathbf{x}_i^T \beta + \epsilon_i \ , \quad i = 1, 2, \cdots, n \ , \tag{1.1}$$

where $\{(\mathbf{x}_i, y_i) : i = 1, 2, \cdots, n\}$ is a sequence of independent and identically distributed (iid) random variables with distribution function $F(\mathbf{x}, y)$, $\mathbf{x}_i$ is a $p \times 1$ random vector, and β is a $p \times 1$ vector of unknown parameters. While, ϵ_i's are iid, independent of $\mathbf{x}_i$ and symmetric about 0 with scale parameter σ.

It is well known that the classical LS estimator is very sensitive to influential observations. On the other hand, the Huber's M-estimator is robust to outliers in y-direction, but still susceptible to leverage points, especially to bad leverage points.

In general, the GM-estimator of β for the regression model (1.1) is defined implicitly by the solution of the equation

$$\sum_{i=1}^{n} \eta \left(\mathbf{x}_i, \ \frac{y_i - \mathbf{x}_i^T \beta}{\sigma} \right) \mathbf{x}_i = \mathbf{0}, \tag{1.2}$$

where $\eta : \Re^p \times \Re \to \Re$.

It is well known that the GM-estimators have bounded influence functions. However, the breakdown points of GM-estimators are at most $1/p$ (see Maronna, Bustos and Yohai, 1979, or Simpson, Ruppert and Carroll, 1992). Thus, in a multiple regression model with several explanatory variables, the breakdown points of the GM-estimators are significantly low, which are viewed as a serious deficiency of the GM-estimators.

To overcome this deficiency of GM-estimators, Simpson *et al.* (1992) and Coakley and Hettmansperger (1993) suggested to use one-step GM-estimators with high breakdown initial estimates. The main point is that the one-step GM-estimators inherit the breakdown points of the initial estimates.

Simpson *et al.* (1992) proposed the one-step Mallows-type GM-estimators with η-function of the form

$$\eta(\mathbf{x}, r/\sigma) = w(\mathbf{x}) \psi(r/\sigma),$$

where $r = y - \mathbf{x}^T \beta$. Note that this Mallows type downweights the leverage points regardless of the contribution to the model of these points. Thus, some good leverage points are inevitably downweighted.

The one-step GM-estimators proposed by Coakley and Hettmansperger (1993) are based on the Schweppe-type η-function, which is of the form

$$\eta(\mathbf{x}, r/\sigma) = w(\mathbf{x}) \psi(r/\sigma w(\mathbf{x})).$$

This Schweppe type is designed to downweight the leverage point only if the corresponding residual is large.

We propose a weight function which depends on both residuals and leverage points simultaneously. This weighting scheme severely downweights the bad leverage points, but seldom downweights the good leverage points. That is, we do not assent to suppress leverage points blindly.

Under some regularity conditions, the proposed GM-estimator has a bounded influence function and a high breakdown point.

2. THE PROPOSED ONE-STEP GM-ESTIMATOR

Note that both the Mallows-type and the Schweppe-type GM-estimators adopt the weights which depend only on the design points $\mathbf{x}$'s. There is no device to distinguish "good" leverage points from "bad" leverage points in the weights. In the Schweppe-type η-function, the residuals are adjusted by weights to avoid downweighting "good" leverage points. But according to our simulation study, some leverage points with moderately small residuals are also significantly downweighted.

To remedy this defect we suggest to use the Schweppe-type η-function with weights which depend on both the residuals and the design points. Thus, the η-function is of the form

$$\eta(\mathbf{x}, r) = w(\mathbf{x}, r)\psi\left(\frac{r}{w(\mathbf{x}, r)}\right) \tag{2.1}$$

or, by using scaled residuals,

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

where

$$w\left(\mathbf{x}, \frac{r}{\sigma}\right) = \min\left(1, \frac{a \cdot \sigma v(\mathbf{x})}{r} sign(r)\right). \tag{2.2}$$

Here, a is a tuning constant and $v(\mathbf{x})$ is a measure of leverageness defined as follows (see Simpson *et al.* 1992):

$$v(\mathbf{x}) = \min\left[1, \left\{\frac{b}{(\mathbf{z} - \mathbf{m_z})^T \mathbf{C_z}^{-1}(\mathbf{z} - \mathbf{m_z})}\right\}^{1/2}\right],$$

where $\mathbf{z}$ is a $(p-1)\times 1$ vector of predictor variables such that $\mathbf{x}^T = (1, \mathbf{z}^T)$, and $\mathbf{m_z}$ and $\mathbf{C_z}$ are the minimum volume ellipsoid(MVE) estimates of location and covariance of $\{\mathbf{z}\}$, respectively, and b is the $(1-\gamma)$ quantile of the chi-squared distribution with $p-1$ degrees of freedom ($\gamma = 0.05$ or 0.025). Note that $||v(\mathbf{x})\mathbf{x}||$ is bounded as a function of $\mathbf{x}$.

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. However, good leverage points are not supposed to be downweighted, if the corresponding residuals obtained from some initial fitting are small enough. This fact also implies that we have to use some robust procedures to obtain initial fitting.

Using the proposed η-function and weights, the corresponding GM-estimator is a solution of the simultaneous equations

$$\sum_{i=1}^{n} w(\mathbf{x}_i, r_i(\beta)/\sigma)\psi\left(\frac{r_i(\beta)}{\sigma w(\mathbf{x}_i, r_i(\beta)/\sigma)}\right)\mathbf{x}_i = \mathbf{0}\ . \tag{2.3}$$

Let $\hat{\beta}_0$ be an initial estimate of β such as the least trimmed squares(LTS) estimate, which was proposed by Rousseeuw (1984). Then the proposed scoring type one-step estimator based on $\hat{\beta}_0$ is given by

$$\hat{\beta} = \hat{\beta}_0 + \hat{\sigma}_0 H_0^{-1} g_0, \tag{2.4}$$

where

$$g_0 = \sum_{i=1}^{n} \eta(\mathbf{x}_i, r_i(\hat{\beta}_0)/\hat{\sigma}_0)\mathbf{x}_i$$

and

$$H_0 = X^T \hat{A} X\ , \quad \hat{A} = diag\left(\frac{1}{n}\sum_{j=1}^{n} \eta'(\mathbf{x}_i, \frac{r_j(\hat{\beta}_0)}{\hat{\sigma}_0})\right). \tag{2.5}$$

Here, X is the $n \times p$ matrix having rows $\mathbf{x}_i^T$, $\eta'(\mathbf{x}, r) = \partial\eta(\mathbf{x}, r)/\partial r$, $\hat{\sigma}_0 = 1.4826 \times MAD\{r_i(\hat{\beta}_0)\}$, the abbreviation MAD stands for *median absolute deviation*, and

$$\eta\left(\mathbf{x}_i, \frac{r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\right) = w(\mathbf{x}_i, r_i(\hat{\beta}_0)/\hat{\sigma}_0)\psi\left(\frac{r_i(\hat{\beta}_0)}{\hat{\sigma}_0 w(\mathbf{x}_i, r_i(\hat{\beta}_0)/\hat{\sigma}_0)}\right).$$

We consider only the scoring type one-step estimator because, according to our simulation study, the Newton-Raphson version of the estimator has larger fluctuation in the correction amounts to the initial value, and consequently becomes unstable and less efficient than the scoring type.

3. PROPERTIES OF THE PROPOSED ESTIMATOR

In this section we investigate some properties of the proposed one-step GM-estimator given by (2.4). Our concerns are focused on: (1) boundedness of the influence function in both x and y directions, (2) high breakdown point, and (3) asymptotic normality of the proposed estimator. We use the following assumptions in Theorem 1-2.

(C1) The initial estimator $\widehat{\beta}_0$ has an influence function $IF(\mathbf{x}, y; \widehat{\beta}_0)$.

(C2) $\psi(\cdot)$ is bounded, $\eta(\mathbf{x}, r)$ is an odd function of r, and $\eta(\mathbf{x}, r) \geq 0$ for all $\mathbf{x} \in \Re^p$ and $r \in \Re^1$.

(C3) Without loss of generality, the first p observations are uncontaminated and $\mathbf{x}_1, \cdots, \mathbf{x}_p$ are linearly independent.

(C4) Assume that, for all $i = 1, \cdots, n$,

$$\sum_{j=1}^{n} \eta'(\mathbf{x}_i, r_j(\widehat{\beta}_0))$$

are positive at least for the first p observations, and nonnegative for the other observations.

(C5) $(1/n)H_0 \xrightarrow{P} D$, where D is a positive definite $p \times p$ matrix.

(C6) $(1/n)(X^T \widehat{V} X) \xrightarrow{P} E$, where $\widehat{V}$ is the diagonal matrix with diagonal elements $\eta^2(\mathbf{x}_i, r_i(\widehat{\beta}_0)/\widehat{\sigma}_0)$, $i = 1, \cdots, n$ and E is a positive definite $p \times p$ matrix.

(C7) $\max_{i,j} |x_{ij}| = o_p(n^{1/2})$, for all $i = 1, \cdots, n$; $j = 1, \cdots, p$.

(C8) $\widehat{\beta}_0 = \beta + O_p(n^{-1/2})$.

(C9) $\widehat{\sigma}_0 = \sigma + O_p(n^{-1/2})$.

First, we derive the influence function of the proposed estimator. For simplicity, let $\sigma = 1$, and for any distribution function F of $(\mathbf{x}, y)$ let $\beta_1(F)$ and $\beta_0(F)$ denote the functionals corresponding to the proposed one-step estimator in (2.4) and the initial estimator, respectively. Then $\beta_1(F)$ can be written as follows:

$$\beta_1(F) = \beta_0(F) + \{H(F)\}^{-1} g(F),$$

where

$$g(F) = E_F\left[\eta(\mathbf{x}, y - \mathbf{x}^T\beta_0(F))\mathbf{x}\right]$$

and

$$H(F) = E_{F^x}\left[E_{F^{y|x}}\left[\eta'(\mathbf{x}, y - \mathbf{x}^T\beta_0(F))\right]\mathbf{x}\mathbf{x}^T\right].$$

Here $F^{\mathbf{x}}$ and $F^{y|\mathbf{x}}$ mean the marginal distribution of $\mathbf{x}$ and the conditional distribution of y given $\mathbf{x}$, respectively. Note that $\widehat{\beta} = \beta_1(F_n)$, $\widehat{\beta}_0 = \beta_0(F_n)$, $g_0 = ng(F_n)$ and $H_0 = nH(F_n)$, where F_n is the empirical distribution function of $\{(\mathbf{x}_i, y_i)\}$.

To obtain the influence function of the proposed estimator $\widehat{\beta}(F)$, let F_0 be the distribution function of the target model satisfying the assumptions in the model (1.1) and F_t be the t-contamination of F_0, i.e., $F_t = (1-t)F_0 + t\delta_{\mathbf{x},y}$, where $\delta_{\mathbf{x},y}$ denotes the pointmass 1 at $(\mathbf{x}, y)$. Note that, under the target model F_0, $\beta_1(F_0) = \beta_0(F_0) = \beta$, i.e., $g(F_0) = \mathbf{0}$.

Now, replace F by F_t in $\beta_1(F)$, then we have

$$\beta_1(F_t) = \beta_0(F_t) + \{H(F_t)\}^{-1}g(F_t). \tag{3.1}$$

And differentiate both sides of (3.1) with respect to t and replace t by 0. Then, under the assumption (C1), we have

$$IF(\mathbf{x}, y; \widehat{\beta}) = IF(\mathbf{x}, y; \widehat{\beta}_0) + [H(F_0)]^{-1}\, IF(\mathbf{x}, y; g) \tag{3.2}$$

Here, the influence function of $g(F)$ can be obtained as follows.

$$\begin{aligned} IF(\mathbf{x}, y; g) &= \lim_{t\downarrow 0}\frac{g(F_t) - g(F_0)}{t} = \lim_{t\downarrow 0}\frac{g(F_t)}{t} \\ &= E_{\delta_{x,y}-F_0}\left[\eta(\mathbf{u}, v - \mathbf{u}^T\beta_0(F_0))\mathbf{u}\right] \\ &\qquad -E_{F_0}\left[\eta'(\mathbf{u}, v - \mathbf{u}^T\beta_0(F_0))\mathbf{u}\mathbf{u}^T\right] IF(\mathbf{x}, y; \widehat{\beta}_0) \\ &= \eta(\mathbf{x}, y - \mathbf{x}^T\beta)\mathbf{x} - [H(F_0)]IF(\mathbf{x}, y; \widehat{\beta}_0). \end{aligned} \tag{3.3}$$

By inserting (3.3) into (3.2), we have

$$IF(\mathbf{x}, y; \widehat{\beta}) = [H(F_0)]^{-1}\eta(\mathbf{x}, y - \mathbf{x}^T\beta)\mathbf{x}. \tag{3.4}$$

Thus, if we choose a ψ-function so that

$$||\eta(\mathbf{x}, r)\mathbf{x}|| = \left\|w(\mathbf{x}, r)\psi\left(\frac{r}{w(\mathbf{x}, r)}\right)\mathbf{x}\right\| < \infty$$

as $||\mathbf{x}||$ or $|r|$ goes to infinity, then the proposed estimator has a bounded influence function.

Next, we consider the finite-sample breakdown point of the proposed estimator. Let $Z_n = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$ be any sample of n data points, and T be an estimator of the regression parameter such that $T(Z_n) = \hat{\beta}$. Denote by $B(m; T, Z_n)$ the maximum bias that can be caused by the contaminated data points Z_n^*, where Z_n^* is obtained by replacing any m of the original data points Z_n by arbitrary values. That is,

$$B(m; T, Z_n) = \sup_{Z_n^*} \; || \; T(Z_n) - T(Z_n^*) \; ||,$$

where the supremum is taken over all possible Z_n^*. Then, the finite-sample breakdown point of the estimator $T(Z_n)$ is defined as

$$\varepsilon_n^*(T, \; Z_n) = \max \left\{ \frac{m}{n} \; ; \; B(m; T, Z_n) < \infty \right\}.$$

That is, ε_n^* is the maximum fraction of outliers that can be allowed in a given sample without spoiling the estimator completely.

We now present a theorem which computes the breakdown point of the proposed estimator. The proof of the theorem is given in Appendix.

Theorem 1. Under the assumptions (C2) through (C4), the proposed one-step estimator has a breakdown point of $([n/2] - p + 1)/n$.

Finally we derive the asymptotic distribution of the proposed estimator under some regularity conditions. Bickel (1975) derived the asymptotic distribution of the one-step M-estimator in the linear model. Maronna and Yohai (1981) discussed the asymptotic behavior of GM-estimators with random carriers assuming general regularity conditions. For the one-step Schweppe-type GM-estimator, the asymptotic normality is described in Coakley and Hettmansperger (1993). Our proof in Appendix is based on Bickel (1975) and Coakley and Hettmansperger (1993).

Theorem 2. Under the assumptions (C2) and (C5) through (C9),

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \Sigma),$$

where $\Sigma = D^{-1} E D^{-1}$ with

$$D = \int \eta' \left(\mathbf{x}, \frac{y - \mathbf{x}^T \beta}{\sigma} \right) \mathbf{x}\mathbf{x}^T dF(\mathbf{x}, y)$$

and

$$E = \int \eta^2 \left(\mathbf{x}, \frac{y - \mathbf{x}^T \beta}{\sigma} \right) \mathbf{x}\mathbf{x}^T dF(\mathbf{x}, y).$$

Proof. The proof is given in Appendix.

For inferences about β, the asymptotic variance of $\sqrt{n}\widehat{\beta}$, $nVar(\widehat{\beta}) = \sigma^2 D^{-1} E D^{-1}$, can be estimated as follows:

$$n\widehat{Var}(\widehat{\beta}) = \widehat{\sigma}_0^2 \, \widehat{D}^{-1} \widehat{E} \, \widehat{D}^{-1}. \tag{3.5}$$

Here, $\widehat{D}$ and $\widehat{E}$ are computed as follows (see Marazzi 1993, p. 145):

$$\widehat{D} = \frac{1}{n}(X^T \widehat{A} X) \ , \quad \widehat{E} = \frac{1}{n}(X^T \widehat{V} X),$$

where $\widehat{V} = diag(\sum_{j=1}^n \eta^2(\mathbf{x}_i, r_j(\widehat{\beta}_0)/\widehat{\sigma}_0)/n)$.

4. SOME MONTE CARLO RESULTS

In this section we want to compare the proposed estimator with the well-known estimators such as the LS, LTS, LMS, Huber-M, Mallows-type GM, and Schweppe-type GM, where LMS is the least median of squares estimator proposed by Rousseeuw (1984). To perform a Monte Carlo study we set up the following four situations:

Case 1) No leverage points and no outliers.

Case 2) No leverage points but with some outliers.

Case 3) Some bad leverage points.

Case 4) Some bad leverage points and some good leverage points.

The simulation model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i, \quad i = 1, \cdots, n. \tag{4.1}$$

The regression parameters are set as $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$ for simplicity. The number of observations is $n = 30$, and the number of replications is 500. The explanatory variables and error terms are generated as follows.

For Case 1, the explanatory variables x_{ij}'s are independently generated from the normal distribution $N(0, 3^2)$, for $i = 1, \cdots, 30$ and $j = 1, 2, 3$. These

explanatory variables are fixed in Case 1 and Case 2. The error term ϵ_i's are independently generated from the standard normal $N(0,1)$, and these errors are newly generated in each replication.

For Case 2, we use the same x_{ij}'s as in Case 1. But, the ϵ_i's are generated from the contaminated normal. The distribution function of ε-contaminated normal(CN(ε, σ)) is given by

$$F(\epsilon) = (1-\varepsilon)\Phi(\epsilon) + \varepsilon\Phi(\epsilon/\sigma),$$

and $\varepsilon = 0.2$ and $\sigma = 5$ are used in our Monte Carlo study. We expect that some mild or extreme outliers are generated from this distribution comparing to $N(0,1)$ of Case 1.

For Case 3, first we generate x_{ij}'s from $N(0,3^2)$ and ϵ_i's from $N(0,1)$. Then we obtain y_i's according to (4.1). Next, to make some bad leverage points, we select 6 points randomly, and replace these $(\mathbf{x}_i, y_i)$ by $(\mathbf{x}_i + 9, y_i)$.

In Case 4, we need some "bad" and some "good" leverage points. We first generate x_{ij}'s from $N(0,1)$ and replace 6(20%) randomly selected $\mathbf{x}_i$ by $\mathbf{x}_i + 30$. Next, we generate ϵ_i's from $N(0,1)$ and obtain y_i's according to (4.1). Thus, we have 6 good leverage points. To make some bad leverage points, we replace 3(10%) randomly selected points $(\mathbf{x}_i, y_i)$ from remaining observations not selected in good leverage points by $(\mathbf{x}_i, y_i)$. The latter 3 points must be bad leverage points.

Note that in Case 3 and Case 4, x_{ij}'s are newly generated in each replication according to these schemes. And, the above four situations are described in Figure 1 in the framework of simple regression.

We compared 7 estimators in our simulation study, and the results are summarized in Table 1. In the table we use the following abbreviations:

LSE : Least squares estimators

LMS : LMS estimators

LTS : LTS estimators

HME : Huber's M-estimators

PRE : Proposed one-step GM-estimators

SCF : Schweppe-type one-step GM-estimators

MAL : Mallows-type one-step GM-estimators

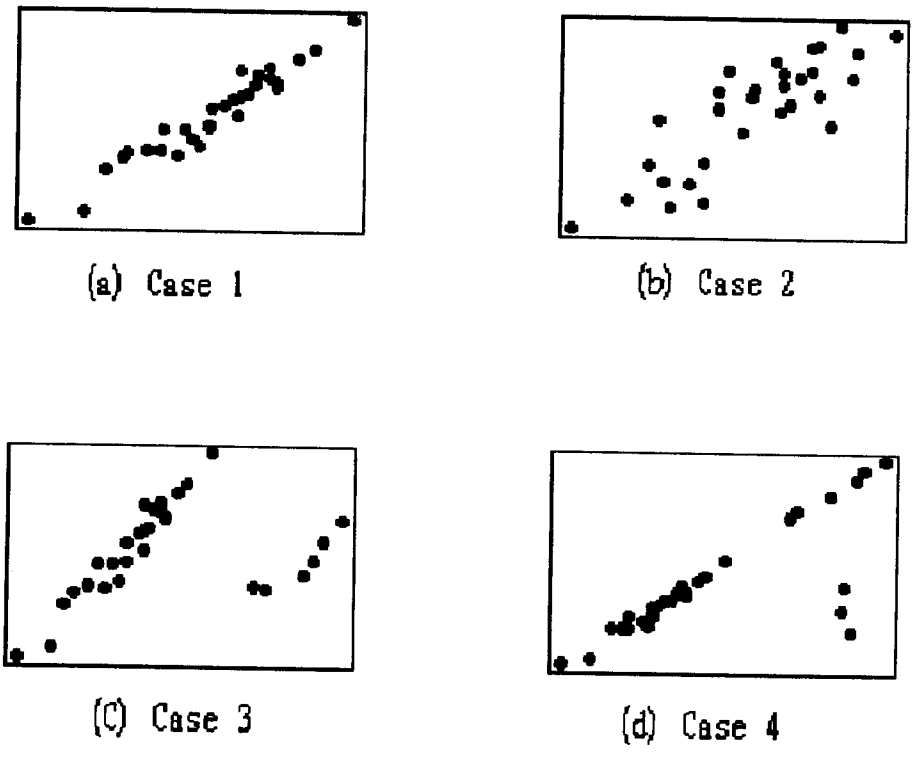

Figure 1. Examples of Simulated Data with 50 Points

We use the Huber's ψ-function with the tuning constant $c = 1.5$, for the M- and GM-estimators. The one-step M- and GM-estimators are obtained by the scoring method. In the proposed estimator, the tuning constant $a = 6.0$ is applied to compute weights. The constant $a = 6.0$ is chosen as a reasonable value through a simulation study in various situations. Note that in Case 1 and Case 2 $v(\mathbf{x}_i) = 1$, for all $i = 1, \cdots, 50$. Therefore, in these cases, the estimates SCF, MAL and HME are all identical. In the measure of leverageness $v(\mathbf{x}_i)$, we used the tuning constant $b = \chi^2_{3,0.975}$ for three types of GM-estimators. All computations in this Monte Carlo simulation were carried out on SUN SPARC10 AXIL311 by using S-PLUS (Ver.3.1: Release for SUN SPARC). The LTS and MVE estimates were obtained by S-PLUS functions *ltsreg* and *cov.mve* respectively. Also the S-PLUS function *rnorm* and *runif* were used to generate normal and contaminated normal random variates.

Table 1. Empirical MEAN and MSE

Est.	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	Est.	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$
Case 1 : No leverage points , $\epsilon \sim N(0,1)$					Case 2 : No leverage points , $\epsilon \sim CN(0.2,5)$				
	Empirical Mean					Empirical Mean			
LSE	0.98491	0.99985	0.99990	1.00318	LSE	1.00676	1.00443	1.00613	0.99150
LMS	0.96694	1.00544	0.99276	1.00448	LMS	1.01267	1.00675	1.00335	0.99262
LTS	0.98052	1.00164	0.99508	1.00263	LTS	1.02778	1.01227	1.00387	0.99343
HMS	0.98357	0.99902	0.99914	1.00411	HME	1.01289	1.00857	1.00065	0.99292
PRE	0.98343	0.99899	0.99906	1.00423	PRE	1.01347	1.00843	1.00036	0.99286
	Empirical MSE					Empirical MSE			
LSE	0.03785	0.00430	0.00333	0.00289	LSE	0.25144	0.02976	0.02558	0.02536
LMS	0.15056	0.01733	0.01219	0.01347	LMS	0.21048	0.02463	0.01782	0.02019
LTS	0.15753	0.01976	0.01217	0.01309	LTS	0.20698	0.02450	0.01702	0.01967
HME	0.04386	0.00604	0.00419	0.00422	HME	0.09404	0.01106	0.00824	0.00906
PRE	0.04407	0.00609	0.00422	0.00428	PRE	0.09357	0.01088	0.00814	0.00903
Case 3 : 6 Bad Leverage Points					Case 4 : Bad & 6 Good Leverage Points				
	Empirical Mean					Empirical Mean			
LSE	0.06845	0.17441	0.18995	0.18876	LSE	0.24803	0.45710	0.43255	0.41657
LMS	1.00556	1.00189	0.99153	1.00574	LMS	0.98901	0.99621	1.00057	1.00329
LTS	1.01285	1.00265	0.99256	0.99958	LTS	0.97735	0.99966	1.00045	0.99806
HME	0.89617	0.92286	0.91850	0.92107	HME	0.92151	0.94314	0.95018	0.94261
PRE	0.96629	0.98860	0.98262	0.98575	PRE	0.97469	0.98364	0.99247	0.98479
SCF	0.91972	0.94952	0.94503	0.94783	SCF	0.93009	0.95369	0.96173	0.95301
MAL	0.90619	0.92765	0.92397	0.92517	MAL	0.92703	0.94392	0.95140	0.94409
	Empirical MSE					Empirical MSE			
LSE	1.82701	0.76114	0.73969	0.73236	LSE	1.25474	0.51278	0.50651	0.54380
LMS	0.18369	0.02265	0.02556	0.02215	LMS	0.19005	0.01677	0.02893	0.02049
LTS	0.18971	0.02098	0.02165	0.03233	LTS	0.17781	0.01516	0.02867	0.01844
HME	0.08661	0.01920	0.02203	0.02420	HME	0.10911	0.01354	0.01789	0.01335
PRE	0.07072	0.01296	0.01508	0.01746	PRE	0.09603	0.00846	0.01412	0.00799
SCF	0.07984	0.01569	0.01799	0.02025	SCF	0.10739	0.01232	0.01655	0.01176
MAL	0.08392	0.01777	0.02036	0.02250	MAL	0.10774	0.01300	0.01719	0.01279

The minimum MSE in each column is underlined.

According to the simulation results in Table 1, the LSE has best performances in both empirical mean and MSE in Case 1, as expected. However, the LSE is easily broken in other cases, especially in Case 3 or 4. The results obtained by the high breakdown estimates(LTS and LMS) are very similar. But these estimates have high MSE, and the efficiencies of these estimates are

very low comparing with other robust estimates. In Case 3 and Case 4, SCF and MAL are only slightly better than the Huber's M-estimate and these two estimates have similar performances.

The proposed estimates (PRE) dominates the others in MSE except for Case 1. And, we can conclude that in general the proposed estimate PRE has better performances in MSE than other competitors under heavy-tailed error distributions or in the presence of leverage points.

APPENDIX

Proof of Theorem 1.

Assume that m $(p \leq m \leq [n/2] - p + 1)$ points are replaced by arbitrary values. Then, to prove that the proposed one-step estimator $\widehat{\beta}$ has a breakdown point of $([n/2] - p + 1)/n$, we have only to show that $||H_0^{-1} g_0||$ remains bounded, because

$$||\widehat{\beta}|| \leq ||\widehat{\beta}_0|| + \widehat{\sigma}_0 ||H_0^{-1} g_0||$$

and the initial (LTS) estimator $\widehat{\beta}_0$ has a breakdown point of $([(n-p)/2])/n$ which is greater than or equal to $([n/2] - p + 1)/n$. Also note that the scale estimator $\widehat{\sigma}_0$ has a breakdown point of $[(n-1)/2]/n$. That is, $||\widehat{\beta}_0||$ and $\widehat{\sigma}_0$ remain bounded with m contaminated points. Now if $\lambda_{min}(H_0)$ is positive, we have

$$||H_0^{-1} g_0|| \leq \frac{||g_0||}{\lambda_{min}(H_0)},$$

where $\lambda_{min}(H_0)$ is the minimum eigenvalue of H_0. Note that the MVE estimators $\mathbf{m_z}$ and $C_\mathbf{z}$ have a breakdown point of $([n/2] - p + 2)/n$. And, we have

$$||g_0|| = \left\| \sum_{i=1}^{n} \eta \left(\mathbf{x}_i, \frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0} \right) \mathbf{x}_i \right\| \leq \sum_{i=1}^{n} \left\| \eta \left(\mathbf{x}_i, \frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0} \right) \mathbf{x}_i \right\|. \qquad (A.1)$$

Here, $\|\eta(\mathbf{x}_i, r_i/\sigma_0)\mathbf{x}_i\| = \|w(\mathbf{x}_i, r_i/\sigma_0)\mathbf{x}_i\||\psi_i| < \infty$, since ψ_i is bounded and $\|w(\mathbf{x}_i, r_i/\sigma_0)\mathbf{x}_i\|$ is bounded unless $r_i = 0$. However $\psi_i = 0$ when $r_i = 0$. Therefore, the right side of (A.1) remains bounded as $|r_i|$ or $\|\mathbf{x}_i\|$ goes to infinity.

Thus, it is enough to show that $\lambda_{min}(H_0)$ is positive. For this, first note the following inequality:

$$\lambda_{min}(P + Q) \geq \lambda_{min}(P) + \lambda_{min}(Q). \qquad (A.2)$$

If Q is a positive semidefinite matrix, then $\lambda_{min}(P+Q) \geq \lambda_{min}(P)$. We can apply the inequality (A.2) in the following with $P = \sum_{i=1}^{p} d_i \mathbf{x}_i \mathbf{x}_i^T$, $Q = \sum_{i=p+1}^{n} d_i \mathbf{x}_i \mathbf{x}_i^T$, and $d_i = \sum_{j=1}^{n} \eta'(\mathbf{x}_i, r_j(\widehat{\beta}_0)/\widehat{\sigma}_0)/n$. That is,

$$\begin{aligned}
\lambda_{min}(H_0) &= \lambda_{min}\left\{\sum_{i=1}^{n} d_i \mathbf{x}_i \mathbf{x}_i^T\right\} \\
&= \lambda_{min}\left\{\sum_{i=1}^{p} d_i \mathbf{x}_i \mathbf{x}_i^T + \sum_{i=p+1}^{n} d_i \mathbf{x}_i \mathbf{x}_i^T\right\} \\
&\geq \lambda_{min}\left\{\sum_{i=1}^{p} d_i \mathbf{x}_i \mathbf{x}_i^T\right\} \\
&\geq \left[\min_{1\leq i\leq p}\{d_i\}\right] \cdot \lambda_{min}\left\{\sum_{i=1}^{p} \mathbf{x}_i \mathbf{x}_i^T\right\} \\
&> 0,
\end{aligned}$$

because the first p d_i's are positive and other d_i's are nonnegative by (C4). Moreover, $\lambda_{min}(\sum_{i=1}^{p} \mathbf{x}_i \mathbf{x}_i^T) > 0$ by (C3). Hence the theorem follows.

Proof of Theorem 2. Without loss of generality, let $\beta = \mathbf{0}$, so $\epsilon_i = y_i$. Note that this is possible since the one-step estimator $\widehat{\beta}$ is obtained only through residuals, so the proposed estimator is regression equivariant.

To prove the theorem, first consider the following fact which stems from Proposition 4.1 in Bickel (1975).

Under the assumptions (C2) and (C5) through (C9),

$$\sup\left\{n^{-1/2}\left|\sigma\sum_{i=1}^{n} x_{ik}[\eta(\mathbf{x}_i, r_i(\mathbf{t})) - \eta(\mathbf{x}_i, \epsilon_i)] + [\sum_{j=1}^{p} t_j \sum_{i=1}^{n} \eta'(\mathbf{x}_i, \epsilon_i) x_{ik} x_{ij}]\right|; \right.$$

$$\left.\max_{1\leq j\leq p} |t_j| \leq M \cdot n^{-1/2}\right\} \xrightarrow{p} 0 \quad (A.3)$$

for each $k = 1, 2, \cdots, p$.

Since $\widehat{\beta}_0 = O_p(n^{-1/2})$ by (C8), we can apply $t_j = \widehat{\beta}_{0j} - \beta_j = \widehat{\beta}_{0j}$, $j = 1, \cdots, p$ to (A.3). In (A.3) $\eta'(\mathbf{x}_i, \epsilon_i)$ can be replaced by $E[\eta'(\mathbf{x}_i, \epsilon_i)]$ because

$$n^{-1/2} \sum_{j=1}^{p} t_j \sum_{i=1}^{n} |\eta'(\mathbf{x}_i, \epsilon_i) - E[\eta'(\mathbf{x}_i, \epsilon_i)]|\, x_{ik} x_{ij} = o_p(1).$$

Here, the expectation E is taken over the error, and by assumption (C5) we can substitite this expectation by d_i, where $d_i = \sum_{j=1}^{n} \eta'(\mathbf{x}_i, r_j(\widehat{\beta}_0)/\widehat{\sigma}_0)/n$.

Thus, we have

$$\sup\left\{n^{-1/2}\left|\sigma\sum_{i=1}^{n}x_{ik}\left[\eta\left(\mathbf{x}_i,\frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0}\right)-\eta\left(\mathbf{x}_i,\frac{\epsilon_i}{\sigma}\right)\right]+\sum_{j=1}^{p}\widehat{\beta}_{0j}\sum_{i=1}^{n}d_ix_{ik}x_{ij}\right|\right\}\xrightarrow{p}0. \tag{A.4}$$

Now, the estimator given in (2.4) can be rewritten as

$$\widehat{\sigma}_0\sum_{i=1}^{n}x_{ik}\eta\left(\mathbf{x}_i,\frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0}\right)=\sum_{j=1}^{p}(\widehat{\beta}_j-\widehat{\beta}_{0j})\sum_{i=1}^{n}d_ix_{ik}x_{ij}, \tag{A.5}$$

for each $k=1,2,\cdots,p$. Thus, using (A.5), (A.4) can be reduced as follows.

$$n^{-1/2}\Big|\sigma\sum_{i=1}^{n}x_{ik}\Big[\eta\left(\mathbf{x}_i,\frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0}\right)-\eta\left(\mathbf{x}_i,\frac{\epsilon_i}{\sigma}\right)\Big]+\sum_{j=1}^{p}\widehat{\beta}_j\Big[\sum_{i=1}^{n}d_ix_{ik}x_{ij}\Big]$$

$$-\widehat{\sigma}_0\sum_{i=1}^{n}x_{ik}\eta\left(\mathbf{x}_i,\frac{r_i(\widehat{\beta}_0)}{\widehat{\sigma}_0}\right)\Big|\xrightarrow{p}0.$$

Note that $\widehat{\sigma}_0=\sigma+O_p(n^{-1/2})$ by (C9). Therefore, we have

$$n^{-1/2}\Big|\sum_{j=1}^{p}\widehat{\beta}_j\sum_{i=1}^{n}d_ix_{ik}x_{ij}-\sigma\sum_{i=1}^{n}x_{ik}\eta\left(\mathbf{x}_i,\frac{\epsilon_i}{\sigma}\right)\Big|\xrightarrow{p}0.$$

That is, in matrix representation with $\Psi(\epsilon/\sigma w)=(\psi(\epsilon_1/\sigma w_1),\cdots,\psi(\epsilon_n/\sigma w_n))^T$ and $W=diag(w_i)$. Here, for simplicity, we denote $w(\mathbf{x}_i,r_i(\widehat{\beta}_0)/\widehat{\sigma}_0)$ and $w(\mathbf{x}_i,\epsilon_i/\sigma)$ by $\widehat{w}_i$ and w_i, respectively.

$$n^{-1/2}\Big[H_0\widehat{\beta}-\sigma X^TW\Psi\left(\frac{\epsilon}{\sigma w}\right)\Big]\xrightarrow{p}0. \tag{A.6}$$

Since H_0/n converges in probability to D by (C5),

$$n^{-1/2}\cdot(H_0/n)^{-1}\left[H_0\widehat{\beta}-\sigma X^TW\Psi\left(\frac{\epsilon}{\sigma w}\right)\right]\xrightarrow{p}0.$$

Thus, we have

$$n^{1/2}\Big[\widehat{\beta}-\beta-\sigma H_0^{-1}X^TW\Psi\left(\frac{\epsilon}{\sigma w}\right)\Big]\xrightarrow{p}0. \tag{A.7}$$

By assumption (C5), (A.7) implies that

$$n^{1/2}[\widehat{\beta}-\beta]\overset{d}{\equiv}n^{1/2}\sigma H_0^{-1}X^TW\Psi\left(\frac{\epsilon}{\sigma w}\right)\overset{d}{\equiv}n^{-1/2}\sigma D^{-1}X^TW\Psi\left(\frac{\epsilon}{\sigma w}\right), \tag{A.8}$$

where $\stackrel{d}{\equiv}$ means *asymptotic equivalence in distribution.*

Thus, to prove the theorem we have only to show that

$$n^{-1/2}X^TW\Psi\left(\frac{\epsilon}{\sigma w}\right) \xrightarrow{d} N\Big(\mathbf{0},\ E\left[w^2(\mathbf{x},\epsilon/\sigma)\ \psi^2(\frac{\epsilon}{\sigma w(\mathbf{x},\epsilon/\sigma)})\mathbf{x}\mathbf{x}^T\right]\Big).$$

To show this, we apply the Cramér-Wold device (Serfling 1980, *p.* 18) and Theorem 10.9 (Lindeberg-Feller central limit theorem) of Arnold (1981, *p.* 156) to the $n^{-1/2}X^TW\Psi$. Here, note that the MVE estimators $\mathbf{m_z}$ and $C_\mathbf{z}$ used in the weight $w(\mathbf{x}_i,\epsilon_i)$ can be replaced by the mean vector and covariance matrix of $\mathbf{z}$, respectively, since the MVE estimators are consistent. Therefore, we may treat $w(\mathbf{x}_i,\epsilon_i)$'s as independent random variables. Now to apply the Cramér-Wold device, consider the following linear combination. For any $\mathbf{s}\in\mathbf{R}^p$,

$$\begin{aligned}
n^{-1/2}\mathbf{s}^TX^TW\Psi &= \sum_{j=1}^{p} n^{-1/2}\sum_{i=1}^{n} s_j w(\mathbf{x}_i,\epsilon_i/\sigma)x_{ij}\psi\left(\frac{\epsilon_i}{\sigma w(\mathbf{x}_i,\epsilon_i/\sigma)}\right)\\
&= \sum_{i=1}^{n} n^{-1/2}w(\mathbf{x}_i,\epsilon_i/\sigma)\left(\sum_{j=1}^{p} s_jx_{ij}\right)\psi\left(\frac{\epsilon_i}{\sigma w(\mathbf{x}_i,\epsilon_i/\sigma)}\right)\\
&= \sum_{i=1}^{n} a_{ni},
\end{aligned}$$

where $a_{ni} = n^{-1/2}w(\mathbf{x}_i,\epsilon_i/\sigma)\psi(\epsilon_i/\sigma w(\mathbf{x}_i,\epsilon_i/\sigma))\ \mathbf{s}^T\mathbf{x}_i$. Note that a_{ni}'s are independent. To check the Lindeberg's conditions, we compute the variance of a_{ni} as follows.

$$\begin{aligned}
\sigma_{ni}^2 &= Var(a_{ni})\\
&= E[Var(a_{ni}|X)] + Var[E(a_{ni}|X)]\\
&= \frac{1}{n}\mathbf{s}^TE\left[E\left(w^2(\mathbf{x}_i,\epsilon_i/\sigma)\psi^2\left(\frac{\epsilon_i}{\sigma w(\mathbf{x}_i,\epsilon_i/\sigma)}\right)\mathbf{x}_i\mathbf{x}_i^T\Big|X\right)\right]\mathbf{s}\\
&= \frac{1}{n}\mathbf{s}^TE\left[w^2(\mathbf{x}_i,\epsilon_i/\sigma)\psi^2\left(\frac{\epsilon_i}{\sigma w(\mathbf{x}_i,\epsilon_i/\sigma)}\right)\mathbf{x}_i\mathbf{x}_i^T\right]\mathbf{s},
\end{aligned}$$

since in the model (1.1) we assumed that the distribution of ϵ_i is symmetric about 0, ϵ_i and $\mathbf{x}_i$ are independent and $w(\mathbf{x}_i,\epsilon_i/\sigma)\psi(\epsilon_i/\sigma w(\mathbf{x}_i,\epsilon_i/\sigma))$ is an odd function of ϵ_i. Thus by (C7), $\lim_{n\to\infty}\max_{1\le i\le n}\sigma_{ni}^2 = 0$ and

$$\sum_{i=1}^{n}\sigma_{ni}^2 = \mathbf{s}^TE\left[w^2(\mathbf{x},\epsilon/\sigma)\psi^2\left(\frac{\epsilon}{\sigma w(\mathbf{x},\epsilon/\sigma)}\right)\mathbf{x}\mathbf{x}^T\right]\mathbf{s}.$$

Finally, for arbitrary positive δ

$$\begin{aligned} C_n &= \sum_{i=1}^{n} \int_{|a_{ni}|>\delta} a_{ni}^2 \, dF(\mathbf{x}_i, y_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} \int_{|a_{ni}|>\delta} \mathbf{s}^T eta^2(\mathbf{x}_i, \epsilon_i/\sigma) \mathbf{x}_i \mathbf{x}_i^T \mathbf{s} \, dF(\mathbf{x}_i, y_i) \\ &= \mathbf{s}^T E[\eta(\mathbf{x}, \epsilon_i/\sigma) \mathbf{x}\mathbf{x}^T I(|a_{n1}| > \delta)] \mathbf{s} \\ &\to 0 \end{aligned}$$

as n goes to infinity, since $\lim_{n\to\infty} \Pr(|a_{n1}| > \delta) = 0$. Thus, from (A.8)

$$n^{1/2}[\widehat{\beta} - \beta] \stackrel{d}{\equiv} \sigma D^{-1} \cdot N(\mathbf{0}, E).$$

Hence the theorem follows.

REFERENCES

(1) Arnold, S.F. (1981). *The Theory of Linear Models and Multivariate Analysis*, New York: John Wiley.

(2) Bickel, P.J. (1975). One-Step Huber Estimates in the Linear Model. *Journal of the American Statistical Association*, 70, 428-434.

(3) 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.

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

(5) Maronna, R.A. and Yohai, V.J. (1981). Asymptotic Behavior of General M-estimates for Regression and Scale with Random Carriers , *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 58, 7-20.

(6) Rousseeuw, P.J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79, 871-880.

(7) Serfling, R.J. (1980). *Approximation Theorems of Mathmetical Statistics*, New York: Wiley.

(8) 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.

(9) Statistical Sciences (1994). *S-PLUS for Windows User's Manual*, Siattle: Statistical Sciences.