ON THE STAHEL–DONOHO ESTIMATOR AND DEPTH-WEIGHTED MEANS OF MULTIVARIATE DATA

BY YIJUN ZUO,[1] HENGJIAN CUI[2] AND XUMING HE[3]

*Michigan State University, Beijing Normal University
and University of Illinois at Urbana-Champaign*

The depth of multivariate data can be used to construct weighted means as robust estimators of location. The use of projection depth leads to the Stahel–Donoho estimator as a special case. In contrast to maximal depth estimators, the depth-weighted means are shown to be asymptotically normal under appropriate conditions met by depth functions commonly used in the current literature. We also confirm through a finite-sample study that the Stahel–Donoho estimator achieves a desirable balance between robustness and efficiency at Gaussian models.

1. Introduction. Depth functions for multivariate data have piqued the interest of researchers in robust and nonparametric statistics for quite some time. A number of data depth measures, including the half-space depth of Tukey (1975), the simplicial depth of Liu (1990) and the projection depth discussed in Liu (1992) and Zuo and Serfling (2000a), have been proposed and discussed for outlier detection, data ranking and robust estimation. General discussions of the properties of data depth can be found in Liu and Singh (1993), He and Wang (1997), Rousseeuw and Hubert (1999) and Zuo and Serfling (2000b). More extensive studies of Tukey's half-space depth and the associated location estimators were made by Donoho and Gasko (1992) on robustness and Massé (1999) on asymptotics. The Stahel–Donoho estimator [Stahel (1981) and Donoho (1982)], a location estimator based on projection depth, has been studied by Tyler (1994) on robustness and by Maronna and Yohai (1995) on asymptotics. The limiting distribution of the Stahel–Donoho estimator has not been worked out, notwithstanding.

The present paper focuses on the asymptotic behavior of the depth-weighted *L*-type location estimators. Following Liu (1990) and Liu, Parelius and Singh (1999), we shall call them *DL*-estimators of location. We give sufficient conditions

under which the *DL*-estimators are asymptotically normal and note that these conditions have already been established in the literature for some commonly used depth measures. We then specialize to a class of projection depth functions and show that the corresponding *DL*-estimators satisfy all the conditions needed for Gaussian limiting distributions. The asymptotic normality of the Stahel–Donoho estimator follows as a special case.

The asymptotic properties established in the paper are basic to statistical inference based on depth or the induced statistics. They also supplement the work of Bai and He (1999), Zuo (2003) and Kim and Hwang (2001) on the asymptotic distributions of maximal depth estimators. The *DL*-estimators we consider here require positive weights to a portion of observations around the maximal depth point, and, consequently, their limiting distributions are still Gaussian. If we take a single point of maximal depth, the limiting distribution is no longer Gaussian. For location estimation, the efficiency consideration always favors using the observations around the "median," not just the "median" itself. The difference in the form of limiting distributions is yet another reason for choosing weighted means.

The rest of the paper is organized as follows. In Section 2, we establish, under appropriate conditions on the depth function and the weight function, a general asymptotic representation theorem for the *DL*-estimators from which the influence function can be identified and a Gaussian limiting distribution can be derived. Section 3 is devoted to the projection depth function and verifies that it satisfies all the conditions needed for the corresponding *DL*-estimators to have Gaussian limiting distributions. We also discuss the finite-sample efficiency and breakdown robustness of such location estimators. Technical proofs of all asymptotic results are provided in the Appendix.

2. Asymptotics of *DL*-estimators. *DL*-estimators were first considered in Liu (1990) and then discussed in great length in Liu, Parelius and Singh (1999). This section deals with the asymptotics of the estimators. We work with a general depth function $D(x, F)$ and the corresponding *DL*-estimators of multivariate location. More specifically, we assume, without loss of generality, that $D(x, F) \in [0, 1]$ and define for a multivariate distribution F and a weight function W the following functional:

$$(2.1) \qquad L(F) = \frac{\int x W(D(x, F)) F(dx)}{\int W(D(x, F)) F(dx)}.$$

Given a random sample $X_1, \ldots, X_n$, we shall write $L(F_n)$ as the empirical version of $L(F)$. The mean or trimmed mean can be viewed as an interesting case within our framework. The maximal depth estimator may be viewed as a limiting case of (2.1) where W places all its weight on the points with maximal depth, but it is excluded in this paper. Let us first impose the following conditions to ensure a

well-defined $L(F)$:

$$(2.2) \qquad \int W(D(x, F))F(dx) > 0, \qquad \int \|x\| \, W(D(x, F))F(dx) < \infty.$$

The first part of (2.2) holds in typical cases and the second part becomes trivial if $E\|X\| < \infty$ or if $W(D(x, F))$ is 0 outside some bounded set.

2.1. $\sqrt{n}$-consistency and asymptotic normality.

The asymptotic properties of $L(F_n)$ depend on those of $D_n(x) = D(x, F_n)$ as demonstrated in the following theorem. For the sake of convenience, we define

$$(2.3) \qquad v_n = \sqrt{n}(F_n - F), \qquad H_n(\cdot) = \sqrt{n}(D_n(\cdot, F_n) - D(\cdot, F))$$

and denote $\|H_n\|_\infty = \sup_{x \in \mathbb{R}^d} \|H_n(x)\|$. Let $D_r = \{x : D(x, F) \geq r\}$ for $r \geq 0$ and let W' be the derivative of W. For convenience, we itemize the assumptions as follows:

(A1) W' is continuous and $W(r) = 0$ for $r \in [0, \alpha r_0]$ with some $\alpha > 1$;
(A2) $\|H_n\|_\infty = O_p(1)$ and $\sup_{x \in D_{r_0}} \|x\| |H_n(x)| = O_p(1)$ for some $r_0 \geq 0$;
(A3) $\int \|x\|^2 (W(D(x, F)))^2 F(dx) < \infty$;
(A4) there exists $h(x, y)$ such that $H_n(x) = \int h(x, y)v_n(dy) + o_p(1)$ uniformly on $S_n \subset D_{r_0}$ with $P(D_{r_0} - S_n) = o(1)$ and $\int [\int \|y\| |W'(D(y, F))h(y, x)|F(dy)]^2 F(dx) < \infty$.

REMARK 2.1. Condition (A1) is rather trivial for a sufficiently smooth weight $W(r)$ that is 0 for r in a neighborhood of 0. Condition (A3) is required for the existence of the covariance matrix of the DL-estimator and becomes trivial if $E\|X\|^2 < \infty$ or if $W(D(x, F))$ is 0 outside some bounded set. The set D_{r_0} in (A2) could be replaced by any bounded set containing D_{r_0} or the whole space $\mathbb{R}^d$, depending on the applications.

THEOREM 2.1. *Assume* (2.2) *and let* $\theta = L(F)$.

1. *Under conditions* (A1) *and* (A2), *we have* $L(F_n) - \theta = O_p(1/\sqrt{n})$.
2. *Under conditions* (A1)–(A4), *we have*

$$L(F_n) - \theta = n^{-1} \sum_{i=1}^{n} (K_\theta(X_i) - EK_\theta(X)) + o_p(n^{-1/2}),$$

where

$$K_\theta(x) = \frac{\int (y - \theta)W'(D(y, F))h(y, x)F(dy) + (x - \theta)W(D(x, F))}{\int W(D(x, F))F(dx)}.$$

The detailed proof of Theorem 2.1 is given in the Appendix, but it is based on an approximation to the numerator of $\sqrt{n}L(F_n)$. Owing to equivariance of the estimator, we may assume without loss of generality that $\theta = L(F) = 0$. We then have

$$\sqrt{n}\int xW(D(x, F_n))F_n(dx)$$

$$= \int xW(D(x, F))v_n(dx) + \sqrt{n}\int x[W(D(x, F_n)) - W(D(x, F))]F_n(dx)$$

$$= \int xW'(\theta_n(x))H_n(x)F_n(dx) + \int xW(D(x, F))v_n(dx)$$

$$= \int \left\{ \int yW'(D(y, F))h(y, x)F(dy) + xW(D(x, F)) \right\} v_n(dx) + o_p(1)$$

for some $\theta_n(x)$ between $D(x, F_n)$ and $D(x, F)$. The last equality above is obtained by replacing $H_n(x)$ by its linear approximation given in (A4). The asymptotic representation in Theorem 2.1 follows from the above arguments. A consequence of the representation is

$$\sqrt{n}(L(F_n) - \theta) \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

with $\Sigma = \text{Cov}(K_\theta(X))$. In addition, by considering the differentiability of the functional $L(F)$, it can be formally proven that the influence function for the *DL*-estimators is of the form $K_\theta(x)$, but the technical details are not pursued here.

REMARK 2.2. Theorem 2.1 is stated for weak convergence but the strong consistency of the estimators can be established in a similar fashion if the in-probability bound of (A2) holds almost surely.

2.2. *A look at some common depth functions.* Theorem 2.1 provides a general means to establish the asymptotic normality for *DL*-estimators. Through the following examples, we shall illustrate that conditions (A2)–(A4) hold for some commonly used depth functions.

Most depth functions satisfy a so-called "vanishing at infinity" property in the sense that $\lim_{\|x\|\to\infty} D(x, F) = 0$; see Liu (1990) and Zuo and Serfling (2000a). When $r_0 > 0$, D_{r_0} is bounded for such depth functions. Consequently, (2.2) and (A3) become trivial for the smooth weight W in (A1), and the second part of (A2) follows directly from the first part. Therefore, if (A1) holds for some $r_0 > 0$, one just needs to focus on (A4) and the first part of (A2) when using Theorem 2.1.

EXAMPLE 2.1. The half-space depth function $HD(x, F) = \inf\{P(H_x)\}$, where H_x is any closed half-space containing x [Tukey (1975)]. Clearly, $HD(\cdot, \cdot)$ vanishes at ∞. Conditions (A2) and (A4) can be verified using the analyses of

Massé (1999). Therefore, for any weight function satisfying (A1) with $r_0 > 0$, the *DL*-estimator constructed with $HD(\cdot, \cdot)$ is asymptotically normal. In this case, Theorem 2.1 agrees with the result obtained by Massé (1999) for half-space depth trimmed means.

EXAMPLE 2.2. The simplicial depth function $SD(x, F) = P(x \in S[X_1, \ldots, X_{d+1}])$, where $X_1, \ldots, X_{d+1}$ is a random sample from F and $S[x_1, \ldots, x_{d+1}]$ denotes the d-dimensional simplex with vertices $x_1, \ldots, x_{d+1}$ [Liu (1990)]. By virtue of the asymptotic results of $SD(\cdot, \cdot)$ in Sections 2 and 3 of Dümbgen (1992), conditions (A2) and (A4) hold for $SD(\cdot, \cdot)$ with $r_0 > 0$. Asymptotics of $SD(\cdot, \cdot)$ are also treated in Arcones and Giné (1993).

It is worth noting that conditions (A2)–(A4) fail for *SD* and *HD* if $r_0 = 0$. Therefore, the asymptotic normality of the corresponding *DL*-estimators is established only when the observations with sufficiently low depth are trimmed. The next example allows for more general weighting.

EXAMPLE 2.3. The Mahalanobis depth $MD(x, F) = (1 + (x - \mu_F)' \Sigma_F^{-1} (x - \mu_F))^{-1}$, where μ_F and Σ_F are some location and scatter functionals at F [see Liu (1992) and Zuo and Serfling (2000a)]. If $n^{1/2}(\mu_{F_n} - \mu_F) = \int h_1(y) \nu_n(dy) + o_p(1)$ and $n^{1/2}(\Sigma_{F_n} - \Sigma_F) = \int h_2(y) \nu_n(dy) + o_p(1)$, it is then straightforward to check that (A2) and (A4) hold for $MD(\cdot, \cdot)$ with $r_0 = 0$ and $h(x, y) = (2(x - \mu_F)' \Sigma_F^{-1} h_1(y) + (x - \mu_F)' \Sigma_F^{-1} h_2(y) \Sigma_F^{-1}(x - \mu_F))/(1 + (x - \mu_F)' \Sigma_F^{-1}(x - \mu_F))^2$. If $W(r)$ is continuously differentiable on $[0, 1]$ and $E_F \|X\|^2 < \infty$, then (A1)–(A4) hold with $r_0 = 0$.

We shall show in the next section that (A1)–(A4) also hold for a class of projection depth functions with $r_0 = 0$.

3. Projection depth and the Stahel–Donoho estimator. In this section, we turn to a class of projection depth functions and the corresponding *DL*-estimators with the Stahel–Donoho estimator as a special case.

The projection depth functions discussed in the paper can be described as follows. Let μ and σ be location and scale functionals in $\mathbb{R}^1$. The outlyingness of a point $x \in \mathbb{R}^d$ with respect to a given distribution function F of $X \in \mathbb{R}^d$, $d \geq 1$, can be defined as

$$(3.1) \qquad O(x, F) = \sup_{\|u\|=1} g(x, u, F),$$

where

$$(3.2) \qquad g(x, u, F) = |u'x - \mu(F_u)|/\sigma(F_u),$$

with F_u the distribution of $u'X$, $u \in \mathbb{R}^d$. If $u'x - \mu(F_u) = \sigma(F_u) = 0$, we define $g(x, u, F)$ to be 0. The projection depth of a point $x \in \mathbb{R}^d$ with respect to the distribution F is then defined as

$$(3.3) \qquad PD(x, F) = 1/(1 + O(x, F)).$$

Throughout our discussions, μ and σ are assumed to exist for all univariate distributions involved. We also assume that μ is *translation* and *scale equivariant* and σ is *scale equivariant* and *translation invariant*, that is, $\mu(F_{sY+c}) = s\mu(F_Y) + c$ and $\sigma(F_{sY+c}) = |s|\sigma(F_Y)$, respectively, for any scalars s and c and random variable $Y \in \mathbb{R}^1$. The empirical versions of $g(x, u, F)$, $O(x, F)$ and $PD(x, F)$ shall be denoted by $g_n(x, u)$, $O_n(x)$ and $PD_n(x)$, respectively. They are obtained by replacing F by the corresponding empirical distribution F_n.

Note that (3.1) has long been used as a measure of outlyingness for multivariate data; see Mosteller and Tukey (1977), Stahel (1981) and Donoho (1982). It is affine invariant, and provides an ordering of points from the center outwards. Any monotone decreasing function of $O(x, F)$ can be taken as a measure of depth, but the particular choice of (3.3) ensures that $0 \le PD(x, F) \le 1$ for $x \in \mathbb{R}^d$. With μ and σ being the median (Med) and the median absolute deviation (MAD), respectively, Liu (1992) suggested $PD(x, F_n)$ as a depth function.

3.1. *Asymptotic normality.*   Since the projection depth function is based on a univariate location and scale functional, basic conditions on μ and σ are given first. We use F_{nu} as the empirical distribution function of $\{u'X_i, i = 1, \ldots, n\}$ for any $u \in \mathbb{R}^d$.

(C0) $\sup_{\|u\|=1}|\mu(F_u)| < \infty$, $\sup_{\|u\|=1}\sigma(F_u) < \infty$ and $\inf_{\|u\|=1}\sigma(F_u) > 0$;
(C1) $\sup_{\|u\|=1}|\mu(F_{nu}) - \mu(F_u)| = o_p(1)$, and $\sup_{\|u\|=1}|\sigma(F_{nu}) - \sigma(F_u)| = o_p(1)$;
(C2) $\sup_{\|u\|=1}\sqrt{n}|\mu(F_{nu}) - \mu(F_u)| = O_p(1)$ and $\sup_{\|u\|=1}\sqrt{n}|\sigma(F_{nu}) - \sigma(F_u)| = O_p(1)$;
(C3) the asymptotic representations

$$(3.4) \qquad \mu(F_{nu}) - \mu(F_u) = \frac{1}{n}\sum_{i=1}^{n} f_1(X_i, u) + o_p(1/\sqrt{n})$$

and

$$(3.5) \qquad \sigma(F_{nu}) - \sigma(F_u) = \frac{1}{n}\sum_{i=1}^{n} f_2(X_i, u) + o_p(1/\sqrt{n})$$

hold uniformly in u, and the graph set of $\{f_j(X, u): \|u\| = 1\}$ forms a polynomial set class with $E(f_j(X, u)) = 0$ for any $\|u\| = 1$, $E\sup_{\|u\|=1} f_j^2(X, u) < +\infty$ and

$$E \sup_{\substack{|u_1-u_2|\le\delta \\ \|u_1\|=\|u_2\|=1}} |f_j(X, u_1) - f_j(X, u_2)|^2 \to 0$$

as $\delta \to 0$ for $j = 1, 2$.

For more details on the polynomial set classes, we refer to Pollard (1984). Note that conditions (C1)–(C3) are nested in the sense that (C3) implies (C2) and (C2) is stronger than (C1). We shall discuss the applicability of these conditions later in the section. First, we give the following lemma that relates to the convergence property of the projection depth function. The proof is omitted here as it is similar to that of Theorem 2.3 of Zuo (2003).

LEMMA 3.1.   *Under* (C0), *we have*:

1. $\sup_{x \in \mathbb{R}^d}(1 + \|x\|) |PD(x, F_n) - PD(x, F)| = o_p(C1)$ *if* (C1) *holds*;
2. $\sup_{x \in \mathbb{R}^d}(1 + \|x\|) |PD(x, F_n) - PD(x, F)| = O_p(n^{-1/2})$ *if* (C2) *holds*.

For any x, let $u(x)$ be the set of directions satisfying $(u'x - \mu(F_u))/\sigma(F_u) = O(x, F)$. If $u(x)$ is a singleton, we also use $u(x)$ as the unique direction. If X is a continuous random variable, the nonuniqueness of $u(x)$ may occur at finitely many points.

We can now apply Theorem 2.1 to estimators based on the projection depth PD. Under (C3), condition (A2) holds with $r_0 = 0$ and $D_{r_0} = \mathbb{R}^d$ as implied by Lemma 3.1, and (A1) is automatically true for smooth weight functions. In the proof of the following theorem (see the Appendix), we show that (A3) and (A4) are indeed true for the projection depth.

THEOREM 3.1.   *Assume* (2.2), (C3) *and that* $W(r)$ *is continuously differentiable with* $W(0) = 0$. *In addition, assume that* $\mu(F_u)$ *and* $\sigma(F_u)$ *are continuous in* u, $\sigma(F_u) > 0$, *and* $u(x)$ *is a singleton except for finitely many points of* x *with zero probability. Then the projection depth-based estimator* $L(F_n)$ *is consistent for* $\theta = L(F)$ *and*

$$\sqrt{n}(L(F_n) - \theta) \xrightarrow{d} N(\mathbf{0}, \operatorname{Cov}(K_\theta(X))),$$

where

$$(3.6) \quad K_\theta(x) = \frac{\int(y - \theta)W'(PD(y, F))f(y, x)F(dy) + (x - \theta)W(PD(x, F))}{\int W(PD(x, F))F(dx)},$$

with

$$f(y, x) = \frac{O(y, F)f_2(x, u(y)) + f_1(x, u(y))}{\sigma(F_{u(y)})(1 + O(y, F))^2}.$$

Maronna and Yohai (1995) discussed the $\sqrt{n}$-consistency of the Stahel–Donoho estimator with μ and σ being Med and MAD functionals. We shall demonstrate that the asymptotic normality result of Theorem 3.1 holds for a wide class of depth-weighted means, including the Stahel–Donoho estimator.

For simplicity of presentation, we focus on any distribution F that is elliptically symmetric about θ. For any unit vector u, there exists a positive definite Σ_0 such

that $Y = (u'\Sigma_0 u)^{-1/2} u'(X - \theta)$ is a univariate symmetric variable with density function $p(y)$. Let $\mathrm{MAD}(Y) = m_0$. In the special case of multivariate normality with $\Sigma_0 = \mathrm{Cov}(X)$, $m_0 = \Phi^{-1}(\frac{3}{4})$, where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.

Under the elliptical symmetry of F, the choice of $\sigma(\cdot)$ has no impact on the covariance matrix. To see this, assume, without loss of generality, that $\theta = 0$ and $\sigma(\cdot)$ is Fisher consistent for a given scale parameter, say, $m_0(u'\Sigma_0 u)^{1/2}$ at any F_u. Because of scale equivariance, $f_2(x, u(y))$ is an even function in y, and so are the outlying function $O(y, F)$ and the depth function $PD(y, F)$. Therefore, $\int y W'(PD(y, F)) f_2(x, u(y)) O(y, F)/(\sigma(F_{u(y)})(1 + O(y, F))^2 F(dy)$ is equal to 0 by symmetry, which means that the function $K_\theta(x)$ does not depend on the choice of the scale functional. For this reason, we shall use the MAD functional throughout the paper.

We now consider any M-functional μ that satisfies $E\psi((Y - \mu)/\mathrm{MAD}(Y)) = 0$ for some odd and bounded score function ψ. For technical reasons, we assume ψ to be differentiable anywhere except at finitely many points. The following lemma is based on Cui and Tian (1994).

LEMMA 3.2. *If $p(y)$ is continuous with $p(0) p(m_0) > 0$, then, for the above choice of μ and $\sigma = \mathrm{MAD}$, conditions (C0)–(C3) are satisfied with*

$$f_1(x, u) = \frac{m_0 \sqrt{u'\Sigma_0 u}}{E\psi'(Y/m_0)} \psi\left(u'(x - \theta)/m_0\sqrt{u'\Sigma_0 u}\right)$$

and

$$f_2(x, u) = \frac{\sqrt{u'\Sigma_0 u}}{2p(m_0)}\left(\frac{1}{2} - I\{|u'(x - \theta)| \leq m_0\sqrt{u'\Sigma_0 u}\}\right).$$

In the special case with $(\mu, \sigma) = (\mathrm{Med}, \mathrm{MAD})$, we have $\psi(x) = \mathrm{sign}(x)$, and the expression for $f_1(x, u)$ specializes to $f_1(x, u) = (\sqrt{u'\Sigma_0 u})/(p(0))(1/2 - I\{u'(x - \theta) \leq 0\})$. This case was handled by Cui and Tian (1994), but the same arguments apply to other M-functionals. We also refer to Zhang (1991) for a general treatment that leads to uniform asymptotic results as in (C3).

If the weight function W is continuously differentiable, Lemma 3.2 implies that the Stahel–Donoho estimator satisfies the conditions of Theorem 3.1 and therefore is asymptotically normal. Lemma 3.2 also indicates that the Stahel–Donoho estimator has a bounded influence function in the form of (3.6). We shall now derive the covariance matrix for the Stahel–Donoho estimator and make asymptotic efficiency comparisons as follows.

Let $V = \|\Sigma_0^{-1/2}(X - \theta)\|/m_0$ and $U = |e_1'\Sigma_0^{-1/2}(X - \theta)|/\|\Sigma_0^{-1/2}(X - \theta)\|$ with $e_1 = (1, 0, \ldots, 0)'$. Let $c_1 = E W(1/(1 + V))$ and $c_2 = m_0(E\psi'(Y/m_0))^{-1} \times E\{V W'(1/(1 + V))/(1 + V)^2\}$. By Theorem 3.1 and using the fact that U is

independent of V, the asymptotic variance–covariance matrix of the Stahel–Donoho estimator takes a simpler form:

$$(3.7) \qquad \Sigma = (dc_1^2)^{-1} E_V \big(m_0 V W (1/(1+V)) + c_2 E_U [U \psi (UV)] \big)^2 \Sigma_0$$

as compared to Σ_0, the covariance matrix of $\sqrt{n}(\overline{X}_n - \theta)$ when X is Gaussian. Note that $U^2 \sim \text{Beta}(1/2, (d-1)/2)$ and, under normality, $(m_0 V)^2 \sim \chi_d^2$. The relative asymptotic efficiency would then depend on the dimension d and the weight function W.

We suggest using $W(r)$, which would assign weight 1 to the half of the points with higher depth. This would balance efficiency with robustness. The other half of the points with lower depth could be viewed as outliers, so a low weight should be given. An ad hoc choice of the weight function with the above property in mind takes the form of

$$W(r) = \begin{cases} \dfrac{\exp(-K(1 - r/C)^2) - \exp(-K)}{1 - \exp(-K)}, & \text{if } r < C, \\ 1, & \text{if } r \geq C, \end{cases}$$

where the parameter K controls how much we would like to discriminate points with different projection depth, but the parameter C is defined to be the median of $PD(X, F)$, which is about 0.36 for the bivariate standard normal model. In practice, a consistent estimate of C, the median of $PD(X_1), \ldots, PD(X_n)$, may be used. See Figure 1 for this weight function with two different sets of tuning parameters.

For the above weight function with $K = 3$, Table 1 gives the relative asymptotic efficiencies of the depth-weighted means with respect to the sample mean at the normal model for two choices of the location functionals, mean and median. These two choices represent the two extremes of Huber-type M-functionals. It is clear that the depth-weighted means are highly efficient at the normal model and the
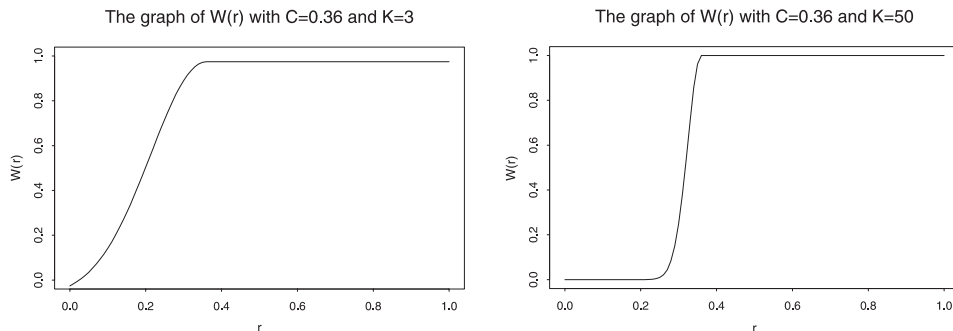

FIG. 1. *Behavior of $W(r)$ with different K values.*

$\mu(\cdot)$	$d = 2$	$d = 3$	$d = 10$
Mean	0.956	0.970	0.992
Median	0.935	0.957	0.990

efficiency varies only slightly with the choice of μ. For instance, the Stahel–Donoho estimator with $\mu = $ Med has an asymptotic relative efficiency of nearly 96% at $d = 3$ as compared to 97% with μ being the mean functional. For any Huber-type M-functional, the efficiency lies in between.

3.2. *Breakdown point.* Here we consider finite-sample robustness of the Stahel–Donoho-type estimators when μ and σ are the Med and modified MAD functionals.

The notion of finite-sample breakdown was introduced in Donoho and Huber (1983). Let $X^n = \{X_1, \ldots, X_n\}$ be a sample of size n from X in $\mathbb{R}^d$, $d \geq 1$. The replacement breakdown point (RBP) of an estimator T at X^n is defined as

$$\text{RBP}(T, X^n) = \min\left\{\frac{m}{n} : \sup_{X_m^n} \|T(X_m^n) - T(X^n)\| = \infty\right\},$$

where X_m^n denotes a contaminated sample by replacing m points of X^n with arbitrary values.

In the following discussion, μ and σ are taken to be the Med and MAD_k, where MAD_k is a modification of MAD resulting in an estimator, denoted by L_n^k, with a breakdown point reaching the upper bound given in Davies (1987). More specifically, define

$$\text{MAD}_k(x^n) = \text{Med}_k\big(\{|x_1 - \text{Med}(x^n)|, \ldots, |x_n - \text{Med}(x^n)|\}\big)$$

and

$$\text{Med}_k(x^n) = \big(x_{(\lfloor (n+k)/2 \rfloor)} + x_{(\lfloor (n+1+k)/2 \rfloor)}\big)/2$$

for $1 \leq k \leq n$, where $\lfloor x \rfloor$ is the largest integer no larger than x, and $x_{(1)} \leq \cdots \leq x_{(n)}$ are the ordered values of $x_1, \ldots, x_n$ in $\mathbb{R}^1$. The usual Med–MAD combination corresponds to L_n^k with $k = 1$.

A random sample X^n is said to be *in general position* if there are no more than d sample points of X^n lying in any $(d-1)$-dimensional subspace. The breakdown point of L_n^k is given in the following theorem.

THEOREM 3.2. *Let $W(r)$ be continuous on $[0, 1]$ and let $0 < W(r) < Mr$ on $(0, 1]$ for some $M > 0$. Then, for any X^n in general position with $n > 2d + k$,*

$$\text{RBP}(L_n^k, X^n) = \begin{cases} \dfrac{\lfloor (n - k + 2)/2 \rfloor}{n}, & \text{if } d = 1, \\[2mm] \min\left\{ \dfrac{\lfloor (n - k + 2)/2 \rfloor}{n}, \dfrac{\lfloor (n + k - 1)/2 \rfloor}{n} \right\}, & \text{if } d = 2, \\[2mm] \min\left\{ \dfrac{\lfloor (n - k + 2)/2 \rfloor}{n}, \dfrac{\lfloor (n - 2d + 1 + k)/2 \rfloor}{n} \right\}, & \text{if } d > 2. \end{cases}$$

The main idea of the proof is as follows. The estimator L_n^k breaks down only if $\text{Med}(F_{nu})$ or $\text{MAD}_k(F_{nu})$ breaks down for some direction u. However, $\text{Med}(F_{nu})$ is uniformly bounded from above over the directions u for any contamination less than 50%. So the RBP of L_n^k depends solely on the smallest RBP of $\text{MAD}_k(F_{nu})$ over the directions u. This corresponds to the uniform breakdown point of MAD_k used by Tyler (1994). Since MAD can be exploded ($\to \infty$) or imploded ($\to 0$), the RBP of L_n^k is determined by two quantities for $d \geq 2$, corresponding to the explosion and implosion of MAD_k, respectively. To explain the difference in the expressions of RBP for $d = 2$ and $d > 2$, we note that there are more ways to place least favorable outliers in three or higher dimensions than in two dimensions. For example, any two lines in $d = 2$ can intersect at one (finite) point, but any two planes in $d = 3$ may intersect at a line. Outliers may be placed on the intersecting line that are arbitrarily far away from the bulk of the data.

REMARK 3.1. (i) Theorem 3.2 focuses on the case $(\mu, \sigma) = (\text{Med}, \text{MAD}_k)$. The result, however, can be extended to general M-functionals μ and σ with the same uniform breakdown points as those of Med and MAD_k. More generally, the RBP of the *DL*-estimators will be no less than the minimum of the uniform RBP's of any functionals μ and σ. (ii) For our choice $(\mu, \sigma) = (\text{Med}, \text{MAD}_k)$, the RBP of L_n^k reaches $\lfloor (n - d + 2)/2 \rfloor / n$ for $d \leq 2$ (when $k = d$) or $\lfloor (n - d + 1)/2 \rfloor / n$ for $d > 2$ (when $k = d$). The latter attains the upper bound on RBP for affine equivariant location and scatter estimators [see Davies (1987)].

Our modification of MAD is similar to that of Gather and Hilker (1997) and slightly more general than that of Tyler (1994). The breakdown point of a similar estimator was given in Tyler (1994) without detailed calculations but a misprint there causes the difference between his and ours. We hope that Theorem 3.2 documents a complete answer to breakdown points of L_n^k, including the Stahel–Donoho estimator.

3.3. *Finite sample efficiency.* There is always a valid concern that a high-breakdown-point estimator may lose efficiency at the Gaussian model. The

CM-estimators of Kent and Tyler (1996) and the cross-checking method considered in He (1991) and He and Wang (1996) are just a few examples to retain high efficiency for high-breakdown estimators. We shall demonstrate via simulations that the Stahel–Donoho estimator can achieve a good balance between robustness and efficiency.

With the specific choice of W in Section 3.1, we conducted a simulation study to see how L_n^1 performs in finite samples at normal and contaminated normal models. We generated 25,000 samples from the bivariate standard normal distribution for different sample sizes $n = 10, 20, \ldots, 100$. An approximate algorithm with time complexity $O(n^3)$ for fixed $d = 2$ was utilized for computing all $PD_n(X_i)$, $i = 1, \ldots, n$, and the weighted mean L_n^1. We calculate for any estimator T the empirical mean squared error (EMSE) with EMSE $= (1/m) \sum_{j=1}^m \|T_j - \theta\|^2$, where $m = 25{,}000$, $\theta = (0, 0)^\tau$ and T_j is the estimate from the jth sample. The relative efficiency (RE) of T is then obtained by dividing the EMSE of the sample mean by that of T.

We also considered the contaminated samples from the model $(1 - \varepsilon) \times N((0, 0)^\tau, \mathbf{I}) + \varepsilon N((\mu_1, \mu_2)^\tau, \sigma^2 \mathbf{I})$ with $\mu_1 = \mu_2 = 10$ and $\sigma = 5$, $\varepsilon = 0\%$, 10% and 20%, where $\mathbf{I}$ is the bivariate identity matrix. Some simulation results are given in Table 2.

The finite-sample relative efficiency of L_n^1 for the bivariate standard normal model is about 93%. This is substantially higher than those of the transformation–retransformation median [Chakraborty, Chaudhuri and Oja (1998)], the half-space median [Rousseeuw and Ruts (1998)] and the projection median [Zuo (2003)] at about 72%, 76% and 79%, respectively. Table 2 shows that the *DL*-estimator is highly competitive with the optimal estimator (sample mean) at the normal

TABLE 2
Empirical mean squared error and relative efficiency

n		$\varepsilon = 0\%$		$\varepsilon = 10\%$		$\varepsilon = 20\%$	
		DL	Mean	DL	Mean	DL	Mean
20	EMSE	0.1038	0.0951	0.1296	2.3350	0.1957	8.5222
	RE	0.92	1.00	18.01	1.00	43.54	1.00
40	EMSE	0.0524	0.0487	0.0692	2.1800	0.1453	8.3873
	RE	0.93	1.00	31.52	1.00	57.73	1.00
60	EMSE	0.0338	0.0313	0.0523	2.101	0.1255	8.1928
	RE	0.92	1.00	40.21	1.00	65.30	1.00
80	EMSE	0.0261	0.0243	0.0412	2.0641	0.1151	8.1115
	RE	0.93	1.00	50.12	1.00	70.47	1.00
100	EMSE	0.0215	0.0200	0.0354	2.0524	0.1124	8.1291
	RE	0.93	1.00	58.00	1.00	72.32	1.00

model but performs much better when there is contamination. This is true for all sample sizes considered. Our results are consistent with a smaller scale simulation conducted by Maronna and Yohai (1995) using a different weight function.

The high efficiency of this estimator at the Gaussian model is directly attributable to the form of the weight function. It is helpful that the weight function decreases to 0 smoothly for outlying observations. If hard trimming is used so that $W(r)$ takes the value 1 for the (deeper) half of the observations but 0 for the other half, the efficiency of the estimator can drop noticeably. In the univariate case, for example, such a depth-based trimming would yield a location estimator that is even less efficient than the maximal depth point (i.e., median). Of course, it will also be much less efficient than the usual 25%-trimmed mean at the Gaussian model.

REMARK 3.2. Like all other high-breakdown estimators, the Stahel–Donoho estimator is computationally intensive. An exact algorithm is available from the first author for $d = 2$ with the computational complexity of $O(n^2)$ for each depth calculation. In higher dimensions, approximate algorithms are usually used to compute depth by searching over a large number of directions perpendicular to hyperplanes that pass through d observations; see Stahel (1981) and Maronna and Yohai (1995). Future research on faster algorithms and the ever-increasing power of computers is expected to make the computation of high-breakdown procedures less painful.

4. Concluding remarks. In the present paper, we provide a set of sufficient conditions on the depth and weight functions under which the depth-weighted means are asymptotically normal. For the half-space depth and the simplicial depth, those conditions have been positively established in the recent literature. A main purpose of the paper is to show that they also hold for a class of projection depth functions. As a consequence, the Stahel–Donoho estimator is shown to have an asymptotically normal distribution, which allows us to study its efficiency and influence function and to carry out large sample inference.

We expect future advances in algorithms and in computer power to make possible routine use of the Stahel–Donoho estimator in high dimensions. However, we have demonstrated in the paper that the estimator enjoys several good properties, including affine equivariance, high breakdown point (in any dimension), high finite-sample and asymptotic efficiencies at the normal model and a bounded influence function.

APPENDIX

PROOF OF THEOREM 2.1. By equivariance, assume, without loss of generality, that $\theta = L(F) = 0$. We shall prove part (2). The proof for part (1) is similar and is omitted.

First, observe that

$$\sqrt{n} \int x\, W(D(x, F_n)) F_n(dx)$$

$$= \sqrt{n} \int x\, W(D(x, F_n)) F_n(dx) - \sqrt{n} \int x\, W(D(x, F)) F(dx)$$

$$= \int x\, W'(\theta_n(x)) H_n(x) F_n(dx) + \int x\, W(D(x, F)) v_n(dx),$$

where $\theta_n(x)$ is a point between $D(x, F_n)$ and $D(x, F)$ and $\sup_{x \in \mathbb{R}^d} |\theta_n(x) - D(x, F)| = O_p(1/\sqrt{n})$.

On one hand, we have

$$\left| \int x\, (W'(\theta_n(x)) - W'(D(x, F))) H_n(x) F_n(dx) \right|$$

$$\leq \int_{\{\theta_n(x)>r_1\} \cup \{D(x,F)>r_1\}} \|x\|\, |H_n(x)|\, |(W'(\theta_n(x)) - W'(D(x, F)))| F_n(dx)$$

$$\leq \int_{\{D(x,F)+O_P(1/\sqrt{n}) \geq r_1\}} \|x\|\, |H_n(x)|\, |(W'(\theta_n(x)) - W'(D(x, F)))| F_n(dx)$$

$$= o_p(1),$$

by (A1) and (A2), and thus

$$\int x\, W'(\theta_n(x)) H_n(x) F_n(dx)$$

$$= \int x\, W'(D(x, F)) H_n(x) F_n(dx) + o_p(1).$$

On the other hand,

$$\left| \int x\, W'(D(x, F)) H_n(x)(F_n - F)(dx) \right|$$

$$\leq \frac{1}{\sqrt{n}} \left| \int_{D_{r_0}} x H_n(x) W'(D(x, F)) v_n(dx) \right|$$

$$+ \frac{1}{\sqrt{n}} \left| \int_{\bar{D}_{r_0}} x\, W'(D(x, F)) H_n(x) v_n(dx) \right|$$

$$= o_p(1),$$

and therefore

$$\int x\, W'(\theta_n(x)) H_n(x) F_n(dx)$$

$$= \int x\, W'(D(x, F)) H_n(x) F(dx) + o_p(1).$$

By Fubini's theorem and (A4),

$$\int x\, W'(\theta_n(x))H_n(x)F_n(dx) = \int x\, W'(D(x,F))H_n(x)F(dx) + o_p(1)$$

$$= \int_{D_{r_0}} x\, W'(D(x,F))\left(\int h(x,y)v_n(dy)\right)F(dx) + o_p(1)$$

$$= \iint y\, W'(D(y,F))h(y,x)F(dy)v_n(dx) + o_p(1).$$

Likewise, we can show that

$$\int W(D(x,F_n))F_n(dx) = \int W(D(x,F))F(dx) + O_p(1/\sqrt{n}).$$

The desired result now follows immediately from (A3), (A4), the central limit theorem and Slutsky's theorem. □

PROOF OF THEOREM 3.1.   We employ Theorem 2.1. By Lemma 3.1 and the given conditions, (A1) and (A2) hold with $D_{r_0} = \mathbb{R}^d$ and $r_0 = 0$. Following the proof of Theorem 2.2 of Zuo (2003), we see that $\|x\|^2(PD(x,F))^2$ is uniformly bounded with respect to $x \in \mathbb{R}^d$. Thus,

$$\int \|x\|^2(W(PD(x,F)))^2 F(dx) \le \int \|x\|^2(PD(x,F))^2\left(\sup_{0 \le r \le 1} W'(r)\right)F(dx)$$

is finite. Hence, (A3) holds. To check (A4), we establish the following two lemmas. For the rest of the paper, we assume, without loss of generality, that $u(x)$ is a singleton except for $x = 0$ and that $P(X = 0) = 0$.

For the rest of the Appendix, we use an alternative definition $g(x,u,F) = (u'x - \mu(F_u))/\sigma(F_u)$ to (1.2). For the projection depth when μ and σ are affine equivariant, this leads to the same depth function $PD(x,F)$. □

LEMMA A.1.   *Let* $g(\delta) = \inf_{1/M \le \|x\| \le M} \inf_{\|u-u(x)\| \ge \delta, \|u\|=1}[g(x,u(x),F) - g(x,u,F)]$, *where* $M = M(\delta)$ *is continuous and monotone in* δ *with* $\lim_{\delta \downarrow 0} M(\delta) = \infty$. *Then, under condition* (1) *of Theorem* 3.1, $g(\cdot)$ *is a continuous and increasing function with* $g(0) = 0$ *and* $g(\delta) > 0$ *for any* $\delta > 0$.

PROOF.   Define $g^*(x,\delta) = \inf_{\|u-u(x)\| \ge \delta, \|u\|=1}(g(x,u(x),F) - g(x,u,F))$, which is monotone in $\delta > 0$. It is easy to see that $g^*(x,\delta)$ is continuous in x and δ, is positive for $\delta > 0$ and $x \ne 0$ and tends to 0 as $\delta \to 0$. Since $g(\delta) = \inf_{1/M \le \|x\| \le M} g^*(x,\delta)$, we conclude that $g(\delta)$ is continuous, nonnegative and monotone with $g(0) = 0$. Now we show that $g(\delta)$ is positive for $\delta > 0$.

If $g(\delta) = 0$ for some $\delta > 0$, then there exists a sequence x_m satisfying $1/M \le \|x_m\| \le M$ such that $g^*(x_m,\delta) \to 0$ as $m \to \infty$. We can then choose a subsequence $\{x_{m'}\}$ of $\{x_m\}$ such that $x_{m'} \to x_0$ ($\|x_0\| \ge 1/M > 0$). By the continuity of $g^*(x,\delta)$ in x, we have $g^*(x_0,\delta) = 0$. This contradicts the positivity of $g^*(x,\delta)$ for $\delta > 0$. □

LEMMA A.2. *Under the conditions of Theorem* 3.1, *there exists a sequence of sets $S_n \subset \mathbb{R}^d$ such that $1 - P\{S_n\} = o(1)$ and $H_n(x) = \int f(x, y)\nu_n(dy) + o_p(1)$ uniformly over S_n with*

$$f(x, y) = \frac{O(x, F)f_2(y, u(x)) + f_1(y, u(x))}{\sigma(F_{u(x)})(1 + O(x, F))^2}.$$

PROOF. By the conditions of Theorem 3.1, (C0) and (C2) hold so that we can apply Lemma 3.1 to verify that

$$
\begin{aligned}
PD&(x, F_n) - PD(x, F) \\
&= -\frac{O(x, F_n) - O(x, F)}{(1 + O(x, F_n))(1 + O(x, F))} \\
&= -\frac{\sup_{\|u\|=1} g(x, u, F_n) - g(x, u(x), F)}{(1 + O(x, F))^2} + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= -\frac{g(x, u(x), F_n) - g(x, u(x), F)}{(1 + O(x, F))^2} - \frac{I_n(x)}{(1 + O(x, F))^2} + o_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where $I_n(x) = \sup_{\|u\|=1} I_n(u, x)$ and

$$
\begin{aligned}
I_n(u, x) &= \left[g(x, u, F_n) - g(x, u, F) - \left(g(x, u(x), F_n) - g(x, u(x), F)\right)\right] \\
&\quad - \left[g(x, u(x), F) - g(x, u, F)\right] \\
&=: J_n(u, x) - \left[g(x, u(x), F) - g(x, u, F)\right] \le J_n(u, x).
\end{aligned}
$$

Denote $t(\delta) = \sum_{j=1}^4 t_j(\delta)$, where

$$
\begin{aligned}
t_j(\delta) &= \left\{E\left[\sup_{\|u_1 - u_2\| \le \delta} |f_j(X, u_1) - f_j(X, u_2)|^2\right]\right\}^{1/2}, \qquad j = 1, 2, \\
t_3(\delta) &= \sup_{\|u_1 - u_2\| \le \delta} |\mu(F_{u_1}) - \mu(F_{u_2})|, \\
t_4(\delta) &= \sup_{\|u_1 - u_2\| \le \delta} |\sigma(F_{u_1}) - \sigma(F_{u_2})|.
\end{aligned}
$$

By the conditions of Theorem 3.1, $t(\delta) \to 0$ as $\delta \to 0$. Let $M(\delta) = [t(\delta) + \delta]^{-1/2}$. Then $M(\delta) \to \infty$ and $M(\delta)t(\delta) \to 0$ as $\delta \to 0$. By Lemma A.1, we can choose a positive sequence $\delta_n \to 0$, such that $\sqrt{n}g(\delta_n) \to \infty$. Let $\tau_n = 1/M(\delta_n)$, $M_n = M(\delta_n)$, and $S_n = \{x : \tau_n \le \|x\| \le M_n\}$. Then $\tau_n \to 0$ and $M_n \to \infty$ and $P\{S_n\} \to 1$ as $n \to \infty$.

Now we show that $I_n(x) = o_p(1/\sqrt{n})$ uniformly for $x \in S_n$. Observe that, for $x \in S_n$,

$$I_n(x) = \max\left\{\sup_{\|u - u(x)\| \le \delta_n} I_n(u, x), \sup_{\|u - u(x)\| \ge \delta_n} I_n(u, x)\right\}.$$

Since

$$\sup_{x \in S_n, \|u - u(x)\| \geq \delta_n} \sqrt{n} I_n(u, x) \leq \sup_{x \in S_n} \sqrt{n} |J_n(u, x)| - \sqrt{n} g(\delta_n) \to -\infty$$

as $n \to \infty$ and $I_n(x) \geq I_n(u(x), x) = 0$, thus, for $x \in S_n$ and n sufficiently large,

$$\sqrt{n} I_n(x) = \sup_{\|u - u(x)\| \leq \delta_n} \sqrt{n} I_n(u, x) \leq \sup_{\|u - u(x)\| \leq \delta_n} \sqrt{n} |J_n(u, x)|.$$

If we can show that $\sup_{x \in S_n, \|u - u(x)\| \leq \delta_n} \sqrt{n} J_n(u, x) = o_p(1)$, then $I_n(x) = o_p(1/\sqrt{n})$ uniformly over S_n.

Write $l_n(u) = \mu(F_{nu}) - \mu(F_u)$ and $s_n(u) = \sigma(F_{nu}) - \sigma(F_u)$. Then, for $x \in S_n$, $\|u - u(x)\| \leq \delta_n$ and for sufficiently large n we have

$$\sqrt{n} J_n(u, x)$$
$$= -\frac{(u'x - \mu(F_u))\sqrt{n} s_n(u) + \sigma(F_u)\sqrt{n} l_n(u)}{\sigma^2(F_u)}$$
$$+ \frac{(u(x)'x - \mu(F_{u(x)}))\sqrt{n} s_n(u(x)) + \sigma(F_{u(x)})\sqrt{n} l_n(u(x))}{\sigma^2(F_{u(x)})} + o_p(1)$$
$$= -\frac{(u'x - \mu(F_u))(1/\sqrt{n}) \sum_{i=1}^n f_2(X_i, u) + \sigma(F_u)(1/\sqrt{n}) \sum_{i=1}^n f_1(X_i, u)}{\sigma^2(F_u)}$$
$$+ \frac{(u(x)'x - \mu(F_{u(x)}))(1/\sqrt{n}) \sum_{i=1}^n f_2(X_i, u(x))}{\sigma^2(F_{u(x)})}$$
$$+ \frac{\sigma(F_{u(x)})(1/\sqrt{n}) \sum_{i=1}^n f_1(X_i, u(x))}{\sigma^2(F_{u(x)})} + o_p(1)$$
$$=: J_{1n}(u, x) + J_{2n}(u, x) + o_p(1),$$

where

$$J_{1n}(u, x) = \left(\frac{1}{\sigma(F_{u(x)})} - \frac{1}{\sigma(F_u)} \right) \frac{1}{\sqrt{n}} \sum_{i=1}^n f_1(X_i, u)$$
$$+ \left(\frac{u(x)'x - \mu(F_{u(x)})}{\sigma^2(F_{u(x)})} - \frac{u'x - \mu(F_u)}{\sigma^2(F_u)} \right) \frac{1}{\sqrt{n}} \sum_{i=1}^n f_2(X_i, u)$$

and

$$J_{2n}(u, x) = \frac{1}{\sigma(F_{u(x)})} \frac{1}{\sqrt{n}} \sum_{i=1}^n (f_1(X_i, u(x)) - f_1(X_i, u))$$
$$+ \frac{u(x)'x - \mu(F_{u(x)})}{\sigma^2(F_{u(x)})} \frac{1}{\sqrt{n}} \sum_{i=1}^n (f_2(X_i, u(x)) - f_2(X_i, u)).$$

By the conditions of Theorem 3.1, we see that

$$\sup_{x \in S_n, \|u-u(x)\| \leq \delta_n} |J_{1n}(u, x)| = o_p(1).$$

By the Cauchy–Schwarz inequality, the maximum inequality [see Pollard (1990), Chapter 5] and the conditions of Theorem 3.1, we have

$$E\left(\sup_{x \in S_n, \|u-u(x)\| \leq \delta_n} |J_{2n}(u, x)|^2\right)$$

$$\leq CE\left[\sup_{\|u_1-u_2\| \leq \delta_n} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [f_1(X_i, u_1) - f_1(X_i, u_2)]\right)^2\right]$$

$$+ CM_n^2 E\left[\sup_{\|u_1-u_2\| \leq \delta_n} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [f_2(X_i, u_1) - f_2(X_i, u_2)]\right)^2\right]$$

$$\leq CE\left[\sup_{\|u_1-u_2\| \leq \delta_n} \left(f_1(X, u_1) - f_1(X, u_2)\right)^2\right]$$

$$+ CM_n^2 E\left[\sup_{\|u_1-u_2\| \leq \delta_n} \left(f_2(X, u_1) - f_2(X, u_2)\right)^2\right]$$

$$\leq C\left[t(\delta_n)^2 + (M_n t(\delta_n))^2\right] \to 0$$

as $n \to \infty$, where C stands for some generic constant that may vary from line to line. It then follows that

$$\sup_{x \in S_n, \|u-u(x)\| \leq \delta_n} |J_{2n}(u, x)| = o_p(1),$$

and therefore $\sqrt{n} I_n(x) = o_p(1)$ uniformly for $x \in S_n$.

Now we have

$$PD(x, F_n) - PD(x, F)$$

$$= -\frac{g(x, u(x), F_n) - g(x, u(x), F)}{(1 + O(x, F))^2} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{(u(x)'x - \mu(F_{u(x)})/\sigma^2(F_{u(x)})) f_2(X_i, u(x))}{(1 + O(x, F))^2}\right.$$

$$\left. + \frac{(1/\sigma(F_{u(x)})) f_1(X_i, u(x))}{(1 + O(x, F))^2}\right) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where $o_p(\cdot)$ is uniform for $x \in S_n$. Hence, $H_n(x) = \int f(x, y)\nu_n(dy) + o_p(1)$ uniformly over S_n with $f(x, y)$ as specified in the lemma. □

PROOF OF THEOREM 3.2.   We focus on the cases with $d = 2$ and $d > 2$. The case with $d = 1$ is simpler and thus omitted here. Recall that $(u'y -$

$\text{Med}(u'Z))/\text{MAD}_k(u'Z) = 0$ if $u'y - \text{Med}(u'Z) = \text{MAD}_k(u'Z) = 0$ for any $Z = \{Z_1, \ldots, Z_n\}$ in $\mathbb{R}^d$.

First, we consider the case with $d = 2$ and $k = 1$.

(i) $m = \lfloor n/2 \rfloor$ *points are sufficient for breakdown of* L_n^k. Let l be a line determined by two sample points from X^n and let X_j be a point not on l. Consider replacing m other points in X^n (not from l or X_j) by y, a point on the line l. Choose two unit vectors u_1 and u_2 perpendicular to l and to the line connecting y and X_j, respectively. Since $m + 1 > n/2$, it follows that $O(Z_i, Z) = \infty$ for all points except at y, where Z denotes the contaminated sample. This means that $L_n^k(Z) = y$, so a breakdown occurs as $\|y\| \to \infty$.

(ii) $m = \lfloor n/2 \rfloor - 1$ *points are not sufficient for breakdown of* L_n^k. Since $m < \lfloor (n+1)/2 \rfloor$ and $n - m > n/2$, it is clear that there exists a constant $C < \infty$ such that $|\mu_n(u'Z)|, \sigma_n(u'Z) \leq C$ uniformly over any unit vector u and contaminated sample Z. It suffices to show that the numerator of L_n^k is always bounded above and the denominator is bounded away from 0. To see the former, note that, for any nonzero $Z_i \in Z$ and $u_0 = Z_i/\|Z_i\|$, we have $\|Z_i\| \leq \sigma_n(u_0'Z)O(Z_i, Z) + |\mu_n(u_0'Z)| \leq C/PD(Z_i, Z)$. By the assumption that $W(r) < Mr$, we have $W(PD(Z_i, Z))\|Z_i\| \leq MC$, implying that the numerator of L_n^k is bounded above.

Now we show that the denominator of L_n^k is uniformly bounded away from 0 for any Z. Otherwise, there is a sequence of contaminating data sets $\{Z_t\}$ with $Z_t = \{Z_{1t}, \ldots, Z_{nt}\}$ such that $W(PD(Z_{it}, Z_t)) \to 0$ as $t \to \infty$ for $1 \leq i \leq n$. Since there are at least $n - m$ points in Z_t such that $|u'Z_{it} - \mu_n(u'Z_t)|$ is uniformly bounded for any u, $\sigma_n(u_t'Z_t)$ must approach 0 for some unit vector u_t as $t \to \infty$. This happens only if $\lfloor n+2 \rfloor/2$ points in Z_t approach (or lie on) the same line as $t \to \infty$. Hence, we have to move at least m points in X^n to a line l_t determined by two points in X^n. However, for points Z_{it} approaching or on l_t to have $W(PD(Z_{it}, Z_t)) \to 0$ as $t \to \infty$, there must be $\lfloor n+2 \rfloor/2$ points of Z_t approaching or on another line l_t' determined by points in X^n. Thus, l_t' must intersect l_t at a point y_t and the m contaminating points must approach y_t as $t \to \infty$. Since there are only finitely many y_t, we can assume w.l.o.g. that $y_t = y$ for sufficiently large t by taking a subsequence of $\{Z_t\}$ if necessary. For simplicity, assume n is odd. For a given unit vector u_t, assume that $|u_t'Z_{k1t} - \mu_n(u_t'Z_t)| \leq \cdots \leq |u_t'Z_{knt} - \mu_n(u_t'Z_t)|$, where $\{Z_{k1t}, \ldots, Z_{knt}\}$ is a permutation of $\{Z_1, \ldots, Z_n\}$. Then there are at least two original points from X^n among $\{Z_{k1t}, \ldots, Z_{kst}\}$ with $s = \lfloor (n+2)/2 \rfloor$. If there are exactly two such points, then $|g(Z_{it}, u_t, Z_t)| \leq 1$ for each of the m contaminating points Z_{it} approaching y_t as $t \to \infty$. If there are at least three such points, say, X_1, X_2, X_3, then it is direct to verify that $\sigma(u_t'Z_t) \geq \frac{1}{2}\max_{1 \leq j \leq k \leq 3}|u_t'(X_j - X_k)| \geq \frac{1}{2}\min_{1 \leq j_1 \leq j_2 \leq j_3 \leq n}\max_{j,k \in \{j_1, j_2, j_3\}}|u_t'(X_j - X_k)|$, and, therefore,

$$|g(Z_{it}, u_t, Z_t)| \leq \sup_u \frac{4 \sup_i |u'y - u'X_i|}{\min_{1 \leq j_1 \leq j_2 \leq j_3 \leq n}\max_{j,k \in \{j_1, j_2, j_3\}}|u'(X_j - X_k)|},$$

bounded by a finite constant (depending only on X^n) for each of the m contaminating points Z_{it} approaching y_t as $t \to \infty$. In either case, $O(Z_{it}, Z_t)$ are bounded uniformly over t and Z_t for m contaminating points, contradicting the assertion that $W(PD(Z_{it}, Z_t)) \to 0$ as $t \to \infty$ for $1 \le i \le n$.

Now we consider the case with $d = 2$ and $k > 1$.

(i) $m = \lfloor (n - k)/2 \rfloor + 1$ *points are sufficient for breakdown of* L_n^k. Move m points in X^n to the same site y far away from the original convex hull of X^n and let $\|y\|$ approach ∞. Since $n - m < (n + k)/2$, $|\sigma_n(u'Z)|$ is unbounded as $\|y\| \to \infty$ for any u not perpendicular to y. Consequently, we have $O(Z_i, Z) \le 2$ as $\|y\| \to \infty$ for all points $Z_i = y$. Therefore, $\|L_n^k(Z)\| \to \infty$ as $\|y\| \to \infty$.

(ii) $m = \lfloor (n - k)/2 \rfloor$ *points are not sufficient for breakdown of* L_n^k. Since $m < \lfloor (n + 1)/2 \rfloor$ and $n - m > (n + k - 1)/2$, it is clear that $|\mu_n(u'Z)|$ and $\sigma_n(u'Z)$ are uniformly bounded above for any unit vector u and contaminated data set Z. Similar to the case with $k = 1$, we see that the numerator of L_n^k is bounded from above. It remains to show that the denominator of L_n^k is uniformly bounded away from 0 for any Z. Otherwise, there is a sequence of contaminating data sets $\{Z_t\}$ with $Z_t = \{Z_{1t}, \ldots, Z_{nt}\}$ such that $W(PD(Z_{it}, Z_t)) \to 0$ as $t \to \infty$ for $1 \le i \le n$. Note that $|u'Z_i - \mu(u'Z)|$ is uniformly bounded for at least $n - m$ uncontaminated points; thus, to have $W(PD(Z_{it}, Z_t)) \to 0$, $\sigma_n(u'Z_t)$ must approach 0 for some u as $t \to \infty$. This happens only if $\lfloor n + k + 1 \rfloor / 2$ points in Z_t approach (or lie on) a line l_t. When $k \ge d + 1$, it is impossible for $\lfloor n + k + 1 \rfloor / 2$ points in Z_t to approach (or lie on) a line l_t as $m + d < \lfloor (n + k)/2 \rfloor$. When $k = d$ and n is odd, $m + d < \lfloor (n + k + 1)/2 \rfloor$. Thus, we need to focus only on the case that $k = d$ and n is even. Applying the same arguments for the case with $k = 1$, we can reach a contradiction to the assertion that $W(PD(Z_{it}, Z_t)) \to 0$ as $t \to \infty$ for $1 \le i \le n$.

Now we move to the case with $d > 2$ and $k \le d$.

(i) $m = \lfloor (n - 2d + 1 + k)/2 \rfloor$ *points are sufficient for breakdown of* L_n^k. Let H_1 and H_2 be two hyperplanes, each containing d points in X^n and intersecting at a line l containing no points in X^n; these two planes exist since $n \ge 2d$ and X^n is in general position. Move m points of X^n not on H_1 and H_2 to a point y on the line l. Since $d + m > (n + k - 1)/2$, we can choose two unit vectors perpendicular to the two hyperplanes, respectively. It then follows that $O(Z_i, Z) = \infty$ for all Z_i not equal to y. Hence, $L_n^k(Z) = y$, and a breakdown occurs as $\|y\| \to \infty$.

(ii) $m = \lfloor (n - 2d + 1 + k)/2 \rfloor - 1$ *points are not sufficient for breakdown of* L_n^k. When $m < \lfloor (n + 1)/2 \rfloor$ and $n - m > (n + k - 1)/2$, we have $|\mu_n(u'Z)|$ and $\sigma_n(u'Z)$ uniformly bounded above for any unit vector u and contaminated data set Z. By similar arguments to those used above for $d = 2$, we can show that the numerator of L_n^k is uniformly bounded above for any contaminated data set Z but the denominator is uniformly bounded away from 0.

The case with $d > 2$ and $k > d$ can be handled using the arguments made above, so we omit the details. $\quad \square$

REFERENCES

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542.

BAI, Z. D. and HE, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *Ann. Statist.* **27** 1616–1637.

CHAKRABORTY, B., CHAUDHURI, P. and OJA, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statist. Sinica* **8** 767–784.

CUI, H. and TIAN, Y. (1994). Estimation of the projection absolute median deviation and its application (in Chinese). *J. Systems Sci. Math. Sci.* **14** 63–72.

DAVIES, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Qualifying paper, Dept. Statistics, Harvard Univ.

DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

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

DÜMBGEN, L. (1992). Limit theorem for the simplicial depth. *Statist. Probab. Lett.* **14** 119–128.

GATHER, U. and HILKER, T. (1997). A note on Tyler's modification of the MAD for the Stahel–Donoho estimator. *Ann. Statist.* **25** 2024–2026.

HE, X. (1991). A local breakdown property of robust tests in linear regression. *J. Multivariate Anal.* **38** 294–305.

HE, X. and WANG, G. (1996). Cross-checking using the minimum volume ellipsoid estimator. *Statist. Sinica* **6** 367–374.

HE, X. and WANG, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495–504.

KENT, J. T. and TYLER, D. E. (1996). Constrained M-estimation multivariate location and scatter. *Ann. Statist.* **24** 1346–1370.

KIM, J. and HWANG, J. (2001). Asymptotic properties of location estimators based on projection depth. Preprint.

LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414.

LIU, R. Y. (1992). Data depth and multivariate rank tests. In L_1-*Statistical Analysis and Related Methods* (Y. Dodge, ed.) 279–294. North-Holland, Amsterdam.

LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260.

LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. Statist.* **27** 783–858.

MARONNA, R. A. and YOHAI, V. J. (1995). The behavior of the Stahel–Donoho robust multivariate estimator. *J. Amer. Statist. Assoc.* **90** 330–341.

MASSÉ, J. C. (1999). Asymptotics for the Tukey depth. Preprint.

MOSTELLER, F. and TURKEY, J. W. (1977). *Data Analysis and Regression*: *A Second Course in Statistics*. Addison Wesley, Reading, MA.

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

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

ROUSSEEUW, P. J. and HUBERT, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.* **94** 388–433.

ROUSSEEUW, P. J. and RUTS, I. (1998). Constructing the bivariate Tukey median. *Statist. Sinica* **8** 828–839.

STAHEL, W. A. (1981). Robust estimation: Infinitesimal optimality and covariance matrix estimators. Ph.D. thesis, ETH, Zurich (in German).

TUKEY, J. W. (1975). Mathematics and the picturing data. In *Proceedings of the International Congress of Mathematicians* **2** 523–531. Canadian Mathematical Congress, Montreal.

TYLER, D. E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.* **22** 1024–1044.

ZHANG, J. (1991). Asymptotic theories for the robust PP estimators of the principal components and dispersion matrix. III. Bootstrap confidence sets, bootstrap tests. *Systems Sci. Math. Sci.* **4** 289–301.

ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490.

ZUO, Y. and SERFLING, R. (2000a). General notions of statistical depth function. *Ann. Statist.* **28** 461–482.

ZUO, Y. and SERFLING, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.* **28** 483–499.

Y. ZUO
DEPARTMENT OF STATISTICS
  AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824
USA
E-MAIL: zuo@msu.edu

H. CUI
DEPARTMENT OF MATHEMATICS
BEIJING NORMAL UNIVERSITY
BEIJING 100875
PEOPLE'S REPUBLIC OF CHINA
E-MAIL: cuihj@hotmail.com

X. HE
DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
725 S. WRIGHT
CHAMPAIGN, ILLINOIS 61820
USA
E-MAIL: he@stat.uiuc.edu