Journal of the Korean
Statistical Society
Vol. 26, No. 1, 1997

A Study on a One-Step Pairwise GM-Estimator in Linear Models†

Moon Sup Song [1] and Jin Ho Kim [2]

Abstract

In the linear regression model $y_i = \alpha + \mathbf{x}_i{}^T\beta + \epsilon_i,\ i = 1, 2, \ldots, n$, the weighted pairwise absolute deviation (WPAD) estimator was defined by minimizing the dispersion function $D(\beta) = \sum\sum_{i<j} w_{ij}|r_j(\beta) - r_i(\beta)|$, where $r_i(\beta)$'s are residuals and w_{ij}'s are weights. This estimator can achieve bounded total influence with positive breakdown by choice of weights w_{ij}. In this paper, we consider a more general type of dispersion function than that of $D(\beta)$ and propose a *pairwise GM-estimator* based on the dispersion function. Under some regularity conditions, the proposed estimator has a bounded influence function, a high breakdown point, and asymptotically a normal distribution. Results of a small-sample Monte Carlo study are also presented.

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

---

†The present studies were supported in part by the Basic Science Research Institute Program, Ministry of Education, Korea, 1996, Project number BSRI-96-1415.

[1] Department of Statistics, Seoul National University, Seoul, 151-742, Korea.
[2] Korea Research Institute of Standards and Science, Taejon, 305-606, Korea.

1. INTRODUCTION

We consider the linear regression model

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

where $(\mathbf{x}_i, y_i), i = 1, 2, \ldots, n$ is a sequence of independent and identically distributed (iid) random variables with distribution function $F(\mathbf{x}, y)$ and density $f(\mathbf{x}, y)$, $\mathbf{x}_i$ is a $p \times 1$ vector of explanatory variables, and $(\alpha, \beta^T)^T$ is a $(p+1) \times 1$ vector of parameters. Assume that the errors (ϵ_i's) are iid, independent of $\mathbf{x}$, and symmetric about 0. In this paper we are mainly interested in the estimation of β. Given an estimator $(\hat{\alpha}, \hat{\beta})$, the residuals are denoted by $r_i(\hat{\alpha}, \hat{\beta}) = y_i - \hat{\alpha} - \mathbf{x}_i^T\hat{\beta}$ and $r_i(\hat{\beta}) = y_i - \mathbf{x}_i^T\hat{\beta}$.

The classical least squares (LS) estimator is optimal when the error terms are normal. However, it is well known that the LS estimator is very sensitive to outliers in the y-direction and high leverage points in the $\mathbf{x}$-direction. To overcome the nonrobustness of the LS estimator, many robust methods have been developed.

In the problem of robust regression, it is desirable to simultaneously achieve (a) a bounded influence function; (b) a high breakdown point (say .5); and (c) a high efficiency (say .95) vs. LS in the case of normal errors (Coakley and Hettmansperger, 1993; Yohai and Zamar, 1988). A bounded influence approach gives good stability and high efficiency against infinitesimal contamination, and thus it is a concept of local behavior. A breakdown point procedure can cope with a large fraction of bad outliers, and it is a concept of global stability.

The M-estimators for the regression model are robust to outliers in the y-direction, but the influence of bad leverage points is not bounded. The generalized M (GM) estimators were proposed to obtain bounded-influence estimators in both $\mathbf{x}$- and y- directions. But, unfortunately these bounded-influence estimators have breakdown points of at most $1/(p+1)$ (Simpson, Ruppert, and Carroll, 1992), which is considered as a serious deficiency.

To achieve high breakdown point, several estimators such as the least median of squares (LMS) and the least trimmed squares (LTS) estimators (Rousseeuw, 1984) have been proposed. But these estimators do not have bounded influence functions. Recently, Simpson *et al.* (1992) and Coakley and Hettmansperger (1993) suggested one-step GM-estimators based on high breakdown initial estimators. These estimators inherit the high breakdown points of initial estimators and have bounded influence functions. Moreover, they can attain the high efficiency of GM-estimators.

On the other hand, a general family of rank estimators was proposed by Jaeckel (1972). The use of Wilcoxon scores achieves good efficiency for a normal distribution and robustness against outlying in the y-direction but remains very sensitive to high leverage points. Hettmansperger and McKean (1978) showed that the Jaeckel's dispersion function with Wilcoxon scores is equivalent to the dispersion function of $\sum\sum_{i<j}|r_j(\beta) - r_i(\beta)|$. Sievers (1983) proposed a weighted rank estimator by minimizing the dispersion function

$$D(\beta) = \sum\sum_{i<j} w_{ij}|r_j(\beta) - r_i(\beta)|, \tag{1.2}$$

where w_{ij}'s are weights which may depend on the design matrix X. Recently, Naranjo and Hettmansperger (1994) proposed weights w_{ij} of the dispersion function (1.2) that achieve bounded total influence with positive breakdown. Hereafter, following them, we refer to the Naranjo-Hettmansperger estimator as the weighted pairwise absolute deviation (WPAD) estimator. This estimator can attain good efficiency for normal (typically 90-95 % in simple regression), but have a breakdown point of at most 1/3.

We consider a more general type of dispersion function than that of (1.2) and propose a *pairwise GM-estimator* based on the dispersion function. Under some regularity conditions, the proposed estimator has a bounded influence function, a high breakdown point, and asymptotically a normal distribution.

2. THE PROPOSED ONE-STEP PAIRWISE GM-ESTIMATOR

We now consider estimating β by the solution of the (vector) equation

$$g(\beta) = \sum\sum_{i<j} \eta\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\beta) - r_i(\beta)}{\sigma}\Big)(\mathbf{x}_j - \mathbf{x}_i) = \mathbf{0}, \tag{2.1}$$

where the η-function is defined by

$$\eta\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\beta) - r_i(\beta)}{\sigma}\Big) = w_{ij}\psi\Big(\frac{r_j(\beta) - r_i(\beta)}{\sigma}\Big)$$

with

$$w_{ij} = w\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\beta) - r_i(\beta)}{\sigma}\Big),$$

ψ is an odd and bounded function, and $\sigma^2 = Var\Big\{r_j(\beta) - r_i(\beta)\Big\} = 2Var\Big\{r_i(\beta)\Big\}$. Note that $g(\beta)$ is free of the intercept α which may be estimated as a second stage.

Assume that the weights satisfy $w_{ij} = w_{ji}$. Then $g(\beta)$ in (2.1) can be rewritten as

$$\begin{aligned} g(\beta) &= \sum\sum_{i<j} w_{ij}\psi\Big(\frac{r_j(\beta) - r_i(\beta)}{\sigma}\Big)(\mathbf{x}_j - \mathbf{x}_i) \\ &= \sum_i \Big\{\sum_j w_{ij}\psi\Big(\frac{r_j(\beta) - r_i(\beta)}{\sigma}\Big)\Big\}\mathbf{x}_i. \end{aligned} \tag{2.2}$$

The equation (2.2) is very similar to the estimating equation of the GM-estimator applied to the pairwise difference of residuals. Thus we shall refer to the estimator as the *pairwise GM-estimator.*

Let $\hat{\beta}_0$ be an initial estimator of β such as LTS estimator, which was proposed by Rousseeuw (1984). Let $\hat{\sigma}_0$ be a robust estimator of σ such as $\hat{\sigma}_0 = \sqrt{2} \times 1.4826 \times MAD\{r_i(\hat{\beta}_0)\}$, where MAD is the median absolute deviation defined as

$$MAD\{r_i(\hat{\beta}_0)\} = med_i\Big\{\Big|r_i(\hat{\beta}_0) - med_j\Big(r_j(\hat{\beta}_0)\Big)\Big|\Big\}.$$

Then the one-step estimator based on $\hat{\beta}_0$ can be obtained by taking a first-order Taylor-series expansion of the left side of (2.1) about β. Thus, the proposed one-step pairwise GM-estimator is given by

$$\hat{\beta} = \hat{\beta}_0 + \hat{\sigma}_0 H_0^{-1} g_0, \tag{2.3}$$

where

$$g_0 = \sum\sum_{i<j}\eta\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i) \tag{2.4}$$

and

$$H_0 = \sum\sum_{i<j}\eta'\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T. \tag{2.5}$$

Here η' denotes $\eta'(\mathbf{x}_i, \mathbf{x}_j, t) = (\partial/\partial t)\eta(\mathbf{x}_i, \mathbf{x}_j, t)$ and H_0 is obtained using the fact $(\partial/\partial\beta)\{r_j(\beta) - r_i(\beta)\} = \mathbf{x}_i - \mathbf{x}_j$.

We have to decide the weights w_{ij} and η- or ψ-function in order to obtain the proposed estimator $\hat{\beta}$ in (2.3). For the ψ-function, any robust M-estimator ψ-functions can be used. We used the Huber's ψ-function in a small-sample Monte Carlo study, which is presented in Section 4. Naranjo and Hettmansperger (1994), in WPAD estimator, suggested to use $w_{ij} = v(\mathbf{x}_i)v(\mathbf{x}_j)$, where $v(\mathbf{x})$ is a measure of leverageness defined as

$$v(\mathbf{x}) = min\Big[1, \Big\{\frac{b}{(\mathbf{x} - m_{\mathbf{x}})^T C_{\mathbf{x}}^{-1}(\mathbf{x} - m_{\mathbf{x}})}\Big\}^{r/2}\Big], \tag{2.6}$$

which was considered by Simpson *et al.* (1992). Here $m_{\mathbf{x}}$ and $C_{\mathbf{x}}$ are the minimum volume ellipsoid (MVE) estimators of location and covariance of $\mathbf{x}$, respectively, and b is a quantile of the chi-squared distribution with p degrees of freedom.

We want to downweight the leverage points and also outliers simultaneously. Thus, we propose to use the weights defined as follows:

$$\begin{aligned} w_{ij} &= w\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\beta) - r_i(\beta)}{\sigma}\Big) \\ &= \begin{cases} v(\mathbf{x}_i)v(\mathbf{x}_j) & , \ |\{r_j(\beta) - r_i(\beta)\}/\sigma| < a \\ 0 & , \ otherwise, \end{cases} \end{aligned} \tag{2.7}$$

with a constant a. Then, the proposed one-step pairwise GM-estimator (2.3) can be written as follows:

$$\hat{\beta} = \hat{\beta}_0 + \hat{\sigma}_0 H_0^{-1} g_0, \tag{2.8}$$

where g_0 is defined by (2.4) and

$$H_0 = \sum\sum\nolimits_{i<j} \hat{w}_{ij} \psi'\Big(\frac{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T \tag{2.9}$$

with

$$\begin{aligned} \hat{w}_{ij} &= w\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\Big) \\ &= \begin{cases} v(\mathbf{x}_i)v(\mathbf{x}_j) & , \ |\{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)\}/\hat{\sigma}_0| < a \\ 0 & , \ otherwise. \end{cases} \end{aligned}$$

Here ψ' denotes $\psi'(t) = (\partial/\partial t)\psi(t)$.

3. PROPERTIES OF THE ONE-STEP PAIRWISE GM-ESTIMATOR

In this section we investigate the equivariance and asymptotic properties of the proposed one-step pairwise GM-estimator. For the asymptotic properties, we are interested in the boundedness of the influence function, high breakdown point, and asymptotic normality of the proposed estimator. The definitions and discussion of affine, regression, and scale equivariance can be found in Rousseeuw and Leroy (1987).

We use the following assumptions in the theorems.

(A1) ψ is odd, bounded, and strictly increasing.

(A1)$'$ ψ is odd and bounded with properties such that ψ' & ψ'' exist, and $|\psi'(t)|$ & $|\psi''(t)|$ are bounded.

(A2) $\hat{\sigma}_0$ is affine invariant.

(A3) $\hat{\sigma}_0$ is scale equivariant.

Theorem 3.1. (a) If $\hat{\beta}_0$ is affine equivariant and Assumptions (A1) and (A2) hold, then the proposed estimator $\hat{\beta}$ is also affine equivariant. (b) If $\hat{\beta}_0$ is regression equivariant, then the proposed estimator $\hat{\beta}$ is also regression equivariant. (c) If $\hat{\beta}_0$ is scale equivariant and Assumptions (A1) and (A3) hold, then the proposed estimator $\hat{\beta}$ is also scale equivariant.

Proof. The proof follows in a straightforward way from the assumptions.

We now consider the influence function of the proposed one-step pairwise GM-estimator of (2.8). Then $(1/\binom{n}{2})g(\beta)$ in (2.1) has the functional form

$$g_0(\beta(F)) = \int\int \eta\Big(\mathbf{x}_1, \mathbf{x}_2, \frac{r_2(\beta(F)) - r_1(\beta(F))}{\sigma}\Big)(\mathbf{x}_2 - \mathbf{x}_1)dF(\mathbf{x}_2, y_2)dF(\mathbf{x}_1, y_1). \tag{3.1}$$

$g_0(F_n) = \mathbf{0}$ yields the pairwise GM-estimator $\hat{\beta}(F_n)$. We let F_0 be a fixed distribution representing the target model and let F_t be a t-contamination of F_0, i.e., $F_t = (1-t)F_0 + t\,\delta$, where $\delta = \delta(\mathbf{x}, y)$ denotes pointmass 1 at $(\mathbf{x}, y)$.

The following theorem gives the influence function of the proposed estimator $\hat{\beta}$ in (2.8) as defined by Hampel (1974). Our proof of Theorem 3.2 is mainly based on Simpson *et al.* (1992) and Song, Park, and Nam (1996).

The following assumptions are used in Theorem 3.2 .

(A4) $(1/\binom{n}{2})\sum\sum_{i<j} w_{ij}\psi'\Big(\frac{r_j(\beta) - r_i(\beta)}{\sigma}\Big)(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T \xrightarrow{p} H_0(\beta(F))$,
where

$$H_0(\beta(F)) = \int\int w_{ij}\psi'\Big(\frac{r_2(\beta(F)) - r_1(\beta(F))}{\sigma}\Big)(\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T \times dF(\mathbf{x}_2, y_2)dF(\mathbf{x}_1, y_1) \tag{3.2}$$

and $H_0(\beta(F))$ is a positive definite $p \times p$ matrix.

(A5) $||w(\mathbf{x}_i, \mathbf{x}_j, t)(\mathbf{x}_j - \mathbf{x}_i)||$ is bounded for any t and $\mathbf{x}$.

(A6) The initial estimator $\hat{\beta}_0$ has an influence function $IF(\mathbf{x}, y; \hat{\beta}_0)$.

Remark 1. For w_{ij} of (2.7) with $r = 2$ in (2.6), Assumption (A5) is satisfied provided that ψ is bounded.

Theorem 3.2. Under Assumptions (A1)′, (A4), (A5) and (A6), the influence function of the proposed estimator $\hat{\beta}$ of (2.8) is

$$IF(\mathbf{x}, y; \hat{\beta}) \;=\; 2\,\sigma_0 \left\{H_0\big(\beta(F_0)\big)\right\}^{-1} g_1(\beta(F_0)), \tag{3.3}$$

where

$$g_1(\beta(F_0)) = \int \eta\Big(\mathbf{x}_1, \mathbf{x}, \frac{r(\beta(F_0)) - r_1(\beta(F_0))}{\sigma_0}\Big)(\mathbf{x} - \mathbf{x}_1)dF_0(\mathbf{x}_1, y_1)$$

with $r(\beta(F_0)) = y - \mathbf{x}^T\beta(F_0)$, $r_1(\beta(F_0)) = y_1 - \mathbf{x}_1^T\beta(F_0)$, and $\sigma_0 = \sigma(F_0)$. And, the proposed estimator has a bounded influence function.

Proof. The proof is given in Appendix.

We now consider the finite sample breakdown point which was introduced by Donoho and Huber (1983). We will show that the proposed estimator (2.8) inherits the breakdown properties of the initial estimator $\hat{\beta}_0$ and the weights. The proof is similar to that of Simpson *et al.* (1992) and Song *et al.* (1996).

We need the following additional assumption to derive the finite sample breakdown point of the proposed estimator.

(**A7**) Without loss of generality, assume that the first $(p+1)$ observations are uncontaminated. Assume also that ψ is nondecreasing and the following assumptions hold for $1 \le i \ne j \le p+1$: (a) $\psi'(\{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)\}/\hat{\sigma}_0) \ge d > 0$ for some finite constant d, (b) $\hat{w}_{ij} > 0$, and (c) $\mathbf{x}_i \ne \mathbf{x}_j$ for at least one $i \ne j$.

Remark 2. For simplicity, let $\alpha = 0$. There are at least $n/2$ observations with $|r_i(\hat{\beta}_0)| \le a_0 \times med_i(|r_i(\hat{\beta}_0)|)$ for $a_0 > 1$. For sufficiently large n, $med_i(|r_i(\hat{\beta}_0)|) \approx MAD\{r_i(\hat{\beta}_0)\}$ because $med_i(r_i(\hat{\beta}_0)) \approx 0$. We can choose a_1 such that there are at least $n/2$ observations with $|r_i(\hat{\beta}_0)| \le a_1 \times MAD\{r_i(\hat{\beta}_0)\}$. Thus, there are approximately at least $n/2$ observations with $|r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)| \le d_1 \times MAD\{r_i(\hat{\beta}_0)\}$, where $2a_1 = d_1$. Moreover, when the Huber's ψ -function with tuning constant c with $d_1 \le \sqrt{2} \times 1.4826 \times c$ is used, there are approximately $n/2$ observations $\psi'(\{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)\}/\hat{\sigma}_0) > 0$. And, the weights w_{ij} of (2.7) also satisfy Assumption (A7) by a proper choice of a.

Theorem 3.3. Assume that $n \geq 2p+1$. Under Assumptions (A1)′, (A5), and (A7), the proposed estimator (2.8) has a breakdown point of $([n/2]-p)/n$.

Proof. The proof is given in Appendix.

Next, we derive the asymptotic distribution of the proposed estimator (2.8). Naranjo and Hettmansperger (1994) derived the asymptotic normality of the WPAD estimator of (1.2), which is based on Sievers (1983). Maronna and Yohai (1981) evaluated the asymptotic behavior of GM-estimators. Coakley and Hettmansperger (1993) and Simpson *et al.* (1992) discussed the asymptotic normality of the one-step GM-estimators.

First, we consider the large sample behavior of the proposed estimator. Our proof of Theorem 3.4 is mainly based on Simpson *et al.* (1992).

We assume the following conditions.

(A8) $w_{ij} = w\left(\mathbf{x}_i, \mathbf{x}_j, \dfrac{r_j(\beta)-r_i(\beta)}{\sigma}\right)$ is an even function of $r_j(\beta)-r_i(\beta)$ such that as $n \to \infty$, (a) $\left(1/\binom{n}{2}\right)\sum\sum_{i<j}\left(\dfrac{r_j(\beta)-r_i(\beta)}{\sigma}\right)^2 w_{ij}\,||\mathbf{x}_i-\mathbf{x}_j|| = O_p(1)$ and (b) $\left(1/\binom{n}{2}\right)\sum\sum_{i<j} w_{ij}\,||\mathbf{x}_i-\mathbf{x}_j||^3 = O_p(1)$.

Remark 3. (1) Take $r=2$ in (2.6). For w_{ij} of (2.7), (a) of Assumption (A8) is satisfied. (2) Note that $||\mathbf{x}||^2 v(\mathbf{x}) \leq k_0\,\lambda_{max}(C_x)$, where $\lambda_{max}(C_x)$ is the maximum eigenvalue of C_x and k_0 is a positive constant (Simpson *et al.*, 1992). w_{ij} in (2.7) gives $||\mathbf{x}_i-\mathbf{x}_j||^2\, w_{ij} \leq k_1\lambda_{max}(C_x)$ with a sufficiently large k_1, $0<k_1<\infty$. So, (b) of Assumption (A8) is satisfied provided that $\lambda_{max}(C_x)=O_p(1)$ and $\left(1/\binom{n}{2}\right)\sum\sum_{i<j}||\mathbf{x}_j-\mathbf{x}_i|| = O_p(1)$.

Theorem 3.4. Assume that Assumptions (A1)′, (A5), and (A8) hold. Suppose $n^{1/2}(\hat{\beta}_0-\beta)=O_p(1)$ and $n^{1/2}(\hat{\sigma}_0-\sigma)=O_p(1)$ with $\sigma>0$. Then for $\hat{\beta}$ given by (2.8),

$$n^{-3/2}H_0(\hat{\beta}-\beta) = n^{-3/2}\sigma\; g(\beta) + O_p(n^{-1/2}), \quad as\ n \to \infty, \tag{3.4}$$

where H_0 and $g(\beta)$ are defined by (2.9) and (2.1), respectively.

Proof. The proof is given in Appendix.

We now state the asymptotic normality of the proposed estimator (2.8). It is derived from (3.4) using the asymptotic normality of $g(\beta)$. For notational simplicity, we let $\epsilon = y-\mathbf{x}^T\beta(F)$ and $F_i = F(\mathbf{x}_i, y_i)$.

Theorem 3.5. Under the conditions of Theorem 3.4, Assumption (A4) and $\left(1/\binom{n}{2}\right)\sum\sum_{i<j}||\mathbf{x}_j - \mathbf{x}_i|| = O_p(1)$, for $\hat{\beta}$ given by (2.8),

$$n^{1/2}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma^2\Sigma), \tag{3.5}$$

where $\Sigma = \left\{H_0(\beta(F))\right\}^{-1} E \left\{H_0(\beta(F))\right\}^{-1}$ with $H_0(\beta(F))$ defined by (3.2) and

$$E = 4\int\Big\{\int\eta\Big(\mathbf{x}_1, \mathbf{x}_2, \frac{\epsilon_2 - \epsilon_1}{\sigma}\Big)(\mathbf{x}_2 - \mathbf{x}_1)dF_2\Big\} \times\Big\{\int\eta\Big(\mathbf{x}_1, \mathbf{x}_2, \frac{\epsilon_2 - \epsilon_1}{\sigma}\Big)(\mathbf{x}_2 - \mathbf{x}_1)dF_2\Big\}^T dF_1.$$

Proof. The proof is given in Appendix.

4. SMALL-SAMPLE MONTE CARLO STUDY

In this section we compare the small-sample behavior of the proposed (PGM) estimator and the well-known estimators such as the least square (LS), least trimmed squares (LTS), Huber-M (HM), Mallows-type one-step GM (MGM), Schweppe-type one-step GM (SGM), and WPAD estimators. To look at the effect of outliers in the y-direction as well as in the $\mathbf{x}$-direction, we consider the following four situations:

Case 1) No leverage points and no outliers.
Case 2) No leverage points but with some outliers only in the y-direction.
Case 3) Some bad leverage points, i.e., outliers in both $\mathbf{x}$- and y-directions.
Case 4) Some bad leverage points and some good leverage points.

We consider the simulation model

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \cdots, n. \tag{4.1}$$

The regression parameters are set as $\alpha = \beta = 1$ for simplicity. The sample size is $n = 30$, and 1,000 replications are performed. The explanatory variables are fixed. The error terms are generated from normal or heavy-tailed distributions. Some good or bad leverage points are constructed by transfering some observations.

For Case 1, the x_i's are fixed $(0.1,\ 0.2, \cdots, 3.0)$ and the error term ϵ_i's are randomly generated from the standard normal $N(0, 1)$. The y_i's are computed according to (4.1).

For Case 2, the x_i's are fixed as in Case 1. But the error term ϵ_i's are randomly generated from contaminated normal distributions. The distribution function of ϵ-contaminated normal $CN(\epsilon, (\mu, \sigma))$ is given by $F(x) = (1-\epsilon)\Phi(x) + \epsilon\Phi((x-\mu)/\sigma)$, where Φ is the standard normal cumulative. We consider two contaminated normal distributions : $CN(0.2, (0, 5))$ and $CN(0.1, (-10, 1))$.

For Case 3, we first generate the same x_i's and y_i's as in Case 1. To make some bad leverage points, we choose three points randomly, and replace these (x_i, y_i) by $(x_i + 5, y_i - 5)$.

For Case 4, first we generate the same x_i's and y_i's as in Case 1. To make some good leverage points, we replace three randomly selected (x_i, y_i) by $(x_i + 5, y_i + 5)$. To make some bad leverage points, we choose three points randomly from the remaining sample not selected as good leverage points, and replace these (x_i, y_i) by $(x_i + 5, y_i - 5)$.

The error terms and leverage points are newly generated in each case and in each replication. Figure 4.1 shows the examples of simulated data according to the given configuration.

For the ψ-function, Huber's ψ with the tuning constant $c = 1.5$ is used. To obtain the measure of leverageness, $v(\mathbf{x})$ in (2.6), $r = 2$ and $b = \chi^2(1, 0.975)$ are used. In the proposed estimator, the tuning constant $a = 2.7$ is applied to compute weights w_{ij}'s in (2.7). The constant $a = 2.7$ is determined as a reasonable value through many simulations in various situations.

The measures of performance evaluated are empirical means (MEAN) and mean squared errors (MSE), which are calculated as follows.

$$MEAN(\hat{\beta}) = \frac{1}{1000} \sum_{k=1}^{1000} \hat{\beta}^k$$

and

$$MSE(\hat{\beta}) = \frac{1}{1000} \sum_{k=1}^{1000} (\hat{\beta}^k - \beta)^2,$$

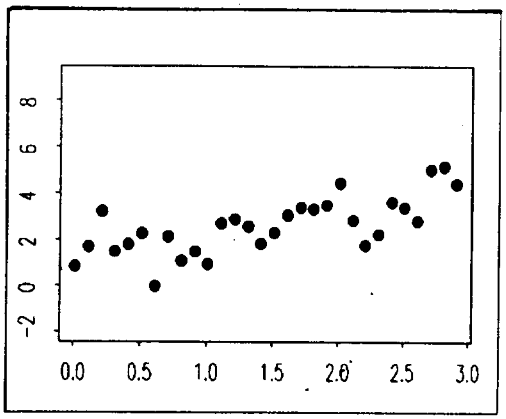
where $\hat{\beta}^k$ is the estimated value of $\hat{\beta}$ at kth replication. $MEAN(\hat{\alpha})$ and $MSE(\hat{\alpha})$ are calculated similarly.

The simulation was performed on a personal computer with a Pentium 100 MHz processor by using S-PLUS (Ver. 3.2 Release 1 for MS Windows 3.1 : 1994). The normal variates were generated by the S-PLUS function *rnorm*. The LS, LTS, and MVE estimators were calculated by S-PLUS functions *lsfit*, *ltsreg*, and *cov.mve*, respectively. The WPAD and proposed estimators were calculated on two stages because $D(\beta)$ of (1.2) and $g(\beta)$ of (2.1) are free of

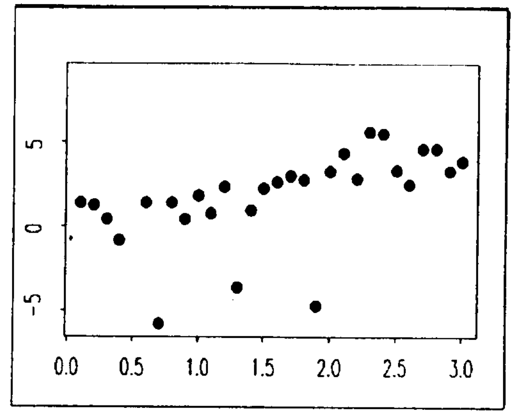
Table 1. Empirical MEAN and MSE

	Case 1		Case 2				Case 3		Case 4	
			$CN(0.2,(0,5))$		$CN(0.1,(-10,1))$					
Est.	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
				Empirical MEAN						
LS	1.00157	0.99333	0.97997	1.01362	-0.01131	1.00528	2.92975	-0.38867	2.14189	0.16421
LTS	1.01284	0.97903	0.95869	1.02235	1.00726	0.99231	1.03772	0.97447	1.01025	0.99220
HM	1.00234	0.99372	0.98174	1.01057	0.74866	1.01021	3.05889	-0.46874	1.79686	0.42798
MGM	〃	〃	〃	〃	〃	〃	1.81111	0.42210	1.51011	0.62926
SGM	〃	〃	〃	〃	〃	〃	1.82039	0.41569	1.43734	0.67994
WPAD	1.00344	0.99204	0.97931	1.00981	0.97416	1.00690	1.76523	0.45216	1.61762	0.61527
PGM	1.00297	0.99250	0.97675	1.00998	0.98411	1.00621	1.05574	0.95544	1.03068	0.97282
				Empirical MSE						
LS	<u>0.37326</u>	<u>0.20842</u>	0.89970	0.51178	1.55929	0.67243	2.01668	1.43738	1.30650	0.92234
LTS	0.83097	0.45622	0.86027	0.48503	0.80549	0.45915	0.89179	0.51847	0.66946	0.32650
HM	0.42955	0.23826	0.55869	0.32815	0.58882	0.29344	2.65062	1.86730	1.42385	1.00263
MGM	〃	〃	〃	〃	〃	〃	1.28860	0.87567	0.87667	0.60661
SGM	〃	〃	〃	〃	〃	〃	1.29700	0.88350	0.84719	0.58275
WPAD	0.38055	0.21348	0.53289	0.30845	0.46908	0.27187	1.05480	0.75348	0.86742	0.57394
PGM	0.39270	0.22143	<u>0.50714</u>	<u>0.29012</u>	<u>0.39756</u>	<u>0.21799</u>	<u>0.51314</u>	<u>0.33401</u>	<u>0.41219</u>	<u>0.23043</u>

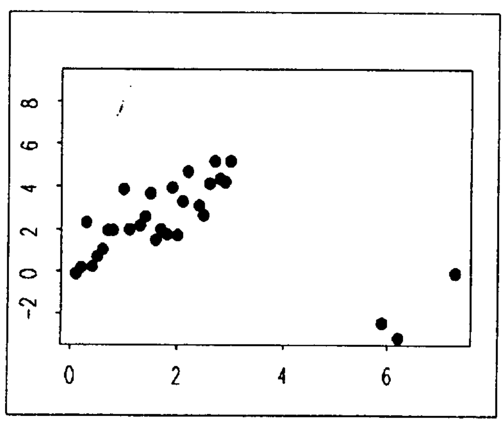
The minimum MSE in each column is underlined.



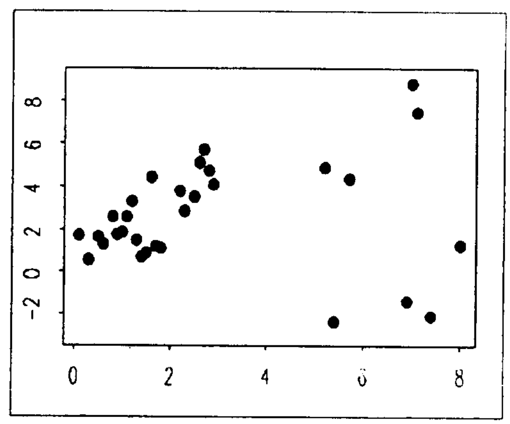
(1) Case 1 : No outliers and No leverage points



(2) Case 2 : Outliers($\varepsilon \sim CN(0.1, (-10, 1))$) and No leverage points



(3) Case 3 : Bad leverage points



(4) Case 4 : Good and bad leverage points

Figure 1. Examples of Simulated Data with 30 Points

the intercept α. The slope parameter β was first estimated, and then the intercept α was estimated as a location parameter at the second stage, which can be obtained by S-PLUS function *robloc*. The WPAD estimator was obtained by S-PLUS function *l1fit*.

The simulation results are given in Table 4.1. Note that in Case 1 and 2, the HM, MGM, and SGM estimators are all identical because the weights are

1, i.e., $v(\mathbf{x}_i) = 1$, for all $i = 1, \cdots, 30$. The LS estimator provides the best results when the errors are normally distributed, but breaks down easily when there are outliers in y or x. The high breakdown point estimator LTS shows good performances in the empirical mean, but has larger MSE than other robust estimators. The Huber's M-estimator has good efficiencies in Case 1 and 2, but is sensitive to outliers in Case 3 and 4. MGM and SGM estimators are slightly better than the Huber's M-estimator. The WPAD estimator has better performances than other robust estimators except the proposed PGM estimator, but still has many problems under extreme outliers in y or x.

The proposed estimator PGM dominates the others in the empirical MSE under extreme outliers or in the presence of leverage points. Moreover, the proposed PGM estimator has good performances in the empirical mean in all cases. In normal error situation, the PGM estimator has a high efficiency comparing with the LS estimator. The PGM estimator has similar performances to the WPAD estimator in the normal case.

APPENDIX : PROOFS OF THEOREMS

Proof of Theorem 3.2

The proposed estimator of (2.8) corresponds to the functional

$$\hat{\beta}(F) = \hat{\beta}_0(F) + \hat{\sigma}_0(F)\Big\{H_0(\hat{\beta}_0(F))\Big\}^{-1} g_0(\hat{\beta}_0(F)), \qquad (A.1)$$

where

$$H_0(\hat{\beta}_0(F)) = \int\int w\Big(\mathbf{x}_1, \mathbf{x}_2, \frac{r_2(\hat{\beta}_0(F)) - r_1(\hat{\beta}_0(F))}{\hat{\sigma}_0(F)}\Big)\psi'\Big(\frac{r_2(\hat{\beta}_0(F)) - r_1(\hat{\beta}_0(F))}{\hat{\sigma}_0(F)}\Big) \times(\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T dF(\mathbf{x}_2, y_2)dF(\mathbf{x}_1, y_1)$$

and

$$g_0(\hat{\beta}_0(F)) = \int\int \eta\Big(\mathbf{x}_1, \mathbf{x}_2, \frac{r_2(\hat{\beta}_0(F)) - r_1(\hat{\beta}_0(F))}{\hat{\sigma}_0(F)}\Big)(\mathbf{x}_2 - \mathbf{x}_1)dF(\mathbf{x}_2, y_2)dF(\mathbf{x}_1, y_1).$$

Replacing F by F_t, (A.1) becomes

$$\hat{\beta}(F_t) = \hat{\beta}_0(F_t) + \hat{\sigma}_0(F_t)\ \Big\{H_0(\hat{\beta}_0(F_t))\Big\}^{-1}\ g_0(\hat{\beta}_0(F_t)). \qquad (A.2)$$

Note that, under target model F_0, $\hat{\beta}(F_0) = \hat{\beta}_0(F_0) = \beta(F_0)$, i.e., $g_0(\hat{\beta}_0(F_0)) = 0$. Hence, taking derivatives of both sides of (A.2) with respect to t and evaluating at $t = 0$ give, under Assumption (A6),

$$IF(\mathbf{x}, y; \hat{\beta}) = IF(\mathbf{x}, y; \hat{\beta}_0) + \sigma_0\ H_0(\beta(F_0))^{-1}\ IF(\mathbf{x}, y; g_0). \qquad (A.3)$$

Next, we drive the influence function of g_0. Replacing F in $g_0(\hat{\beta}_0(F))$ by F_t gives

$$\begin{aligned}
&g_0(\hat{\beta}_0(F_t)) \\
&= (1-t)^2 \int\int \eta\Big(\mathbf{x}_1, \mathbf{x}_2, \frac{r_2(\hat{\beta}_0(F_t)) - r_1(\hat{\beta}_0(F_t))}{\hat{\sigma}_0(F_t)}\Big)(\mathbf{x}_2 - \mathbf{x}_1) dF_0(\mathbf{x}_2, y_2) dF_0(\mathbf{x}_1, y_1) \\
&\quad + t(1-t) \int \eta\Big(\mathbf{x}, \mathbf{x}_2, \frac{r_2(\hat{\beta}_0(F_t)) - r(\hat{\beta}_0(F_t))}{\hat{\sigma}_0(F_t)}\Big)(\mathbf{x}_2 - \mathbf{x}) dF_0(\mathbf{x}_2, y_2) \\
&\quad + t(1-t) \int \eta\Big(\mathbf{x}_1, \mathbf{x}, \frac{r(\hat{\beta}_0(F_t)) - r_1(\hat{\beta}_0(F_t))}{\hat{\sigma}_0(F_t)}\Big)(\mathbf{x} - \mathbf{x}_1) dF_0(\mathbf{x}_1, y_1).
\end{aligned}$$

Thus, for weights w_{ij} of (2.7),

$$\begin{aligned}
&IF(\mathbf{x}, y; g_0) \\
&= \lim_{t\to 0} \frac{g_0(\hat{\beta}_0(F_t)) - g_0(\hat{\beta}_0(F_0))}{t} = \lim_{t\to 0} \frac{g_0(\hat{\beta}_0(F_t))}{t} \\
&= 2 \int \eta\Big(\mathbf{x}_1, \mathbf{x}, \frac{r(\beta(F_0)) - r_1(\beta(F_0))}{\sigma_0}\Big)(\mathbf{x} - \mathbf{x}_1) dF_0(\mathbf{x}_1, y_1) \\
&\quad -2 \int\int \eta\Big(\mathbf{x}_1, \mathbf{x}_2, \frac{r_2(\beta(F_0)) - r_1(\beta(F_0))}{\sigma_0}\Big)(\mathbf{x}_2 - \mathbf{x}_1) dF_0(\mathbf{x}_2, y_2) dF_0(\mathbf{x}_1, y_1) \\
&\quad + \int\int IF(\mathbf{x}, y; \hat{w}_{12}) \psi\Big(\frac{r_2(\beta(F_0)) - r_1(\beta(F_0))}{\sigma_0}\Big)(\mathbf{x}_2 - \mathbf{x}_1) dF_0(\mathbf{x}_2, y_2) dF_0(\mathbf{x}_1, y_1) \\
&\quad -\sigma_0^{-1} \int\int w_{12} \psi'\Big(\frac{r_2(\beta(F_0)) - r_1(\beta(F_0))}{\sigma_0}\Big) IF(\mathbf{x}, y; \hat{\beta}_0)(\mathbf{x}_2 - \mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)^T \\
&\qquad\qquad \times dF_0(\mathbf{x}_2, y_2) dF_0(\mathbf{x}_1, y_1) \\
&\quad -\sigma_0^{-1} \int\int w_{12} \psi'\Big(\frac{r_2(\beta(F_0)) - r_1(\beta(F_0))}{\sigma_0}\Big)\Big(\frac{r_2(\beta(F_0)) - r_1(\beta(F_0))}{\sigma_0}\Big) IF(\hat{\sigma}_0) \\
&\qquad\qquad \times dF_0(\mathbf{x}_2, y_2) dF_0(\mathbf{x}_1, y_1) \\
&= 2\ g_1(\beta(F_0)) - \sigma_0^{-1}\ H_0(\beta(F_0))\ IF(\mathbf{x}, y; \hat{\beta}_0),
\end{aligned} \qquad (A.4)$$

because $\eta(\mathbf{x}_1, \mathbf{x}_2, t)$ and $w_{12}\psi'(t)t$ are odd functions about t by Assumption (A1)$'$ and (2.7), and the distribution of $r_2(\beta(F_0)) - r_1(\beta(F_0))$ is symmetric about 0. Here $w_{12} = w(\mathbf{x}_1, \mathbf{x}_2, (r_2(\beta(F_0)) - r_1(\beta(F_0)))/\sigma_0)$.

The influence function can be obtained by inserting (A.4) into (A.3).

Moreover, the influence function in (3.3) is bounded in Euclidean norm since $g_1(\beta(F_0))$ is bounded by Assumptions (A1)$'$ and (A5), and $H_0(\beta(F_0))$ is a positive definite matrix by Assumption (A4), which implies that $\lambda_{min}(H_0(\beta(F_0)))$ > 0, where λ_{min} is the minimum eigenvalue of $H_0(\beta(F_0))$. That is,

$$||IF(\mathbf{x}, y; \hat{\beta})|| \ < \ \infty. \qquad (A.5)$$

Proof of Theorem 3.3

We have $||\hat{\beta}|| \ \leq \ ||\hat{\beta}_0|| + \hat{\sigma}_0 \ ||H_0^{-1} \ g_0||$. And if $\lambda_{min}(H_0)$ is positive, we have $||\hat{\beta}|| \ \leq \ ||\hat{\beta}_0|| + \hat{\sigma}_0 \ ||g_0||/\lambda_{min}(H_0)$.

Assume that the $n - m$ observations are the good ones and that the remaining m observations ($p \leq m \leq [n/2] - p$) are replaced by arbitrary values. The initial LTS estimator and the scale estimator $\hat{\sigma}_0$ have breakdown points of $([n-p-1]/2)/n$ and $[(n-1)/2]/n$, respectively, which are greater than or equal to $([n/2]-p)/n$. Thus, $||\hat{\beta}_0||$ and $\hat{\sigma}_0$ remain bounded with m arbitrary values. The MVE estimator m_x and C_x have a breakdown point $([n/2]-p+1)/n$ (Rousseeuw and Leroy, 1987). Note also that by Assumptions (A1)$'$ and (A5)

$$||g_0|| \leq \sum\sum\nolimits_{i<j} ||\eta\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i)|| \ < \ \infty.$$

Thus, to prove that the proposed estimator $\hat{\beta}$ has a breakdown point of $([n/2]-p)/n$, it is sufficient to show that $\lambda_{min}(H_0)$ is positive. Note that $\lambda_{min}(P+Q) \ \geq \ \lambda_{min}(P) + \lambda_{min}(Q)$. Thus, if Q is a positive semidefinite matrix, then $\lambda_{min}(P+Q) \geq \lambda_{min}(P)$.

We have, for weights w_{ij} of (2.7),

$$\begin{aligned} \lambda_{min}(H_0) \ &= \ \lambda_{min}\Big\{\sum\sum\nolimits_{i<j} \hat{w}_{ij}\psi'\Big(\frac{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T\Big\} \\ &\geq \ \lambda_{min}\Big\{\sum\sum\nolimits_{1 \ \ i<j\leq p+1} \hat{w}_{ij}\psi'\Big(\frac{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T\Big\} \\ &\geq \ \Big[min_{1\leq i<j\leq p+1} \hat{w}_{ij}\psi'\Big(\frac{r_j(\hat{\beta}_0) - r_i(\hat{\beta}_0)}{\hat{\sigma}_0}\Big)\Big] \\ &\qquad \times \Big[\lambda_{min}\Big\{\sum\sum\nolimits_{1\leq i<j\leq p+1} (\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T\Big\}\Big] \\ &> \ 0 \end{aligned}$$

by Assumption (A7).

Proof of Theorem 3.4

We denote $r_i(\hat{\beta}_0) = y_i - \mathbf{x}_i^T\hat{\beta}_0$ and $r_i(\beta) = y_i - \mathbf{x}_i^T\beta$ by r_i and ϵ_i, respectively. Let $g(\beta, \hat{\sigma}_0) = \hat{\sigma}_0\sum\sum_{i<j}\eta(\mathbf{x}_i, \mathbf{x}_j, (\epsilon_j - \epsilon_i)/\hat{\sigma}_0)(\mathbf{x}_j - \mathbf{x}_i)$.

Taking a second-order Taylor series expansion about $\hat{\beta}_0$ under (A1)$'$ yields

$$\begin{aligned}
g(\beta, \hat{\sigma}_0) &= \hat{\sigma}_0\sum\sum_{i<j}\eta\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j - r_i}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i) \\
&+\sum\sum_{i<j}\eta'\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j - r_i}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T(\hat{\beta}_0 - \beta) \\
&+\frac{1}{2}\hat{\sigma}_0^{-1}\sum\sum_{i<j}\eta''\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\tilde{\beta}_0) - r_i(\tilde{\beta}_0)}{\hat{\sigma}_0}\Big)(\mathbf{x}_j - \mathbf{x}_i)\Big[(\mathbf{x}_j - \mathbf{x}_i)^T(\hat{\beta}_0 - \beta)\Big]^2 \\
&= H_0\ (\hat{\beta} - \beta) + O_p\Big(\hat{\sigma}_0^{-1}||\hat{\beta}_0 - \beta||^2\sum\sum_{i<j}\eta''\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{r_j(\tilde{\beta}_0) - r_i(\tilde{\beta}_0)}{\hat{\sigma}_0}\Big) \\
&\qquad \times||\mathbf{x}_j - \mathbf{x}_i||^3\Big), \qquad (A.6)
\end{aligned}$$

where $\tilde{\beta}_0$ is between β and $\hat{\beta}_0$. The last equality can be easily obtained by using (2.8) and Schwarz inequality.

Applying Taylor series expansion about σ yields, after some simplication,

$$\begin{aligned}
g(\beta, \hat{\sigma}_0) &= \hat{\sigma}_0\sum\sum_{i<j}\eta\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{\epsilon_j - \epsilon_i}{\sigma}\Big)(\mathbf{x}_j - \mathbf{x}_i) \\
&-(\hat{\sigma}_0 - \sigma)\sum\sum_{i<j}\eta'\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{\epsilon_j - \epsilon_i}{\sigma}\Big)\Big(\frac{\epsilon_j - \epsilon_i}{\sigma}\Big)(\mathbf{x}_j - \mathbf{x}_i) \\
&+\frac{1}{2}(\hat{\sigma}_0 - \sigma)^2\tilde{\sigma}^{-1}\sum\sum_{i<j}\eta''\Big(\mathbf{x}_i, \mathbf{x}_j, \frac{\epsilon_j - \epsilon_i}{\tilde{\sigma}}\Big)\Big(\frac{\epsilon_j - \epsilon_i}{\tilde{\sigma}}\Big)^2(\mathbf{x}_j - \mathbf{x}_i), \qquad (A.7)
\end{aligned}$$

where $\tilde{\sigma}$ is between σ and $\hat{\sigma}_0$.

Equating (A.6) and (A.7), for weights w_{ij} of (2.7),

$$\begin{aligned}
n^{-3/2}\ H_0\ (\hat{\beta} - \beta) &= n^{-3/2}\hat{\sigma}_0\sum\sum_{i<j}w_{ij}\psi\Big(\frac{\epsilon_j - \epsilon_i}{\sigma}\Big)(\mathbf{x}_j - \mathbf{x}_i) \\
&-n^{-3/2}\ (\hat{\sigma}_0 - \sigma)\sum\sum_{i<j}w_{ij}\psi'\Big(\frac{\epsilon_j - \epsilon_i}{\sigma}\Big)\Big(\frac{\epsilon_j - \epsilon_i}{\sigma}\Big)(\mathbf{x}_j - \mathbf{x}_i) \\
&+\frac{1}{2}n^{-3/2}\ (\hat{\sigma}_0 - \sigma)^2\tilde{\sigma}^{-1}\sum\sum_{i<j}\tilde{w}_{ij}\psi''\Big(\frac{\epsilon_j - \epsilon_i}{\tilde{\sigma}}\Big)\Big(\frac{\epsilon_j - \epsilon_i}{\tilde{\sigma}}\Big)^2(\mathbf{x}_j - \mathbf{x}_i) \\
&+O_p\Big(n^{-3/2}\ \hat{\sigma}_0^{-1}||\hat{\beta}_0 - \beta||^2\sum\sum_{i<j}\bar{w}_{ij}\psi''\Big(\frac{r_j(\tilde{\beta}_0) - r_i(\tilde{\beta}_0)}{\hat{\sigma}_0}\Big)||\mathbf{x}_j - \mathbf{x}_i||^3\Big),
\end{aligned}$$

$$(A.8)$$

where $\tilde{w}_{ij} = w(\mathbf{x}_i, \mathbf{x}_j, (\epsilon_j - \epsilon_i)/\tilde{\sigma})$ and $\bar{w}_{ij} = w(\mathbf{x}_i, \mathbf{x}_j, (r_j(\tilde{\beta}_0) - r_i(\tilde{\beta}_0))/\hat{\sigma}_0)$.

Next, we consider the asymptotic behavior of the right side of (A.8). $(1/\binom{n}{2})\sum\sum_{i<j} w_{ij}\psi((\epsilon_j - \epsilon_i)/\sigma)(\mathbf{x}_j - \mathbf{x}_i)$ is the U-type statistics with the kernel $w_{ij}\psi((\epsilon_j - \epsilon_i)/\sigma)(\mathbf{x}_j - \mathbf{x}_i)$ of degree 2 based on the ϵ's. We can derive the following fact which can be made by applying Theorem 3 of Section 3.4.2 in Lee (1990, p122).

$$(1/\tbinom{n}{2})\sum\sum_{i<j} w_{ij}\psi(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i) \xrightarrow{p} E_{F(\mathbf{x}_i, y_i)\times(\mathbf{x}_j, y_j)}\Big[w_{ij}\psi(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i)\Big].$$

Here,

$$\begin{aligned} &E_{F(\mathbf{x}_i, y_i)\times(\mathbf{x}_j, y_j)}\Big[w_{ij}\psi(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i)\Big] \\ &= E_{F\mathbf{x}}\Big[E_{F(\mathbf{x}_i, y_i)\times(\mathbf{x}_j, y_j)|\mathbf{x}}\Big[w_{ij}\psi(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i)\Big]\Big] \\ &= 0, \end{aligned} \qquad (A.9)$$

since the distribution of $\epsilon_j - \epsilon_i$ is the symmetric about 0 and $w_{ij}\psi((\epsilon_j - \epsilon_i)/\sigma)$ is an odd function of $\epsilon_j - \epsilon_i$ by Assumption (A1)$'$ and (2.7). Hence, by the assumption on $\hat{\sigma}_0$,

$$n^{-3/2}(\hat{\sigma}_0 - \sigma)\sum\sum_{i<j} w_{ij}\psi(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i) = O_p(n^{-1/2}).$$

Similarly, we can drive

$$(1/\tbinom{n}{2})\sum\sum_{i<j} w_{ij}\psi'(\frac{\epsilon_j - \epsilon_i}{\sigma})(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i)$$

$$\xrightarrow{p} E_{F(\mathbf{x}_i, y_i)\times(\mathbf{x}_j, y_j)}\Big[w_{ij}\psi'(\frac{\epsilon_j - \epsilon_i}{\sigma})(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i)\Big] = \mathbf{0},$$

since $w(\mathbf{x}_i, \mathbf{x}_j, t)\psi'(t)t$ is an odd function of t by Assumption (A1)$'$ and (2.7). Hence, by the assumption on $\hat{\sigma}_0$,

$$n^{-3/2}\ (\hat{\sigma}_0 - \sigma)\sum\sum_{i<j} w_{ij}\psi'(\frac{\epsilon_j - \epsilon_i}{\sigma})(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i) = O_p(n^{-1/2}),$$

Moreover, under by the assumptions on $\hat{\sigma}_0$, $\hat{\beta}_0$, (A1)$'$, and (A8),

$$n^{-3/2}\ (\hat{\sigma}_0 - \sigma)^2\tilde{\sigma}^{-1}\sum\sum_{i<j} \tilde{w}_{ij}\psi''(\frac{\epsilon_j - \epsilon_i}{\tilde{\sigma}})(\frac{\epsilon_j - \epsilon_i}{\tilde{\sigma}})^2(\mathbf{x}_j - \mathbf{x}_i) = O_p(n^{-1/2})$$

and

$$n^{-3/2}\ \hat{\sigma}_0^{-1}||\hat{\beta}_0-\beta||^2\sum\sum\nolimits_{i<j}\bar{w}_{ij}\psi''\Big(\frac{r_j(\tilde{\beta}_0)-r_i(\tilde{\beta}_0)}{\hat{\sigma}_0}\Big)||\mathbf{x}_j-\mathbf{x}_i||^3=O_p(n^{-1/2}).$$

Hence the theorem follows.

The following lemma will be used in the proof of Theorem 3.5. The proof of the lemma is very similar to that of Lemma A.1 of Simpson *et al.* (1992).

Lemma A.1. Suppose $n^\tau(\hat{\beta}_0-\beta)=O_p(1)$ and $n^\tau(\hat{\sigma}_0-\sigma)=O_p(1)$ with $\sigma>0$ and $\tau>0$. Let $\{c_{ij}\}$ be a sequence of finite constants. If a measurable function u and h satisfies the following condition

$$\Big|(u(s)-u(t))h(\mathbf{x}_i,\mathbf{x}_j)\Big|\le L\ |s-t|/(1+|t|),\quad all\ (s,t)$$

for a finite constant L, then

$$\begin{aligned}\sum\sum\nolimits_{i<j}|c_{ij}|\ &\Big|u(\frac{r_j-r_i}{\hat{\sigma}_0})h(\mathbf{x}_i,\mathbf{x}_j)-E_{F(\mathbf{x}_i,y_i)\times(\mathbf{x}_j,y_j)}\Big[u(\frac{\epsilon_j-\epsilon_i}{\sigma})h(\mathbf{x}_i,\mathbf{x}_j)\Big]\Big|\\ \le\sum\sum\nolimits_{i<j}|c_{ij}|\ &\Big|u(\frac{\epsilon_j-\epsilon_i}{\sigma})h(\mathbf{x}_i,\mathbf{x}_j)-E_{F(\mathbf{x}_i,y_i)\times(\mathbf{x}_j,y_j)}\Big[u(\frac{\epsilon_j-\epsilon_i}{\sigma})h(\mathbf{x}_i,\mathbf{x}_j)\Big]\Big|\\ &+O_p\Big(n^{-\tau}\sum\sum\nolimits_{i<j}|c_{ij}|\ (1+||\mathbf{x}_j-\mathbf{x}_i||)\Big)\end{aligned}$$

with $r=y-\mathbf{x}^T\hat{\beta}_0$.

Proof. The condition implies

$$\begin{aligned}&\Big|\Big(u(\frac{r_j-r_i}{\hat{\sigma}_0})-u(\frac{\epsilon_j-\epsilon_i}{\hat{\sigma}_0})\Big)h(\mathbf{x}_i,\mathbf{x}_j)\Big|\\ &\le L\ \hat{\sigma}_0^{-1}|(\mathbf{x}_j-\mathbf{x}_i)^T(\beta-\hat{\beta}_0)|\Big/(1+\hat{\sigma}_0^{-1}|\epsilon_j-\epsilon_i|)\\ &\le L\ ||\mathbf{x}_j-\mathbf{x}_i||\ ||\beta-\hat{\beta}_0||\ \hat{\sigma}_0^{-1}.\end{aligned}$$

And,

$$\begin{aligned}&\Big|\Big(u(\frac{\epsilon_j-\epsilon_i}{\hat{\sigma}_0})-u(\frac{\epsilon_j-\epsilon_i}{\sigma})\Big)h(\mathbf{x}_i,\mathbf{x}_j)\Big|\\ &\le L\ \Big|\frac{\epsilon_j-\epsilon_i}{\hat{\sigma}_0}-\frac{\epsilon_j-\epsilon_i}{\sigma}\Big|\Big/(1+\sigma^{-1}|\epsilon_j-\epsilon_i|)\\ &\le L\ |\sigma-\hat{\sigma}_0|\ \hat{\sigma}_0^{-1}.\end{aligned}$$

Hence,

$$\sum\sum_{i<j}|c_{ij}|\ \Big|\Big(u(\frac{r_j - r_i}{\hat{\sigma}_0}) - u(\frac{\epsilon_j - \epsilon_i}{\sigma})\Big)h(\mathbf{x}_i, \mathbf{x}_j)\Big|$$
$$\leq\ L\ \hat{\sigma}_0^{-1}\sum\sum_{i<j}\Big\{|c_{ij}|\ (||\mathbf{x}_j - \mathbf{x}_i||\ ||\beta - \hat{\beta}_0|| + |\sigma - \hat{\sigma}_0|)\Big\}$$
$$\leq\ L\ \hat{\sigma}_0^{-1}\ (||\beta - \hat{\beta}_0|| + |\sigma - \hat{\sigma}_0|)\ \sum\sum_{i<j}\Big\{|c_{ij}|\ (1 + ||\mathbf{x}_j - \mathbf{x}_i||)\Big\}.$$

Thus, the lemma can be easily derived by using last inequality.

Proof of Theorem 3.5

First, we derive the asymptotic normality of $n^{-3/2}g(\beta)$ in (3.4) as Theorem 1 of Section 3.2.1 in Lee (1990, p76). Without loss of generality, let $\sigma = 1$. Let $H_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n} h^{(1)}(\epsilon_i)$ and $H_n^{(2)} = (1/\binom{n}{2})\sum\sum_{i<j}h^{(2)}(\epsilon_i, \epsilon_j)$, where

$$h^{(1)}(\epsilon_i) = E_{F(\mathbf{x}_j, y_j)}\Big[\eta(\mathbf{x}_i, \mathbf{x}_j, \epsilon_j - \epsilon_i)(\mathbf{x}_j - \mathbf{x}_i)\Big]$$

and

$$h^{(2)}(\epsilon_i, \epsilon_j) = \eta(\mathbf{x}_i, \mathbf{x}_j, \epsilon_j - \epsilon_i)(\mathbf{x}_j - \mathbf{x}_i) - h^{(1)}(\epsilon_i) - h^{(1)}(\epsilon_j).$$

We may write

$$(1/\tbinom{n}{2})g(\beta) = 2H_n^{(1)} + H_n^{(2)}. \qquad (A.10)$$

We have

$$Cov(h^{(1)}(\epsilon_1), h^{(2)}(\epsilon_1, \epsilon_2)) = \int h^{(1)}(\epsilon_1)\Big\{\int h^{(2)}(\epsilon_1, \epsilon_2)^T dF_2\Big\}dF_1 = 0,$$

since $E_{F(\mathbf{x}_1, y_1)}\Big[h^{(1)}(\epsilon_1)\Big] = 0$ by (A.9) and $E_{F(\mathbf{x}_2, y_2)}\Big[h^{(2)}(\epsilon_1, \epsilon_2)\Big] = 0$. Hence, $Cov(H_n^{(1)}, H_n^{(2)}) = 0$. And,

$$\begin{aligned} Var(H_n^{(1)}) &= Cov\Big(\frac{1}{n}\sum_{i=1}^{n} h^{(1)}(\epsilon_i),\ \frac{1}{n}\sum_{i=1}^{n} h^{(1)}(\epsilon_i)\Big) \\ &= \frac{1}{n}\ E_{F(\mathbf{x}_1, y_1)}\Big[h^{(1)}(\epsilon_1)\ h^{(1)}(\epsilon_1)^T\Big] \end{aligned}$$

because $(\mathbf{x}_i, y_i)$'s are independent and $E_{F(\mathbf{x}_1, y_1)}\Big[h^{(1)}(\epsilon_1)\Big] = 0$. And,

$$\begin{aligned} Var(H_n^{(2)}) &= (1/\tbinom{n}{2}^2)\sum\sum_{i<j}\sum\sum_{k<l}Cov(h^{(2)}(\epsilon_i, \epsilon_j)\ ,\ h^{(2)}(\epsilon_k, \epsilon_l)) \\ &= (1/\tbinom{n}{2})\binom{2}{1}\binom{n-2}{1}\sigma_1^2 + (1/\tbinom{n}{2})\sigma_2^2, \end{aligned}$$

where $\sigma_1^2 = Cov(h^{(2)}(\epsilon_1, \epsilon_2)\ ,\ h^{(2)}(\epsilon_1, \epsilon_3))$ and $\sigma_2^2 = Cov(h^{(2)}(\epsilon_1, \epsilon_2)\ ,\ h^{(2)}(\epsilon_1, \epsilon_2))$. But, $\sigma_1^2 = 0$ and $\sigma_2^2 < \infty$ by Assumptions (A1)$'$ and (A5). So, $Var(n^{1/2}\ H_n^{(2)}) = O_p(n^{-1})$. Thus, by the multivariate Slutsky's theorem, the asymptotic behaviour of $2 \times n^{-3/2} g(\beta)$ is the same as that of $2 \times n^{1/2} H_n^{(1)}$. From the multivariate version of the central limit theorem for the independent random variate $h^{(1)}(\epsilon_1), \cdots, h^{(1)}(\epsilon_n)$,

$$2n^{-3/2} g(\beta) \xrightarrow{d} N_p(\mathbf{0}, E).$$

Second, let $u(t) = w(\mathbf{x}_i, \mathbf{x}_j, t)\psi'(t)$ with w_{ij} of (2.7), $h(\mathbf{x}_i, \mathbf{x}_j) = (x_{jk} - x_{ik})(x_{jl} - x_{il})$, and $c_{ij} = \binom{n}{2}^{-1}$. Then condition in Lemma 1 is satisfied by Assumptions (A1)$'$ and (A8). So, the Lemma A.1 gives that

$$\begin{aligned}
&(1/\tbinom{n}{2})\sum\sum\nolimits_{i<j} \hat{w}_{ij}\psi'(\frac{r_j - r_i}{\hat{\sigma}_0})(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T \\
&\overset{d}{=} E_{F(\mathbf{x}_i, y_i)\times(\mathbf{x}_j, y_j)}\left[w_{ij}\psi'(\frac{\epsilon_j - \epsilon_i}{\sigma})(\mathbf{x}_j - \mathbf{x}_i)(\mathbf{x}_j - \mathbf{x}_i)^T\right] + O_p(n^{-1/2})
\end{aligned}$$

by Assumption (A4). Here $\overset{d}{=}$ means asymptotic equivalence in distribution. Thus, $2n^{-3/2} H_0 \overset{d}{=} H_0(\beta(F))$.

Hence, the theorem follows.

REFERENCES

(1) Andrew, D.F. (1974). A Robust Method for Mutiple Linear Regression. *Technometric*, **16**, 523-531.

(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) De Jongh, P.J., De Wet, T., and Welsh, A.H. (1987). Mallows-Type Bounded-Influence Trimmed Means, *Journal of the American Statistical Association*, **84**, 805-810.

(5) Donoho, D.L. and Huber, P.J. (1983). The Notion of Breakdown Point. In *A Festschrift for Erich Lehmann*, edited by P. Bickel, K. Doksum and J.L. Hodges, Jr., Belmont, CA: Wadsworth.

(6) Hampel, F.R. (1971). A General Qualitative Definition of Robustness, *Annals of Mathmetical Statistics*, **42**, 1887-1896.

(7) Hampel, F.R. (1974). The Influence Curve and Its Role in Robust Estimation, *Journal of the American Statistical Association*, **69**, 383-393.

(8) Hettmansperger, T.P. and McKean, J.W. (1978). Statistical Inference Based on Ranks, *Psychometrica*, **43**, 69-79.

(9) Huber, P.J. (1973). Robust Regression: Asymptotics, Conjectures, and Monte Carlo, *The Annals of Statistics,* **1**, 799-821.

(10) Jaeckel, L.A. (1972). Estimation Regression Coefficients by Minimizing the Dispersion of the Residuals, *Annals of Mathmetical Statistics,* **43**, 1441-1458.

(11) Jurekova, J., and Portnoy, S. (1987). Asymptotics for One-Step M-Estimators in Regression with Application to Combining Efficiency and High Breakdown Point, *Communications in Statistics, Theory and Methods*, **16**(8), 2187-2199.

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

(13) Lee, A. J. (1990). *U-statistics Theory and Practice*, Marcel Dekker.

(14) 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.

(15) Naranjo, J.D. and Hettmansperger, T.P. (1994). Bounded Influence Rank Regression, *Journal of Royal Statistical Society, Series B*, **56**(1), 209-220.

(16) Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection,* New York: John Wiley.

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

(18) Rousseeuw, P.J. and Yohai, V. (1984). Robust Regression by Means of S-Estimators. In *Robust and Nonlinear Time Series*, edited by J. Franke, W. Hardle, and R.D. Martin, Lectures Notes in Statistics No. 26, 256-272. New York: Springer-Verlag.

(19) Sievers G. L. (1983). A Weighted Dispersion Function for Estimation in Linear Models, *Communication in Statistics - Theory and Methods*, **12**(10), 1161-1179.

(20) 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.

(21) Song, M.S., Park, Ch.S., and Nam, H.S. (1996). A High Breakdown and Efficient GM-Estimator in Linear Models, *Journal of the Korean Statistical Society*, in Proceeding.

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

(23) Yohai, V.J., and Zamar, R.H. (1988). High Breakdown Point Estimates of Regression by Means of the Minimization of an Efficient Scale, *Journal of the American Statistical Association*, **83**, 406-413.