The Annals of Statistics
1998, Vol. 26, No. 3, 1170–1189

BREAKDOWN PROPERTIES OF LOCATION *M*-ESTIMATORS[1]

BY JIAN ZHANG AND GUOYING LI

Academia Sinica

In this article, we consider the asymptotic behavior of three kinds of sample breakdown points. It is shown that for the location M-estimator with bounded objective function, both the addition sample breakdown point and the simplified replacement sample breakdown point strongly converge to the gross-error asymptotic breakdown point, whereas the replacement sample breakdown point strongly converges to a smaller value. In addition, it is proved that under some regularity conditions these sample breakdown points are asymptotically normal. The addition sample breakdown point has a smaller asymptotic variance than the simplified replacement sample breakdown point. For the commonly used redescending M-estimators of location, numerical results indicate that among the three kinds of sample breakdown points, the replacement sample breakdown point has the largest asymptotic variance.

1. Introduction. An important concept in robust statistics is the breakdown point, which measures the ability of a statistic to resist the outliers contained in the data set. The asymptotic and finite sample versions of this concept are owing to Hampel (1968, 1971) and Donoho and Huber (1983), respectively. The concept was originally used to study the global robustness of an estimator. Recently, this concept has been improved and extended to other situations including test statistics, nonlinear regression and test decisions [see, e.g., He, Simpson and Portnoy (1990), Coakley and Hettmansperger (1992), Stromberg and Ruppert (1992), Sakata and White (1995) and Zhang (1996)]. There has been much recent interest in positive-breakdown estimators for location, covariance and regression parameters as well [see, e.g., Rousseeuw (1994) for a recent survey]. However, there is not much work on the breakdown points of the location M-estimators with bounded objective functions. These are important and frequently adopted robust estimators; see Mosteller and Tukey [(1977), Chapters 10 and 14], Huber [(1981), Section 4.8] and Hampel, Ronchetti, Rousseeuw and Stahel [(1986), Section 2.6]. A recent application was provided by Okafor (1990). In this article, we are concerned with the asymptotic behavior of three kinds of sample breakdown points of location M-estimators, especially those of the redescending M-estimators.

To motivate the questions studied in this article, we first introduce some notation. Suppose that $X = (x_1, \dots, x_n)$ is a sample of size n. We consider the problem of estimating or testing a hypothesis concerning a p-dimensional parameter θ. In the following text, $T = T(X)$ denotes an estimator or a test

Received April 1994; revised October 1997.

[1]Supported in part by the National Natural Science Foundation of China and the Postdoctoral Science Foundation of China.

AMS 1991 *subject classification*. Primary 62F35.

Key words and phrases. Sample breakdown point, redescending M-estimator, asymptotics.

statistic of θ. In practical situations, the data set X often suffers from recording errors, transmission errors, "fudging" and so on. So what the data analysts obtained may not be X, but a contaminated sample, say Y. It seems difficult to clean Y because the unfortunate data analysts do not know which values in Y are contaminated [see Coakley and Hettmansperger (1992) for a vivid description]. There are two approaches to imitating the contamination of a sample: the replacement contamination and the addition contamination [see Donoho and Huber (1983)]. Intuitively, the former seems to be more consistent with what happens in practical situations, but the latter is simpler.

How large is the difference between the sample breakdown points based on the above contamination approaches? This question of high importance to data analysts has not been convincingly answered yet. We study this problem for the one-dimensional location M-estimators. An important finding in this study is that for the redescending M-estimators of location, the addition and simplified replacement sample breakdown points are strongly consistent with the same asymptotic breakdown point, whereas the replacement sample breakdown point converges to a smaller asymptotic breakdown point. This implies that, compared with the addition sample breakdown point, the replacement is relatively conservative in describing the breakdown robustness even in the large sample case. Similar results hold for the sample breakdown points of some tests discussed in Zhang (1996). These results have some natural extensions to the p-dimensional case.

In the following discussion, we take the replacement contamination as an example to define the sample breakdown point. Let $\mathscr{X}_m(X)$ be the set of all Y which result from replacing m observations in X by any values. For the sake of notational simplicity, we also let $\mathscr{X}_m$ denote the class of all measurable vector maps $Y = Y(X)$ from R^n (the n-dimensional Euclidean space) to R^n which change at most m components of X. Following Donoho and Huber (1983), for each subset Θ_0 of R^p (the p-dimensional Euclidean space with the L_2 norm $\|\cdot\|$), we can define the replacement sample breakdown point

$$\begin{aligned}\hat{\varepsilon}_{Rn}(\Theta_0) &= \hat{\varepsilon}_{Rn}(X, T, \Theta_0)\\ &= \frac{1}{n}\min\{0 \le m \le n\colon T(Y) \notin \Theta_0 \text{ for some } Y \in \mathscr{X}_m(X)\}.\end{aligned}$$

In the case of location estimation, we choose $\Theta_0 = \{\theta_0 \in R^p\colon \|\theta_0 - T(X)\| \le t_0\}$ with t_0 being equal to infinity or some large positive constant. In the case of a confidence set, if T is used to construct a confidence set, say $\{\theta\colon \|\theta - T\| \le b\}$, then we choose $\Theta_0 = \{\theta_0\colon \|\theta_0 - \theta\| \le b\}$. Similarly, Θ_0 can be found for the situations considered in Stromberg and Ruppert (1992), Sakata and White (1995) and Zhang (1996). It is obvious that the variance of $\hat{\varepsilon}_{Rn}(\Theta_0)$ being zero is equivalent to $\hat{\varepsilon}_{Rn}(\Theta_0)$ being universal (that is, independent of the initial sample X) almost surely. So, the asymptotic variance of the sample breakdown point can show the degree of its dependence on the initial sample. The distribution of the sample breakdown point plays an important role in describing the robust behavior of the corresponding statistical procedures. For example, in the case of a confidence set, by the same argument as Zhang (1997), we can show that the value of the distribution fuction of $\hat{\varepsilon}_{Rn}(\Theta_0)$ at m is just the maximum of the 1-*confidence level* over the contamination neighborhood $\mathscr{X}_m$.

In the one-dimensional location case, Huber (1984) and Chao (1986) showed that for the M-estimator with certain unbounded objective function ρ the addition sample breakdown point (with $\Theta_0 = R^1$) is about 1/2 and is independent of both ρ and the sample, whereas, for the M-estimator with a bounded ρ function, Huber (1984) found that the addition sample breakdown point is surprisingly complicated and depends on the shape of ρ, the tuning constant and the sample configuration. Lopuhaä (1992) gave an extreme example to show, for the M-estimator with a particular bounded ρ function, that the sample breakdown point can vary from $1/n$ to 1/2 because of the different sample values. A similar phenomenon happens in the case of nonlinear regression [see, e.g., Stromberg and Ruppert (1992)]. He, Jureková, Koenker and Portnoy (1990) showed that the monotonicity of the location estimator is a sufficient condition for $\hat{\varepsilon}_{Rn}(R^1)$ being universal. Further views on the monotonicity can be found in Basset (1991) and Rousseeuw (1994). On the other hand, for the case that $\Theta_0 \neq R^p$, the sample breakdown point often depends on the initial sample, and its distribution is difficult to compute [see Zhang (1996)] even if the corresponding estimator is monotone. The reason for using $\hat{\varepsilon}_{Rn}(\Theta_0)$ with $\Theta_0 \neq R^P$ is presented in He, Simpson and Portnoy (1990) and Rousseeuw and Croux (1994). The difficulty in analyzing the dependence of the sample breakdown point on the initial sample and in obtaining the distribution of the sample breakdown point can be handled by using an asymptotic theory.

The following two questions are central to asymptotic theory: Which (asymptotic) distributions do the sample breakdown points follow? How large are their asymptotic variances? To answer these questions, some alternative descriptions and variations of the replacement sample breakdown point are introduced in Section 2. The results of Section 3 show that for the location M-estimator with certain bounded ρ function, both the addition, simplified replacement and replacement sample breakdown points are asymptotically normal. Compared with the simplified replacement sample breakdown point, the addition sample breakdown point has a smaller asymptotic variance. Numerical results suggest that the replacement sample breakdown point has the largest asymptotic variance among the three kinds of sample breakdown points. The proofs of these results are given in Section 4.

2. Alternative descriptions of breakdown points. The way the sample breakdown point was defined in the previous section is traditional but not convenient to developing the asymptotic theory for the sample breakdown points. In this section, we give some alternative descriptions of the sample breakdown points which are the bases of the theory in the next section. For convenience, we consider only the case that $p = 1$ and $\Theta_0 = R^1$.

Let $X = (x_1, \dots, x_n)$ be a sample of size n from distribution F defined on a Borel measurable space $(R^1, \mathscr{B})$, and let $T(X)$ be a statistic. Let F_n be the empirical distribution based on X. Denote by $X_k = (x_1, \dots, x_k)$ the first k

components of X. For $Y = (y_1, \dots, y_k)$ and $Z = (z_1, \dots, z_l)$, denote

$$Y \cup Z = (y_1, \dots, y_k, z_1, \dots, z_l).$$

For $v = 1, 2, \dots,$ write

$$\mathscr{D}_v = \left\{ \bigcup_{i=1}^{v} I_i\colon I_i \cap I_j = \varnothing, i \neq j; I_i = (a_i, b_i] \text{ or } [a_i, b_i], a_i, b_i \in R^1, 1 \leq i \leq v \right\}.$$

For $C \in \mathscr{B}$, define

$$\mathscr{X}(C) = \left\{ Y = (y_1, \dots, y_n)\colon y_i = x_i \text{ if } x_i \notin X \cap C; y_j \in R^1 \text{ if } x_j \in X \cap C \right\}.$$

Donoho and Huber's (1983) addition sample breakdown point $\hat{\varepsilon}_A$ and replacement breakdown point $\hat{\varepsilon}_{Rn}$ of T at X are, respectively, of the form

$$\begin{aligned} \hat{\varepsilon}_A &= \hat{\varepsilon}_A(X, T) = \hat{\varepsilon}_A(X, T, R^1) \\ &= \min\left\{ \frac{k}{n+k}\colon \sup_{Y \in \mathscr{R}^k} \left| T(X \cup Y) - T(X) \right| = \infty \right\}, \\ \hat{\varepsilon}_{Rn} &= \hat{\varepsilon}_{Rn}(X, T) = \hat{\varepsilon}_{Rn}(X, T, R^1) \\ &= \min\left\{ \frac{k}{n}\colon \max_{\substack{D \in \mathscr{D}_n \\ \#(D \cap X) = k}} \sup_{Y \in \mathscr{X}(D)} |T(Y) - T(X)| = \infty \right\}, \end{aligned}$$

where $\#(D \cap X)$ is the cardinal number of $D \cap X$.

Note that $\sup_{D \in \mathscr{D}_n} |F_n(D) - F(D)| \not\to 0$ [Pollard (1984), page 22]. This makes $\hat{\varepsilon}_{Rn}$ very hard to compute. Fortunately, in the next section for the commonly used location M-estimator we can asymptotically reduce $\mathscr{D}_n$ to $\mathscr{D}_v$, where $v < n$ is a positive constant independent of n. Furthermore, we show that $\hat{\varepsilon}_{Rn} = \hat{\varepsilon}_{Rv} + o_p(1/\sqrt{n})$, where for $v \leq n$,

$$\begin{aligned} \hat{\varepsilon}_{Rv} &= \hat{\varepsilon}_{Rv}(X, T) = \hat{\varepsilon}_{Rv}(X, T, R^1) \\ &= \min\left\{ \frac{k}{n}\colon \max_{\substack{D \in \mathscr{D}_v \\ \#(D \cap X) = k}} \sup_{Y \in \mathscr{X}(D)} |T(Y) - T(X)| = \infty \right\}. \end{aligned}$$

For the general case, we introduce the simplified replacement sample breakdown point which is defined by

$$\begin{aligned} \hat{\varepsilon}_{SR} &= \hat{\varepsilon}_{SR}(X, T) = \hat{\varepsilon}_{SR}(X, T, R^1) \\ &= \min\left\{ \frac{k}{n}\colon \sup_{Y \in R^k} |T(X_{n-k} \cup Y) - T(X)| = \infty \right\}. \end{aligned}$$

This kind of simplification was in Ylvisaker (1977) and Zhang (1996).

For T and X, set

$$m = m(X,T) = \min\left\{k\colon \sup_{Y\in R^k} |T(X_{n-k}\cup Y)| = \infty\right\},$$

$$m_1 = \min\left\{k\colon \sup_{Y\in R^k} |T(X_{n-m}\cup Y)| = \infty\right\},$$

$$r = \min\left\{k\colon \max_{\substack{D\in\mathscr{D}_n\\ \#(D\cap X)=k}} \sup_{Y\in\mathscr{X}(D)} |T(Y)| = \infty\right\}.$$

Then these sample breakdown points have the relationship

$$\hat{\varepsilon}_{SR}(X,T) \geq \hat{\varepsilon}_A(X_{n-m},T) \geq \hat{\varepsilon}_{SR}(X_{n-m+m_1},T),$$
$$\hat{\varepsilon}_{Rn}(X,T) = \hat{\varepsilon}_{Rr}(X,T) \leq \hat{\varepsilon}_{SR}(X,T).$$

In general, $m \neq m_1$. For example, let $T(X) = \mathrm{med}(X)$ and $n = 2s-1$. Then $m = s$ and $m_1 = s-1$.

To study the asymptotic behavior of these sample breakdown points, we recall the asymptotic breakdown point ε_A^* based on gross-error neighborhood, and we introduce another asymptotic breakdown point ε_R^* as follows.

Let $\mathscr{M}$ denote the set of all one-dimensional distributions and let F be the underlying distribution of X. Suppose that the statistic $T(X)$ is a functional of the empirical distribution, F_n, of X. For notational simplicity, we use the same T for the functional, that is, $T(X) = T(F_n)$. For $F\in\mathscr{M}$, the gross-error neighborhood of F is $C(F,\varepsilon) = \{(1-\varepsilon)F + \varepsilon H\colon H\in\mathscr{M}\}$. Then the asymptotic breakdown point of T based on $C(F,\varepsilon)$ is of the form

$$\varepsilon_A^* = \varepsilon_A^*(F,T) = \inf\left\{\varepsilon\geq 0\colon \sup_{G\in C(F,\varepsilon)} |T(G)-T(F)| = \infty\right\}.$$

For $C\in\mathscr{B}$, define F_{R^1-C} by

$$F_{R^1-C}(A) = F((R^1-C)\cap A)/F(R^1-C) \quad \text{for } A\in\mathscr{B},$$

and define F_{n,R^1-C} similarly. Let

$$C_R(F,\varepsilon) = \big\{(1-F(C))F_{R^1-C} + F(C)H\colon C\in\mathscr{B}, F(C)\leq\varepsilon, H\in\mathscr{M}\big\}.$$

$C_R(F,\varepsilon)$ is increasing in ε and is called the replacement corruption neighborhood of F. The asymptotic breakdown point ε_R^* based on $C_R(F,\varepsilon)$ can be expressed as

$$\varepsilon_R^* = \varepsilon_R^*(F,T) = \inf\left\{\varepsilon\geq 0\colon \sup_{G\in C_R(F,\varepsilon)} |T(G)-T(F)| = \infty\right\}.$$

To find the connection between ε_R^* and $\hat{\varepsilon}_{Rn}$, we introduce a finite-sample version of the neighborhood $C_R(F,\varepsilon)$. For $Z = (z_1,\dots,z_l)\in R^l$, let $H(Z)$

denote the empirical distribution of Z. Define

$$C_R(F_n, k/n) = \{(1 - F_n(C))F_{n, R^1 - C} + F_n(C)H(Z) \colon C \in \mathscr{B}, F_n(C) \le k/n, \\ Z \in R^l, l = nF_n(C)\}$$

for $1 \le k \le n$. For

$$G_n = (1 - F_n(C))F_{n, R^1 - C} + F_n(C)H(Z) \in C_R(F_n, k/n),$$

we denote by $x_{i_1}, \dots, x_{i_l}$ the data points in $X \cap C$. Define $y_j = x_j$ if $x_j \in X \cap (R^1 - C)$, $y_{i_j} = z_j$, $1 \le j \le l$, and $Y = (y_1, \dots, y_n)$. Then $T(Y) = T(G_n)$ and

$$\hat{\varepsilon}_{Rn} = \inf\left\{\frac{k}{n} \colon \sup_{G_n \in C_R(F_n, k/n)} |T(G_n) - T(F_n)| = \infty\right\}.$$

Hence, $\hat{\varepsilon}_{Rn}$ can be viewed as a finite-sample version of ε_R^*.

3. Asymptotic properties of sample breakdown points. Let $X = (x_1, \dots, x_n)$ be an independent and identically distributed sample with common distribution $F(\cdot) = F_0(\cdot - \theta)$ and ρ a real function defined in R^1. The location M-estimator $T(X)$ of θ corresponding to a function ρ is defined as the set of the solutions of the minimization problem

$$\sum_{i=1}^{n} \rho(x_i - T(X)) = \min!. \tag{3.1}$$

The functional $T(\cdot)$ associated with the M-estimator and defined by (3.1) is the set of all solutions to the minimization problem

$$\int (\rho(x - T(F)) - \rho(x)) dF(x) = \min!. \tag{3.2}$$

Note that M-estimators may not be uniquely determined by the corresponding implicit equation. In this case, all the definitions in the previous section are still suitable if $T(X)$ denotes the set of all the solutions and define $|T(X)| = \sup\{|t| \colon t \in T(X)\}$. In the following text we will study the M-estimators induced by ρ functions satisfying, respectively, the following conditions:

(Bρ) $\rho(x)$ attains its minimum -1 at $x = 0$; ρ is nonincreasing for $x < 0$ and nondecreasing for $x > 0$. Furthermore, $\rho(x) \to 0$ as $|x| \to \infty$.

(Uρ) $\rho(x)$ attains its minimum 0 at $x = 0$; $\rho(x)$ is symmetric about 0; $\rho(x)$ is nondecreasing for $x > 0$ and $\lim_{|x| \to \infty} \rho(x) = \infty$; $\psi = \rho'$ is continuous in R^1, and there exists $x_0 \ge 0$ such that ψ is nondecreasing in $(0, x_0]$ and nonincreasing in (x_0, ∞).

(Cρ) $\rho(x)$ is convex, $\psi(x) := \rho'(x)$ exists everywhere and $\psi(-\infty) < 0 < \psi(+\infty)$.

Obviously, any bounded ρ function can be transformed to satisfy $-1 \le \rho \le 0$ by a linear mapping. In this sense, all the commonly used redescending M-estimators of location satisfy the conditions (Bρ). The Huber type location M-estimators satisfy both conditions (Uρ) and (Cρ). Also, under (Cρ), (3.1) and (3.2) are, respectively, equivalent to the problems of solving the equations

$$\frac{1}{n}\sum_{x \in X} \psi(x - T(X)) = 0, \qquad \int \psi(x - T(F))\, dF(x) = 0.$$

3.1. *Breakdown points.* This subsection continues the works of Huber (1981, 1984), He, Jureková, Koenker and Portnoy (1990) and Lopuhaä (1992). For completeness, their results are also included in the following theorems. Throughout this article $\lfloor b \rfloor$ denotes the integer part of b and $\lceil b \rceil$ denotes the smallest integer larger than or equal to b.

THEOREM 3.1. *Suppose ρ satisfies condition* (Uρ). *Then for each sample of size n and $1 \le v \le n$, $\hat{\varepsilon}_A = 1/2$ and $\hat{\varepsilon}_{SR} = \hat{\varepsilon}_{Rv} = \lfloor (n+1)/2 \rfloor / n$.*

THEOREM 3.2. *Suppose ρ satisfies condition* (Bρ). *For a given sample $X = (x_1, \dots, x_n)$, set*

$$A_0 = 0, \quad B^v_{n,0} = 0,$$

$$A_k = \sup_{|t| < \infty} \sum_{i=1}^{k} -\rho(x_i - t), \qquad k = 1, 2, \dots, n;$$

$$B^v_{n,n-k} = \min_{\substack{D \in \mathscr{D}_v \\ \#(D \cap X) = k}} \sup_{|t| < \infty} \sum_{x \in X - X \cap D} -\rho(x - t), \qquad 1 \le k, v \le n,$$

$$m_n = \min\{k\colon k \ge A_{n-k}\}, \qquad r_{nv} = \min\{k\colon k \ge B^v_{n,n-k}\}.$$

Let a_n be an integer satisfying $\lceil A_n \rceil \le a_n \le \lfloor A_n \rfloor + 1$. Then $\hat{\varepsilon}_A = a_n/(n + a_n)$, $m_n/n \le \hat{\varepsilon}_{SR} \le (m_n+1)/n$, $r_{nv}/n \le \hat{\varepsilon}_{Rv} \le (r_{nv}+1)/n$ and $\hat{\varepsilon}_{Rn} = \hat{\varepsilon}_{Rr_{nn}}$ with $r_{nn} \le (n+3)/2$. In addition, if there is a $0 < c < \infty$ such that $\rho(x) = 0$ for $|x| \ge c$, then $a_n = \lceil A_n \rceil$, $\hat{\varepsilon}_{SR} = m_n/n$ and $\hat{\varepsilon}_{Rv} = r_{nv}/n$.

THEOREM 3.3. *For ρ satisfying* (Cρ), *any sample of size n and $1 \le v \le n$,*

$$\hat{\varepsilon}_{SR} = \hat{\varepsilon}_{Rv} = \hat{\varepsilon}_A = \min\left\{\frac{-\psi(-\infty)}{\psi(\infty) - \psi(-\infty)}, \frac{\psi(\infty)}{\psi(\infty) - \psi(-\infty)}\right\},$$

if ψ is bounded, and $\hat{\varepsilon}_A = 1/(n+1)$, $\hat{\varepsilon}_{SR} = \hat{\varepsilon}_{Rv} = 1/n$ if ψ is unbounded.

Theorem 3.4 below will give the asymptotic breakdown points of location M-estimators for the bounded ρ and unbounded ρ in the sense of the gross-error neighborhood and replacement corruption neighborhood, respectively. To state this theorem, we define a function $B(\cdot)$ on $[0, 1)$ by

$$B(\varepsilon) = \inf_{\substack{C \in \mathscr{B}, \\ F(C) \le \varepsilon}} \sup_{|t| < \infty} \int_{R^1 - C} -\rho(x - t)\, dF(x).$$

Denote

$$\varepsilon_0 = \inf\{\varepsilon \ge 0\colon \varepsilon \ge B(\varepsilon)\}.$$

Observe that, for ρ satisfying (Bρ), $B(0) > 0$, $B(\varepsilon) \le 1 - \varepsilon$ and $B(\varepsilon)$ is decreasing. Thus, it is easy to show that $0 \le \varepsilon_0 \le 1/2$ and that $\varepsilon > B(\varepsilon)$ if $\varepsilon > \varepsilon_0$ and $\varepsilon < B(\varepsilon)$ if $\varepsilon < \varepsilon_0$. Therefore, ε_0 is the unique solution of the equation $\varepsilon = B(\varepsilon)$ if $B(\varepsilon)$ is continuous.

THEOREM 3.4. (i) *Under condition* (Bρ), *we have* $\varepsilon_A^*(F, T) = A^*/(1 + A^*)$ *with* $A^* = \sup_{t \in R^1} \int(-\rho(x - t))\, dF(x)$ *and* $\varepsilon_R^*(F, T) = \varepsilon_0$.

(ii) *If* ρ *satisfies* (Uρ) *and* $\int |x|\, dF < \infty$, *then* $\varepsilon_A^*(F, T) = \varepsilon_R^*(F, T) = 1/2$.

(iii) *Assume that* ρ *satisfies* (Cρ). *If* ψ *is unbounded, then* $\varepsilon_A^*(F, T) = \varepsilon_R^*(F, T) = 0$. *If* ψ *is bounded, then*

$$\varepsilon_A^*(F, T) = \varepsilon_R^*(F, T) = \min\left\{ \frac{-\psi(-\infty)}{\psi(+\infty) - \psi(-\infty)}, \frac{\psi(+\infty)}{\psi(+\infty) - \psi(-\infty)} \right\}.$$

The methods used in the proofs of Theorems 3.1–3.4 are similar to Huber (1984), so the details are omitted here.

3.2. *Asymptotic properties.* As seen in the previous subsection, under either (Uρ) or (Cρ), $\hat{\varepsilon}_A$, $\hat{\varepsilon}_{Rv}$, $1 \le v \le n$, and $\hat{\varepsilon}_{SR}$ are nonrandom, and converge to the same asymptotic breakdown point. However, when ρ satisfies (Bρ), the sample breakdown points are random and may have different asymptotic behavior which is studied in this subsection. Throughout the remainder of this paper, $\to_d$ stands for the weak convergence in the sense of Pollard [(1990), page 44].

We begin by introducing the following condition for F and ρ:

(Fρ) ρ is centrosymmetric with a derivative function $\psi(x)$ satisfying

$$\left(\int \rho(x - \tau)\, dF(x)\right)' = -\int \psi(x - \tau)\, dF(x), \qquad \int |\psi(x - \tau)|\, dF(x) < \infty;$$

$\psi(x) \ge 0$ for $x \ge 0$ and $F(\cdot) = F_0(\cdot - \theta)$ has density $f(x) = f_0(x - \theta)$, where $f_0(x)$ is centrosymmetric, strictly decreasing for $x \ge 0$ and continuous at $F_0^{-1}((1 + \varepsilon_R^*)/2)$.

At first, we show consistency for $\hat{\varepsilon}_{SR}$, $\hat{\varepsilon}_A$ and $\hat{\varepsilon}_{Rn}$.

THEOREM 3.5 (Consistency). *If* ρ *satisfies condition* (Bρ), *then for any* $F \in \mathscr{M}$, *both* $\hat{\varepsilon}_{SR}$ *and* $\hat{\varepsilon}_A$ *converge to the same* ε_A^* *almost surely. If* F *and* ρ *satisfy conditions* (Fρ) *and* (Bρ), *then* $\hat{\varepsilon}_{Rn}$ *converges to* ε_R^* *almost surely.*

REMARK 3.1. It follows from Theorem 3.5 that $\varepsilon_R^* \le \varepsilon_A^*$. In addition, numerical results show that for the commonly used bounded ρ functions, $\varepsilon_R^* < \varepsilon_A^*$. Thus, Theorem 3.5 suggests that for the location M-estimator with a bounded ρ, the replacement sample breakdown points may be different from both the addition sample breakdown point and the simplified replacement sample breakdown point in the large sample case.

THEOREM 3.6. *Suppose that* ρ *satisfies condition* (Bρ) *and* $\int(\rho(x-\tau_1)-\rho(x-\tau_2))^2\,dF \to 0$ *as* $\tau_1-\tau_2\to 0$. *Let* $\mathscr{F}=\{-\rho(\cdot-\tau): \tau\in R^1\}$. *Then there exists an* F-*bridge* W *indexed by* $\mathscr{F}$ [*Pollard* (1984), *page* 149] *such that*

$$n^{1/2}(\hat{\varepsilon}_A-\varepsilon_A^*) \longrightarrow_d \frac{1}{(1+A^*)^2}\sup_{\tau\in T(F)} W(-\rho(\cdot-\tau)),$$

$$n^{1/2}(\hat{\varepsilon}_{SR}-\varepsilon_A^*) \longrightarrow_d \frac{1}{(1+A^*)^{3/2}}\sup_{\tau\in T(F)} W(-\rho(\cdot-\tau)).$$

Suppose that F *and* ρ *satisfy conditions* (Fρ) *and* (Bρ), *that* ρ *is continuous and that there exists an interval* (a,b) *such that* $0<a<b<\infty$, $(a,b)\cap(q^*,\infty)\neq\varnothing$ *and* $\psi(x)>0$ *for* $x\in(a,b)$, *where* $q^*=F_0^{-1}((1+\varepsilon_R^*)/2)$. *Then*

$$n^{1/2}(\hat{\varepsilon}_{Rn}-\varepsilon_R^*) \longrightarrow_d N(0,V_R),$$

where $N(0,V_R)$ *is a normal distribution with zero mean and variance*

$$V_R=\frac{1}{(1-\rho(q^*))^2}\left\{2\int_{q^*}^{\infty}\rho^2(x)\,dF_0(x)+\varepsilon_R^*\rho^2(q^*)\right\}-\varepsilon_R^{*2}.$$

REMARK 3.2. (1) If ρ is bounded and continuous, then, as $\tau_1-\tau_2\to 0$,

$$\int(\rho(x-\tau_1)-\rho(x-\tau_2))^2\,dF(x)\to 0.$$

(2) Suppose F and ρ satisfy condition (Fρ). Then $T(F)=\{\theta\}$ is unique and the limit distributions of $\hat{\varepsilon}_A$ and $\hat{\varepsilon}_{SR}$ are normal with asymptotic variances

$$V_A=\frac{2\int_0^\infty\rho^2(x)\,dF_0(x)-A^{*2}}{(1+A^*)^4},\qquad V_S=\frac{2\int_0^\infty\rho^2(x)\,dF_0(x)-A^{*2}}{(1+A^*)^3},$$

respectively.

Obviously, for the ρ functions which have appeared in the literature, such as the biweight, Hampel, Andrews and tanh functions [see Hampel, Ronchetti, Rousseeuw and Stahel (1986)], and for the commonly used normal, Laplace and t distributions, the conditions in Theorem 3.6 hold.

REMARK 3.3. Theorem 3.6 indicates that $\hat{\varepsilon}_A$ has a smaller asymptotic variance than $\hat{\varepsilon}_{SR}$. The reason is that the behavior of both $\hat{\varepsilon}_A$ and $\hat{\varepsilon}_{SR}$ is determined by their corrupted samples, respectively. In the case of $\hat{\varepsilon}_A$, the sample size after corruption is $n(1+\hat{\varepsilon}_A/(1-\hat{\varepsilon}_A))$ [approximately $n(1+A^*)$]. In contrast, in the case of $\hat{\varepsilon}_{SR}$ the sample size after corruption is still n.

TABLE 1
The limits and asymptotic standard variances of $\hat{\varepsilon}_A$, $\hat{\varepsilon}_{SR}$ and $\hat{\varepsilon}_{Rn}$

			Biweight				
c	3	4	5	6	7	8	9
ε_A^*	0.4310	0.4577	0.4719	0.4801	0.4852	0.4886	0.4909
ε_R^*	0.3838	0.4260	0.4498	0.4640	0.4730	0.4791	0.4833
$\sqrt{V_A}$	0.0840	0.0546	0.0375	0.0271	0.0203	0.0158	0.0126
$\sqrt{V_S}$	0.1113	0.0741	0.0516	0.0375	0.0283	0.0221	0.0177
$\sqrt{V_R}$	0.1262	0.0894	0.0642	0.0475	0.0363	0.0285	0.0229
			Hampel				
h_1	1	1.31	1.5	2	2.5	3	3
h_2, h_3	2, 3	2.039, 4	2.5, 5	3, 6	3.5, 7	4, 8	4.5, 9
ε_A^*	0.4415	0.4599	0.4721	0.4817	0.4872	0.4906	0.4919
ε_R^*	0.3996	0.4293	0.4499	0.4668	0.4766	0.4827	0.4852
$\sqrt{V_A}$	0.0740	0.0533	0.0380	0.0259	0.0184	0.0135	0.0116
$\sqrt{V_S}$	0.0990	0.0725	0.0524	0.0360	0.0257	0.0190	0.0162
$\sqrt{V_R}$	0.1153	0.0883	0.0655	0.0461	0.0333	0.0248	0.0212
			Andrews				
$a\pi$	π	4	5	6	7	8	9
ε_A^*	0.4454	0.4645	0.4766	0.4835	0.4877	0.4905	0.4925
ε_R^*	0.4062	0.4372	0.4579	0.4700	0.4776	0.4827	0.4862
$\sqrt{V_A}$	0.0688	0.0467	0.0316	0.0226	0.0169	0.0131	0.0105
$\sqrt{V_S}$	0.0924	0.0638	0.0437	0.0315	0.0236	0.0184	0.0147
$\sqrt{V_R}$	0.1083	0.0781	0.0549	0.0402	0.0305	0.0238	0.0191
			tanh				
a	2.032	1.22	1.488	0.820	1.194	1.5	1.5
b, c	2.451, 3	1.289, 4	1.55, 4	0.865, 5	1.217, 5	1.8, 6	2, 7
ε_A^*	0.4221	0.4609	0.4646	0.4674	0.4770	0.4957	0.4999
ε_R^*	0.3690	0.4310	0.4372	0.4428	0.4588	0.4920	0.4981
$\sqrt{V_A}$	0.1173	0.0506	0.0481	0.0375	0.0291	0.0060	0.0015
$\sqrt{V_S}$	0.1543	0.0689	0.0657	0.0513	0.0402	0.0084	0.0021
$\sqrt{V_R}$	0.1758	0.0836	0.0810	0.0613	0.0501	0.0109	0.0028

REMARK 3.4. Consider the biweight, Hampel, Andrews and tanh ρ functions with tuning constants c, (h_1, h_2, h_3), a and (a, b, c), respectively. The definitions of the derivatives of these functions can be found in Huber [(1981), pages 100–102] and Hampel, Ronchetti, Rousseeuw and Stahel [(1986), page 151]. Assume F_0 is a normal distribution. Then the asymptotic breakdown points ε_A^* and ε_R^* and the asymptotic standard variances $\sqrt{V_A}$, $\sqrt{V_S}$ and $\sqrt{V_R}$ of $\hat{\varepsilon}_A$, $\hat{\varepsilon}_{SR}$ and $\hat{\varepsilon}_{Rn}$, for different tuning constants, are given in Table 1.

The numerical results suggest that among the three kinds of sample breakdown points, $\hat{\varepsilon}_{Rn}$ and $\hat{\varepsilon}_A$ have the largest and smallest asymptotic variances, respectively. The asymptotic variances of these sample breakdown points are decreasing functions of c. This implies that in the large sample case the degree of the dependence of these sample breakdown points on the initial sample is decreasing in the tuning constants.

REMARK 3.5. In principle, the asymptotic approach just developed can be used to handle the sample breakdown points of the expressions similar to those in Theorem 3.2. In light of this point, we can obtain the asymptotic distributions of the replacement sample breakdown points for some tests and the corresponding confidence intervals in Zhang (1996). The details are omitted here.

4. Proofs of the main results. In this section, we will prove Theorems 3.5 and 3.6. Lemmas 4.1–4.7 below are auxiliary lemmas. The proofs of Lemmas 4.3 and 4.7 are given in the Appendix. The proofs of Remarks 3.1 and 3.2 and Lemmas 4.1, 4.2 and 4.4–4.6 are omitted.

LEMMA 4.1. *Suppose ρ satisfies* (Bρ) *and F is continuous. Then, for $\varepsilon \geq 0, \delta \geq 0, \varepsilon + \delta \leq 1$,*

$$0 \leq B(\varepsilon) - B(\varepsilon + \delta) \leq \delta,$$

and for $0 \leq \varepsilon \leq 1$,

$$B(\varepsilon) = \lim_{v \to \infty} \inf_{\substack{C: F(C) \leq \varepsilon \\ C \in \mathscr{D}_v}} \sup_{|t| < \infty} \int_{R^1 - C} -\rho(x - t)\, dF(x).$$

LEMMA 4.2. *Define*

$$\mathscr{F}_v = \{ I_{R^1 - D}\rho(\cdot - t) \colon D \in \mathscr{D}_v, t \in R^1 \},$$

where I_{R^1-D} is the indicator function. Then for ρ satisfying (Bρ) *and some nonrandom integers $\{v_n\}$ with $v_n/n \to 0$,*

$$\sup_{f \in \mathscr{F}_{v_n}} \left| \int f\, dF_n - \int f\, dF \right| \to 0$$

almost surely.

LEMMA 4.3. *Assume that F is a continuous distribution and that ρ satisfies* (Bρ). *Then, for a sequence of nonrandom integers $\{v_n\}$ satisfying $v_n/n \to 0$ and $v_n \to \infty$, $\hat{\varepsilon}_{Rv_n} \to \varepsilon_R^*$ almost surely.*

For $t \in R^1$, $0 \leq \varepsilon \leq 1$ and F_0 in (Fρ), *set*

$$\begin{aligned} \Lambda(t, \varepsilon) = {} & \int_{-\infty}^{-F_0^{-1}((1+\varepsilon)/2)} -\rho(x - t) f_0(x)\, dx \\ & + \int_{F_0^{-1}((1+\varepsilon)/2)}^{\infty} -\rho(x - t) f_0(x)\, dx. \end{aligned}$$

LEMMA 4.4. *If F and ρ satisfy* (Fρ) *and* (Bρ), *then*

$$B(\varepsilon) = 2\int_{F_0^{-1}((1+\varepsilon)/2)}^{\infty} -\rho(x)\,dF_0(x).$$

Furthermore, for ε fixed, if there exists an interval (a, b) such that $(a, b) \cap (F_0^{-1}((1+\varepsilon)/2), \infty) \neq \varnothing$ and $\psi(x) > 0$ for $x \in (a, b)$, then $\Lambda(t, \varepsilon)$ attains the maximum only at $t = 0$.

LEMMA 4.5. *For ρ centrosymmetric and satisfying* (Bρ), *$\theta \in R^1$ and $0 \le k \le n$,*

$$\begin{aligned}\inf_{\substack{\#(X_\theta \cap I)=k\\ I=(-a,a] \text{ or } [-a,a]}} \sum_{x \in X_\theta - I} -\rho(x) &\le \inf_{\substack{D \in \mathscr{D}_n\\ \#(D\cap X)=k}} \sup_{|t|<\infty} \sum_{x \in X-D} -\rho(x-t)\\ &\le \inf_{\substack{\#(X_\theta \cap I)=k\\ I=(-a,a] \text{ or } [-a,a]}} \sup_{|t|<\infty} \sum_{x\in X_\theta - I} -\rho(x-t),\end{aligned}$$

where $X_\theta = (x_1 - \theta, \dots, x_n - \theta)$.

The combination of Lemmas 4.4 and 4.5 leads to the reduction of $\mathscr{D}_n$ to $\mathscr{D}_1$.

PROOF OF THEOREM 3.5. Without loss of generality, we assume the true parameter $\theta = 0$. Choose a sequence of integers $\{v_n\}$ such that $v_n/n \to 0$ and $v_n \to \infty$. It follows from the left-hand side of the inequality in Lemma 4.5 that

$$\hat{\varepsilon}_{Rn} \le \hat{\varepsilon}_{Rv_n} \quad \text{for } v_n \ge 1.$$

This together with Lemma 4.3 implies that as $n \to \infty$,

$$\lim \hat{\varepsilon}_{Rn} \le \lim \hat{\varepsilon}_{Rv_n} = \varepsilon_R^* \quad \text{almost surely.}$$

On the other hand, if we let

$$S_{nk} = \inf_{\substack{\#(X\cap I)=k\\ I=(-a,a] \text{ or } [-a,a]}} \sum_{x\in X-I} -\rho(x),$$

$$\hat{\varepsilon}_{0n} = \min\left\{\frac{k}{n}\colon k \ge S_{nk}\right\},$$

then it follows from the right-hand side of the inequality in Lemma 4.5 that $\hat{\varepsilon}_{Rn} \ge \hat{\varepsilon}_{0n}$. By the same argument as Lemma 4.3, we prove that $\hat{\varepsilon}_{0n}$ converges to ε_R^* almost surely. We obtain the desired result for $\hat{\varepsilon}_{Rn}$.

It remains to show the results for $\hat{\varepsilon}_{SR}$ and $\hat{\varepsilon}_A$. We take $\hat{\varepsilon}_{SR}$ as our example because both terms are similar. At first, we show that

$$\left|\frac{A_n}{n} - A^*\right| \to 0 \quad \text{almost surely} \tag{4.1}$$

as $n \to \infty$. Since

$$\left|\frac{A_n}{n} - A^*\right| \le \sup_t \left|\frac{1}{n}\sum_{i=1}^{n} -\rho(x_i - t) - \int\big(-\rho(x-t)\big)\,dF(x)\right|, \tag{4.2}$$

it suffices to show that the right-hand side of (4.2) converges to zero almost surely. This follows from Theorem 24 and Lemma 25 in Pollard [(1984), pages 25 and 27] and the fact that the graphs of functions in $\{-\rho(\cdot - t)\colon t \in R^1\}$ form a polynomial class of sets [Pollard (1984), page 17]. Set

$$f(n,k) = \frac{n-k}{n(1+A^*)}\left(\frac{A_{n-k}}{n-k} - A^*\right).$$

By Theorem 3.2, (4.1) and the definition of $\hat{\varepsilon}_{SR}$ we have

$$|\hat{\varepsilon}_{SR} - \varepsilon^*| \le \sup_{0 \le k \le n} f(n,k) + \frac{1}{n} \to 0 \quad \text{almost surely}$$

as $n \to \infty$. The proof is complete. □

LEMMA 4.6. *Suppose function ρ satisfies the condition* (Bρ). *Let*

$$D_n = \left\{t \in R^1\colon \sum_{i=1}^{n} -\rho(x_i - t) = \sup_{\tau \in \mathscr{R}^1} \sum_{i=1}^{n} -\rho(x_i - \tau)\right\}, \qquad n = 1, 2, \ldots. \tag{4.3}$$

Then $T(F)$ is a bounded closed set and

$$\sup_{t \in D_n} d(t, T(F)) := \sup_{t \in D_n} \inf_{\tau \in T(F)} |t - \tau| \to 0 \quad \textit{almost surely.}$$

LEMMA 4.7. *Suppose function ρ satisfies the conditions in the first part of Theorem* 3.6. *Denote by $C(\mathscr{F}, F)$ the set of all uniformly continuous functionals with respect to seminorm $(\int g^2\,dF)^{1/2}$ defined on $\mathscr{F}$. Let τ_n be a sequence of positive integer-valued random variables with τ_n/n converging to a positive constant τ_0 in probability. Define*

$$W_n = \left\{\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(-\rho(x_i - t) + \int \rho(x-t)\,dF(x)\right)\colon t \in R^1\right\}.$$

Then $C(\mathscr{F}, F)$ is separable and completely regular, and there exists an F bridge, say $\{W(g)\colon g \in \mathscr{F}\}$, whose paths belong to $C(\mathscr{F}, F)$ almost surely and such that $W_n \to_d \{W(g)\colon g \in \mathscr{F}\}$ and $W_{\tau_n} \to_d \{W(g)\colon g \in \mathscr{F}\}$.

PROOF OF THEOREM 3.6. We begin with the proof of the assertion for $\hat{\varepsilon}_{SR}$. The key point is applying the representation theorem of random elements [Pollard (1990), page 45]. First, it follows from Theorem 3.4 and condition (Bρ) that

$$|A_{n-m+1} - m| \le 1,$$

$$|\hat{\varepsilon}_{SR} - \varepsilon_A^*| \le \left|\frac{m}{n} - \frac{A^*}{1+A^*}\right| + 0\left(\frac{1}{n}\right) \to 0 \quad \text{almost surely.}$$

Hence, as $n \to \infty$,

$$n^{1/2}\left(\frac{A_{n-m+1}}{n} - \varepsilon_A^*\right)$$
$$= -n^{1/2}\left(\frac{A_{n-m+1}}{n} - \varepsilon_A^*\right)A^*(1+o(1))$$
$$+ \frac{1}{(1+A^*)^{1/2}}(n-m+1)^{1/2}\left(\frac{A_{n-m+1}}{n-m+1} - A^*\right)(1+o(1)) + o(1),$$

which results in

$$n^{1/2}\left(\frac{A_{n-m+1}}{n} - \varepsilon_A^*\right) = \frac{(n-m+1)^{1/2}}{(1+A^*)^{3/2}}\left(\frac{A_{n-m+1}}{n-m+1} - A^*\right)(1+o(1)) + o(1).$$

Thus, if

$$(4.4) \qquad (n-m+1)^{1/2}\left(\frac{A_{n-m+1}}{n-m+1} - A^*\right) \longrightarrow_d \sup_{\tau \in T(F)} W(-\rho(\cdot - \tau)),$$

then, by Theorem 3.4 and Lemma 3.1 in Zhang and Li [(1993), page 329] as well as the representation theorem of random elements just mentioned, Theorem 3.6 is proved. To prove (4.4), we assume, without loss of generality, that there exists an interval $[-b, b]$ which contains D_n defined by (4.3) because of Lemma 4.6. Now from Lemma 4.7 we deduce (4.4).

The proof of the assertion for $\hat{\varepsilon}_A$ is similar.

The proof of the assertion for $\hat{\varepsilon}_{Rn}$ is based on Lemma 4.5. The key step is to show that the statistics in the left- and right-hand sides of the inequality in Lemma 4.5 converge weakly to the same distribution by using the functional limit theorem of noncentral empirical processes [see Lemma 3.1 of Zhang and Li (1993)]. As before, for simplicity, we assume $\theta = 0$.

At first, we introduce the two distributions, G_n and G, by

$$G_n(x) = F_n((-x, x)), \qquad G(x) = F((-x, x)) \quad \text{for } 0 \le x < \infty;$$
$$G_n^{-1}(u) = \inf\{x \ge 0\colon G_n(x) \ge u\}, \qquad G^{-1}(u) = \inf\{x \ge 0\colon G(x) \ge u\},$$

for $0 \le u \le 1$. For $a \ge 0$, set $I_a = (-a, a]$ or $I_a = [-a, a]$. For $0 \le k \le n$, define $a_{nk} = G_n^{-1}(k/n)$, $I_{nk} = I_{a_{nk}} = (-a_{nk}, a_{nk}]$ or $I_{nk} = I_{a_{nk}} = [-a_{nk}, a_{nk}]$ such that $F_n(I_{nk}) = k/n$. For notational simplicity, in the following discussion we use the symbol $o_p(1)$ to denote any random variable which converges weakly to zero.

STEP 1. Let $q^* = F^{-1}((1+\varepsilon_R^*)/2)$,

$$\mathcal{T} = \left\{\tau\colon \int I(x \notin I_{q^*})(-\rho(\cdot - \tau))\,dF = \sup_{|t|<\infty} \int I(x \notin I_{q^*})(-\rho(\cdot - t))\,dF\right\},$$

$$\mathcal{T}_n = \left\{\tau\colon \int I(x \notin I_{nk_1})(-\rho(\cdot - \tau))\,dF_n = \sup_{|t|<\infty} \int I(x \notin I_{nk_1})(-\rho(\cdot - t))\,dF_n\right\},$$

where $I(\cdot)$ is the indicator of a set, $k_1 = n\hat{\beta}_n$ and $\hat{\beta}_n$ is any function of X which converges to ε_R^* almost surely. Then, by Lemma 4.4 and the assumption,

$\mathscr{T} = \{0\}$. Analogous to Lemma 4.6, we can prove that

$$\sup_{t \in \mathscr{T}_n} d(t, \mathscr{T}) \to 0 \quad \text{almost surely.}$$

Thus, without loss of generality, t is restricted to $[-\tau_0, \tau_0]$ for a sufficiently large value τ_0.

STEP 2. To obtain the limit process of

$$W_{1n}(t) = n^{1/2} \int I(x \notin I_{nk_1})\big(-\rho(x-t)\big)\, dF_n(x) - c_n(q^*, t), \qquad |t| \le \tau_0, \tag{4.5}$$

where

$$c_n(a, t) = n^{1/2} \int I(x \notin I_a)(-\rho(x-t))\, dF(x),$$

$k_1 = n\hat{\beta}_n$ and $\hat{\beta}_n$ is any function of X which converges to ε_R^* almost surely.

Consider the empirical process

$$W_{0n}(a, t) = n^{1/2} \int I(x \notin I_a)\big(-\rho(x-t)\big)\big(dF_n(x) - dF(x)\big),$$
$$|t| \le \tau_0,\ |a - q^*| \le \delta,$$

where $I_a = [-a, a]$ or $I_a = (-a, a]$ and $\delta > 0$ is a constant. By Pollard's functional central limit theorem [Pollard (1990), page 53] and the condition (Bρ), there exists a Gaussian random element, say $W_0(a, t)$, $|t| \le \tau_0$, $|a - q^*| \le \delta$, such that

$$W_{0n} \longrightarrow_d W_0$$

and almost surely the sample paths of W_0 are continuous in (a, t). Note that, by Theorem 3.5 and that $\sup_{0 \le u \le 1} |G_n^{-1}(u) - G^{-1}(u)| \to 0$, we have that $a_{nk_1} \to q^*$ almost surely. Therefore,

$$\big\{W_{0n}(a_{nk_1}, t)\colon |t| \le \tau_0\big\} = \big\{W_{0n}(q^*, t) + o_p(1)\colon |t| \le \tau_0\big\}. \tag{4.6}$$

On the other hand, by Theorem 3.5, $\hat{\varepsilon}_{Rn} \to \varepsilon_R^*$ almost surely. Then, uniformly for $|t| \le \tau_0$,

$$\begin{aligned}
c_n(a_{nk_1}, t) &- c_n(q^*, t) \\
&= n^{1/2}\bigg(\int_{-q^*}^{-a_{nk_1}} + \int_{a_{nk_1}}^{q^*}\bigg)\big(-\rho(x-t)\big)\, dF(x) \\
&= -n^{1/2} f(q^*)\big(\rho(q^*+t) + \rho(q^*-t)\big)\big[G^{-1}(\varepsilon_R^*) - G_n^{-1}(\hat{\beta}_n)\big]\big(1 + o_p(1)\big) \\
&= -n^{1/2} f(q^*)\big(\rho(q^*+t) + \rho(q^*-t)\big)\big[G^{-1}(\varepsilon_R^*) - G^{-1}(\hat{\beta}_n)\big] \\
&\quad - n^{1/2} f(q^*)\big(\rho(q^*+t) + \rho(q^*-t)\big)\big[G^{-1}(\hat{\beta}_n) - G_n^{-1}(\hat{\beta}_n)\big] \\
&\quad \times \big(1 + o_p(1)\big) + o_p(1)
\end{aligned} \tag{4.7}$$

$$= \frac{(\rho(q^*+t)+\rho(q^*-t))}{2} \times \left[n^{1/2}(G_n(q^*)-\varepsilon_R^*)+n^{1/2}(\hat{\beta}_n-\varepsilon_R^*)(1+o_p(1))\right]+o_p(1),$$

Substituting (4.6) and (4.7) into (4.5), we obtain that

$$\begin{aligned} W_{1n}(t) &= W_{0n}(q^*,t)+\frac{(\rho(q^*+t)+\rho(q^*-t))}{2} \\ &\quad \times \left[n^{1/2}(G_n(q^*)-\varepsilon_R^*)+n^{1/2}(\hat{\beta}_n-\varepsilon_R^*)(1+o_p(1))\right]+o_p(1) \\ &= n^{1/2}\int\left[I(x\notin I_{q^*})(-\rho(x-t))+\frac{(\rho(q^*+t)+\rho(q^*-t))}{2}I(x\in I_{q^*})\right] \\ &\quad \times (dF_n(x)-dF(x)) \\ &\quad + o_p(1)+n^{1/2}(\hat{\beta}_n-\varepsilon_R^*)(1+o_p(1))\frac{(\rho(q^*+t)+\rho(q^*-t))}{2} \end{aligned} \tag{4.8}$$

uniformly for $|t| \le \tau_0$.

STEP 3. Define

$$\begin{aligned} S_{nk} &= \inf_{\substack{\#(X\cap I)=k,\\ I=(-a,a]\text{ or }[-a,a]}} \sum_{x\in X-I} -\rho(x), \\ \hat{\varepsilon}_{0n} &= \min\left\{\frac{k}{n} : k \ge S_{nk}\right\}; \\ U_{nk} &= \inf_{\substack{\#(X\cap I)=k,\\ I=(-a,a]\text{ or }[-a,a]}} \sup_{|t|<\infty} \sum_{x\in X-I} -\rho(x-t), \\ \hat{\varepsilon}_{1n} &= \min\left\{\frac{k}{n} : k \ge U_{nk}\right\}. \end{aligned}$$

Letting $\hat{\beta}_n = \hat{\varepsilon}_{0n}$ and $\hat{\varepsilon}_{1n}$ in (4.8), respectively, and by the definitions of $\hat{\varepsilon}_{0n}$ and $\hat{\varepsilon}_{1n}$, we have

$$\begin{aligned} n^{1/2}(\hat{\varepsilon}_{1n}-\varepsilon_R^*) &= n^{1/2}\sup_{|t|\le\tau_0}\Bigg[\int\bigg[I(x\notin I_{q^*})\big(-\rho(x-t)\big) \\ &\qquad + \frac{\big(\rho(q^*+t)+\rho(q^*-t)\big)}{2}I\big(x\in I_{q^*}\big)\bigg] \\ &\qquad \times \big(dF_n(x)-dF(x)\big)+o_p(1) \\ &\qquad + n^{1/2}(\hat{\varepsilon}_{1n}-\varepsilon_R^*)\big(1+o_p(1)\big) \\ &\qquad \times \frac{\big(\rho(q^*+t)+\rho(q^*-t)\big)}{2}+c_n(q^*,t)\Bigg] \\ &\quad - \sup_{|t|\le\tau_0} c_n(q^*,t), \end{aligned} \tag{4.9}$$

$$n^{1/2}(\hat{\varepsilon}_{0n} - \varepsilon_R^*) = n^{1/2} \int \big[I(x \notin I_{q^*})\big(-\rho(x)\big) + \rho(q^*) I(x \in I_{q^*}) \big] \times \big(dF_n(x) - dF(x)\big) + o_p(1) + n^{1/2}(\hat{\varepsilon}_{0n} - \varepsilon_R^*)(1 + o_p(1))\rho(q^*).$$

Let

$$\tilde{W}_n(t) = n^{1/2} \int \bigg[I(x \notin I_{q^*})\big(-\rho(x-t)\big) + \frac{(\rho(q^*+t) + \rho(q^*-t))}{2} I(x \in I_{q^*}) \bigg] \times \big(dF_n(x) - dF(x)\big) + o_p(1)$$

for $|t| \le \tau_0$. Let $\tilde{W}$ denote the limit of $\tilde{W}_n$. Then, using the representation theorem, we have a sequence of random elements, say $\{W_n^*(t)\colon |t| \le \tau_0\}$, α_n, $n \ge 1$, and $\{W^*(t)\colon |t| \le \tau_0\}$, such that $W_n^* \to W^*$, $\alpha_n \to 0$ almost surely and W_n^*, W^* and α_n are the representations of $\tilde{W}_n$, $\tilde{W}$ and $o_p(1)$, respectively.

Let β_{0n} and β_{1n} be, respectively, the solutions of the equations

$$\beta_{0n} = W_n^*(0) + \rho(q^*)\beta_{0n}(1 + \alpha_n),$$

$$\beta_{1n} = \sup_{|t| \le \tau_0} \bigg[W_n^*(t) + \beta_{1n}(1+\alpha_n) \frac{(\rho(q^*+t) + \rho(q^*-t))}{2} + c_n(q^*, t) \bigg] - \sup_{|t| \le \tau_0} c_n(q^*, t).$$

By using the above equations, the inequality $-1 \le \rho(t) \le 0$, the fact that $\mathscr{T} = \{0\}$ proved in Step 1 and Lemma 3.1 in Zhang and Li (1993), we can show that (i) $\{\beta_{0n}\}$ and $\{\beta_{1n}\}$ are bounded sequences and (ii) every convergent subsequence of $\{\beta_{0n}\}$ and $\{\beta_{1n}\}$ converges to the solution of the equation

$$\beta_0 = W^*(0) + \rho(q^*)\beta_0. \tag{4.10}$$

This implies that both β_{0n} and β_{1n} converge to the solution of the equation (4.10). Thus, from (4.9), (4.10) and the representation theorem of random elements, we conclude that $n^{1/2}(\hat{\varepsilon}_{0n} - \varepsilon_R^*)$ and $n^{1/2}(\hat{\varepsilon}_{1n} - \varepsilon_R^*)$ converge weakly to the same limit $\tilde{W}(0)/(1 - \rho(q^*))$.

Note that, by Lemma 4.5,

$$n^{1/2}\big(\hat{\varepsilon}_{0n} - \varepsilon_R^*\big) \le n^{1/2}\big(\hat{\varepsilon}_{Rn} - \varepsilon_R^*\big) \le n^{1/2}\big(\hat{\varepsilon}_{1n} - \varepsilon_R^*\big).$$

Consequently,

$$n^{1/2}\big(\hat{\varepsilon}_{Rn} - \varepsilon_R^*\big) \longrightarrow_d \tilde{W}(0)/\big(1 - \rho(q^*)\big),$$

which follows a normal distribution with zero mean and variance

$$V_R = \frac{1}{\big(1 - \rho(q^*)\big)^2} \bigg\{ 2 \int_{q^*}^{\infty} \rho^2(x)\, dF(x) + \varepsilon_R^* \rho^2(q^*) \bigg\} - {\varepsilon_R^*}^2.$$

This completes the proof. □

APPENDIX

PROOF OF LEMMA 4.3. Let

$$\Delta_{n1} = \sup_{C \in \mathscr{D}_{v_n}} \left| F_n(C) - F(C) \right|,$$

$$\Delta_{n2} = \sup_{\substack{|t| < \infty \\ C \in \mathscr{D}_{v_n}}} \left| \int_{R^1 - C} -\rho(x-t)\, dF_n(x) - \int_{R^1 - C} -\rho(x-t)\, dF(x) \right|.$$

Then

$$F(C) - \Delta_{n1} \le F_n(C) \le F(C) + \Delta_{n1}.$$

By Lemma 4.2, there exists a subset Ω_1 with $P_r(\Omega_1) = 1$ such that

$$\Delta_{n1} \to 0 \quad \text{and} \quad \Delta_{n2} \to 0 \quad \text{for } w \in \Omega_1. \tag{A.1}$$

For notational simplicity, we still use $\hat{\varepsilon}_{Rv_n}$ to denote the value of $\hat{\varepsilon}_{Rv_n}$ at each fixed $w \in \Omega_1$ below. By the condition (Bρ), $0 \le r_{nv_n}/n \le (n+1)/2n$. Thus, to complete the proof, it remains to show that every convergent subsequence of $\{\hat{\varepsilon}_{Rv_n}\}$ converges to ε^*_R. Without confusion, we use the same symbol, $\hat{\varepsilon}_{Rv_n}$, to represent a convergent subsequence of $\hat{\varepsilon}_{Rv_n}$ with a limit value ε^*. In the following discussion, we show that $\varepsilon^* = \varepsilon^*_R$.

From (A.1) and the assumption just made it follows that for each $\delta > 0$, there exists an n_0 such that for $n \ge n_0$,

$$\hat{\varepsilon}_{Rv_n} < \inf\left\{ \sup_t \int_{R^1 - C} -\rho(x-t)\, dF(x) \colon F(C) \le \varepsilon^* - \frac{3\delta}{4}, C \in \mathscr{D}_{v_n} \right\} + \frac{\delta}{2},$$

$$\hat{\varepsilon}_{Rv_n} \ge \inf\left\{ \sup_t \int_{R^1 - C} -\rho(x-t)\, dF(x) \colon F(C) \le \varepsilon^* - \frac{3\delta}{4}, C \in \mathscr{D}_{v_n} \right\} - \frac{\delta}{2}.$$

Then using Lemma 4.1 and letting $n \to \infty$ we have

$$\varepsilon^* \le B\left(\varepsilon^* - \frac{\delta}{4} \right) + \frac{\delta}{2}, \qquad \varepsilon^* \ge B\left(\frac{\delta}{2} + \varepsilon^* \right) - \frac{\delta}{2}.$$

Apply Lemma 4.1 again and let $\delta \to 0^+$. Then the desired result follows. □

PROOF OF LEMMA 4.7. We give only an outline of this proof because it is a special case of Zhang (1993). Since ρ satisfies the condition (Bρ) and $\int(\rho(x - \tau_1) - \rho(x - \tau_2))^2\, dF(x) \to 0$ as $\tau_1 - \tau_2 \to 0$, it is easy to show $C(\mathscr{F}, F)$ is separable and completely regular [Pollard (1984), page 67]. We construct a new empirical process

$$\left\{ \frac{1}{n^{1/2}} \sum_{i=1}^{n} \left(-\rho(x_i - t) + \int \rho(x_i - t)\, dF(x) \right) I_{[\tau \ge i/n]} \colon t \in \mathscr{R}^1, 0 < \tau \le 1 \right\}.$$

Set

$$f_{ni}(x_i,t,\tau) = \frac{-1}{n^{1/2}}\rho(x_i-t)I_{[\tau\ge i/n]},$$

$$\vec{F}_n := \left(\sup_t \frac{|\rho(x_i-t)|}{n^{1/2}}\right)_{1\le i\le n},$$

$$\mathscr{F}_n = \{\vec{f}_n(t) = (f_{ni}(x_i,t,\tau))_{1\le i\le n}\colon t\in[-b,b], 0\le\tau\le 1\},$$

$$Y_n(t,\tau) = \sum_{i=1}^{n}(f_{ni}(x_i,t,\tau) - Ef_{ni}(x_1,t,\tau)).$$

From Pollard's functional central limit theorem [Pollard (1990), page 53], it can be shown that there exists a Gaussian random element $\{Y(t,\tau)\colon t\in[-b,b], 0\le\tau\le 1\}$ to which $\{Y_n(t,\tau)\colon t\in[-b,b], 0\le\tau\le 1\}$ converges weakly. Using the results and Example 9.2 in Pollard [(1990), page 44], we have that $(Y_n, \tau_n/n)$ converges weakly to (Y, τ_0) in the sense of Pollard [(1990), page 44]. Combining this with Example 9.5 of Pollard (1990) and the usual truncation trick for τ_n, we can show that $W_{\tau_n} = (n/\tau_n)^{1/2}Y_n(\cdot, \tau_n/n)$ converges weakly to $\tau_0^{-1/2}Y(\cdot,\tau_0)$ in the sense just mentioned. Now the desired result follows from the fact that the processes $\tau_0^{-1/2}Y(\cdot,\tau_0)$ and $Y(\cdot,1)$ have the same distribution. $\square$

Acknowlegments We are grateful to the referee, the Associate Editor and the Editor, L. D. Brown, for their very valuable and detailed comments that helped to improve the presentation. Part of this manuscript was prepared when the first author was a postdoctoral fellow at the Institute of Applied Mathematics, Academia Sinica, Beijing, and the Department of Statistics, Chinese University of Hong Kong.

REFERENCES

BASSETT, G. W., JR. (1991). Equivariant, monotonic, 50% breakdown estimators. *Amer. Statist.* **45** 135–137.

CHAO, M. (1986). On M and P estimators that have breakdown point equal to $\frac{1}{2}$. *Statist. Probab. Lett.* **4** 127–131.

COAKLEY, C. W. and HETTMANSPERGER, T. P. (1992). Breakdown bounds and expected test resistance. *J. Nonparametr. Statist.* **1** 267–276.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

HE, X., SIMPSON, D. G. and PORTNOY, S. L. (1990). Breakdown robustness of tests. *J. Amer. Statist. Assoc.* **85** 446–452.

HE, X., JURECKOVÁ, J., KOENKER, R. and PORTNOY, S. L. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* **58** 1195–1214.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

HUBER, P. J. (1984). Finite sample breakdown point of M- and P-estimators. *Ann. Statist.* **12** 119–126.

LOPUHAÄ, H. P. (1992). Highly efficient estimators of multivariate location with high breakdown point. *Ann. Statist.* **20** 398–413.

MOSTELLER, F. and TUKEY, J. W. (1977). *Data Analysis and Regression.* Addison-Wesley, Reading, MA.

OKAFOR, R. (1990). A biweight approach to estimate currency exchange rate: The Nigerian example. *J. Appl. Statist.* **17** 73–82.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* IMS, Hayward, CA.

ROUSSEEUW, P. J. (1994). Unconventional features of positive-breakdown estimators. *Statist. Probab. Lett.* **19** 417–431.

ROUSSEEUW, P. J. and CROUX, C. (1994). The bias of k-step M-estimators. *Statist. Probab. Lett.* **20** 411–420.

SAKATA, S. and WHITE, H. (1995). An alternative definition of finite-sample breakdown point with applications to regression model estimators. *J. Amer. Statist. Assoc.* **90** 1099–1106.

STROMBERG, A. J. and RUPPERT, D. (1992). Breakdown in nonlinear regression. *J. Amer. Statist. Assoc.* **87** 991–998.

YLVISAKER, D. (1977). Test resistance. *J. Amer. Statist. Assoc.* **72** 551–556.

ZHANG, J. (1993). Anscombe type theorem for empirical processes. In *Proceedings of the First Scientific Conference of Chinese Postdoctors* (E. Feng, G. Gai and H. Zeng, eds.) 1244–1245. National Defence Industry Publishing House, Beijing.

ZHANG, J. (1996). The sample breakdown points of tests. *J. Statist. Plann. Inference* **52** 161–181.

ZHANG, J. (1997). The optimal breakdown M-test and score test. *Statistics.* To appear.

ZHANG, J. and LI, G. (1993). A new approach to asymptotic distributions of maximum likelihood ratio statistics. In *Statistical Science and Data Analysis: Proceedings of the Third Pacific Area Statistical Conference* (K. Matusita, M. L. Puri and T. Hayakawa, eds.) 325–336. VSP, The Netherlands.

INSTITUTE OF SYSTEMS SCIENCE
ACADEMIA SINICA
BEIJING 100080
CHINA
E-MAIL: jzhang@iss01.iss.ac.cn