The Annals of Statistics
1999, Vol. 27, No. 5, 1616–1637

ASYMPTOTIC DISTRIBUTIONS OF THE MAXIMAL DEPTH ESTIMATORS FOR REGRESSION AND MULTIVARIATE LOCATION

BY ZHI-DONG BAI[1] AND XUMING HE[2]

National University of Singapore and University of Illinois

We derive the asymptotic distribution of the maximal depth regression estimator recently proposed in Rousseeuw and Hubert. The estimator is obtained by maximizing a projection-based depth and the limiting distribution is characterized through a max–min operation of a continuous process. The same techniques can be used to obtain the limiting distribution of some other depth estimators including Tukey's deepest point based on half-space depth. Results for the special case of two-dimensional problems have been available, but the earlier arguments have relied on some special geometric properties in the low-dimensional space. This paper completes the extension to higher dimensions for both regression and multivariate location models.

1. Introduction. Multivariate ranking and depth have been of interest to statisticians for quite some time. The notion of depth plays an important role in data exploration, ranking, and robust estimation; see Liu, Parelius and Singh (1999) for some recent advances. The location depth of Tukey (1975) is the basis for a multivariate median; see Donoho and Gasko (1992). Recently, Rousseeuw and Hubert (1999) introduced a notion of depth in the linear regression setting. Both measures of depth are multivariate in nature and defined as the minimum of an appropriate univariate depth over all directions of projection. The maximal depth estimator is then obtained through a max–min operation which complicates the derivation of its asymptotic distribution. The present paper focuses on the asymptotics of maximal depth estimators.

First, we recall the definition of regression depth. Consider a regression model in the form of $y_i = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta}_1 + e_i$ where $\mathbf{x}_i \in R^{p-1}$, $\boldsymbol{\beta}' = (\beta_0, \boldsymbol{\beta}_1') \in R^p$ and e_i are regression errors. A regression fit $\boldsymbol{\beta}$ is said to be a nonfit to the given data $\mathbf{Z}_n = \{(\mathbf{x}_i, y_i),\ i = 1, 2, \ldots, n\}$ if and only if there exists an affine hyperplane V in the design space such that no $\mathbf{x}_i$ belongs to V and such that the residuals $r_i > 0$ for all $\mathbf{x}_i$ in one of its open half-spaces and $r_i < 0$ for all $\mathbf{x}_i$ in the other open half-space. Then, the regression depth $rdepth(\boldsymbol{\beta}, \mathbf{Z}_n)$ is the smallest number of observations that need to be removed (of whose residuals need to change sign) to make $\boldsymbol{\beta}$ a nonfit. To put it into mathematical

Received August 1998; revised August 1999.
[1]Supported in part by National University of Singapore Grant RP397212.
[2]Supported in part by NSF Grant SBR 96-17278 and by the Wavelets Strategic Research Program (SWRP) funded by the National Science and Technology Board and the Ministry of Education of Singapore under Grant RP960 601/A.
AMS 1991 *subject classifications*. Primary 62G35, 62F12; secondary 62J05, 62H12.
Key words and phrases. Asymptotic distribution, consistency, estimator, median, multivariate location, regression depth, robustness.

formulation, let $\mathbf{w}_i' = (1, \mathbf{x}_i')$, $r_i(\boldsymbol{\beta}) = y_i - \mathbf{w}_i'\boldsymbol{\beta}$. Following Rousseeuw and Hubert (1999), we define

$$
\begin{aligned}
&rdepth(\boldsymbol{\beta}, \mathbf{Z}_n) \\
&= \inf_{\substack{\|\mathbf{u}\|=1 \\ v \in R}} \min \left\{ \sum_{i=1}^{n} I(r_i(\boldsymbol{\beta})(\mathbf{u}'\mathbf{x}_i - v) > 0), \sum_{i=1}^{n} I(r_i(\boldsymbol{\beta})(\mathbf{u}'\mathbf{x}_i - v) < 0) \right\}.
\end{aligned} \tag{1.1}
$$

The maximal depth estimate $\hat{\boldsymbol{\beta}}_n$ maximizes $rdepth(\boldsymbol{\beta}, \mathbf{Z}_n)$ over $\boldsymbol{\beta} \in R^p$. For convenience, we reformulate the objective function (1.1) as follows. Denote $S^p = \{\boldsymbol{\gamma} \in R^p,\ \|\boldsymbol{\gamma}\| = 1\}$ as the unit sphere in R^p. Then it is easy to show that

$$
rdepth(\boldsymbol{\beta}, \mathbf{Z}_n) = \frac{n}{2} + \frac{1}{2} \inf_{\boldsymbol{\gamma} \in S^p} \sum_{i=1}^{n} \operatorname{sgn}(y_i - \mathbf{w}_i'\boldsymbol{\beta}) \operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}), \tag{1.2}
$$

where $\operatorname{sgn}(x)$ is the sign of x. In the rest of the paper, we consider the problem of

$$
\sup_{\boldsymbol{\beta} \in R^p} \inf_{\boldsymbol{\gamma} \in S^p} \sum_{i=1}^{n} \operatorname{sgn}(y_i - \mathbf{w}_i'\boldsymbol{\beta}) \operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}). \tag{1.3}
$$

Note that the deepest point based on Tukey depth for multivariate data has a similar formulation. Given n observations $\mathbf{X}_n = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ in R^p, the deepest point $\hat{\boldsymbol{\theta}}_n$ solves

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta} \in R^p} \inf_{\boldsymbol{\gamma} \in S^p} \sum_{i=1}^{n} I(\boldsymbol{\gamma}'(\mathbf{x}_i - \boldsymbol{\theta}) > 0) \\
&= \frac{1}{n} + \frac{1}{2} \sup_{\boldsymbol{\theta} \in R^p} \inf_{\boldsymbol{\gamma} \in S^p} \sum_{i=1}^{n} \operatorname{sgn}(\boldsymbol{\gamma}'(\mathbf{x}_i - \boldsymbol{\theta})).
\end{aligned} \tag{1.4}
$$

Both (1.3) and (1.4) involve a max–min operation applied to a sum of data-dependent functions. Common techniques can be used to derive the asymptotic distributions of these estimators. In fact, the asymptotic distributions of both estimators have been derived for the case of $p = 2$ by He and Portnoy (1998) and Nolan (1999), respectively. The limiting distribution can be characterized by the random variable that solves $\max_{\boldsymbol{\beta}} \min_{\boldsymbol{\gamma} \in S^p}(W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta})$ for some Gaussian process W and smooth function $\boldsymbol{\mu}$. The difficulty in treating the higher-dimensional case lies mainly in proving uniqueness of the solution $\boldsymbol{\beta}$ to the above max–min problem. Both works cited above used arguments based on two-dimensional geometry and direct extensions to higher dimensions appear difficult. See Nolan (1999) for an explicit account of the difference between the two-dimensional and the higher-dimensional structures.

Limiting distributions as characterized by an arg-max or arg-min functional are not that uncommon in the statistics literature. A good recent reference is Kim and Pollard (1990). The problem we are concerned with here is complicated by the additional optimization over $\boldsymbol{\gamma} \in S^p$. This type of limiting distribution comes up naturally from the use of projections. We focus on the maximal depth regression and the deepest point (as a location estimate) in the present paper due to their importance as a natural generalization of *median* for regression and multivariate data. Both estimators enjoy some of the desirable properties that we expect from the median. For example, they are affine equivariant, have positive breakdown point (higher than that of an M-estimator), and are root-n consistent to their population counterparts. For confidence bands based on depth, see He (1999).

In Section 2, we show that the maximal depth regression estimate is consistent for the conditional median of y given $\mathbf{x}$ if it is linear. The conditional distribution of y given $\mathbf{x}$ may vary with $\mathbf{x}$. This property is shared with the least absolute deviation regression (LAD), commonly interpreted as the median regression; see Koenker and Bassett (1978). Because the breakdown robustness of the LAD is design-dependent [cf. He, Koenker and Portnoy (1990)], the maximal depth regression has the advantage of being robust against data contamination at the leverage points.

In Section 3, we derive the asymptotic distribution of the maximal depth estimate. In line with most other published results on the asymptotic distributions of regression estimators and to avoid being overshadowed by notational and technical complexity, we work with a more restrictive regression model with i.i.d. errors in this section. An almost sure LIL-type result for the estimator is also provided in this section. We then present the limiting distribution of the deepest point for multivariate data in Section 4, extending the work of Nolan (1999). The Appendix provides all the proofs needed in the paper. In particular, we provide a means to establish the uniqueness of solution to a max–min problem that arises from the projection-based depth in regression as well as multivariate location models. For computation of the regression and location depth, we refer to Rousseeuw and Struyf (1998).

2. Consistency of maximal depth regression. We assume that the conditional median of y given $\mathbf{x}$ is linear, that is, there exists $\boldsymbol{\beta}^* \in R^p$ such that

$$\text{Median}(y|\mathbf{x}) = \mathbf{w}'\boldsymbol{\beta}^*, \tag{2.1}$$

where $\mathbf{w}' = (1, \mathbf{x}')$. For a set of n design points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, independent observations of y_i are drawn from the conditional distributions of y given $\mathbf{x} = \mathbf{x}_i$.

If the conditional distribution of $y - \mathbf{w}'\boldsymbol{\beta}^*$ given $\mathbf{x}$ is the same for all $\mathbf{x}$, then the data can be modeled by the usual regression with i.i.d. errors. The above framework includes the case of random designs so that the data $(\mathbf{x}_i, y_i)$ come from the joint distribution of $(\mathbf{x}, y)$ as well as nonstochastic designs.

Since the maximal depth estimate $\hat{\boldsymbol{\beta}}_n$ is regression invariant, we assume without loss of generality that $\boldsymbol{\beta}^* = 0$ so that the conditional median of y is zero. To show that $\hat{\boldsymbol{\beta}}_n \to 0$, conditions on the design points and the error distributions are needed. For this purpose, let F_i be the conditional c.d.f. of y

given $\mathbf{x} = \mathbf{x}_i$. Also define for any $c > 0$,

$$Q_n(c) = \inf_{\mathbf{1} \in S^p} n^{-1} \sum_{i=1}^{n} I(|\mathbf{w}_i' \mathbf{1}| > c). \tag{2.2}$$

We now state our assumptions as follows. If the design points are random, then all the statements involving $\mathbf{w}_i$ are meant to be in the almost sure sense:

(D1) For some $b < \infty$, $\max_{1 \le i \le n} \|\mathbf{w}_i\| = O(n^b)$.
(D2) For any sequence $a_n \to 0$, $\lim_{n\to\infty} Q_n(a_n) = 1$.
(D3) For some $A < \infty$, $n^{-1} \sum_{i=1}^{n} \{1 - F_i(n^A) + F_i(-n^A)\} \to 0$ and $\max_{i \le n} \sup_x (F_i(x + n^{-A}) - F_i(x - n^{-A})) \to 0$ as $n \to \infty$.
(D4) For any $r > 0$, $\eta(r) = \inf_{i \ge 1} \min\{|1 - 2F_i(r)|, |1 - 2F_i(-r)|\} > 0$.

Condition (D2) is to avoid the degenerate case for the design points. This condition is satisfied if $\{\mathbf{x}_i\}$ is a random sample from a continuous multivariate distribution. Condition (D3) includes a weak requirement of the average tail thickness and a weak uniform continuity of all the conditional distribution functions, but (D4) requires that the error mass around the median is not too thin, which is satisfied if each F_i has a density with a common positive lower bound around the median. The following lemma is the basis for our consistency result.

LEMMA 2.1. *Under conditions* (D1)–(D3), *we have with probability* 1,

$$\sup_{\boldsymbol{\beta} \in R^p,\, \boldsymbol{\gamma} \in S^p} \left| n^{-1} \sum_{i=1}^{n} \{\operatorname{sgn}(y_i - \mathbf{w}_i' \boldsymbol{\beta}) \operatorname{sgn}(\mathbf{w}_i' \boldsymbol{\gamma}) - E \operatorname{sgn}(y_i - \mathbf{w}_i' \boldsymbol{\beta}) \operatorname{sgn}(\mathbf{w}_i' \boldsymbol{\gamma})\} \right| \to 0,$$

as $n \to \infty$.

Lemma 2.1 is a standard uniform approximation result except that the approximation is now over the whole space for $\boldsymbol{\beta}$. This is made possible by the fact that when $\|\boldsymbol{\beta}\|$ is large the function $\operatorname{sgn}(y_i - \mathbf{w}_i' \boldsymbol{\beta})$ does not change much. A proof of Lemma 2.1 for the possibly nonstochastic designs $\mathbf{w}_i$ is given in the Appendix.

By (D2) and (D4), for any given $c > 0$, there is a constant $r > 0$ such that $Q_n(r/c) > 1/2$ for sufficiently large n. Consequently, we have

$$\begin{aligned} &\inf_{\|\boldsymbol{\beta}\| \ge c} n^{-1} \sum_{i=1}^{n} \left|1 - 2F_i(\mathbf{w}_i' \boldsymbol{\beta})\right| \\ &\quad \ge \inf_{\mathbf{1} \in S^p} n^{-1} \sum_{i=1}^{n} \min\{|1 - 2F_i(r)|, |1 - 2F_i(-r)|\} I(|\mathbf{w}_i' \mathbf{1}| > r/c) \\ &\quad \ge \eta(r) Q_n(r/c) > \tfrac{1}{2} \eta(r) \end{aligned} \tag{2.3}$$

and with $\boldsymbol{\beta}^* = 0$ we have

$$n^{-1}\sum_{i=1}^{n} E\,\mathrm{sgn}(y_i - \mathbf{w}_i'\boldsymbol{\beta})\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\beta}) = -n^{-1}\sum_{i=1}^{n}\left|1 - 2F_i(\mathbf{w}_i'\boldsymbol{\beta})\right| < -\tfrac{1}{2}\eta(r)$$

for sufficiently large n. Thus, $\inf_{\boldsymbol{\gamma}\in S^p} n^{-1}\sum_{i=1}^{n} E\{\mathrm{sgn}(y_i - \mathbf{w}_i'\boldsymbol{\beta})\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\} < 0$.

On the other hand, $E\{\mathrm{sgn}(y_i)\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\} = 0$ for any $\boldsymbol{\gamma} \in S^p$, so

$$\inf_{\boldsymbol{\gamma}\in S^p}\{n^{-1}\sum_{i=1}^{n} E\,\mathrm{sgn}(y_i)\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\} = 0.$$

Therefore, the maximal depth estimator has to be in the ball $\{\boldsymbol{\beta}: \|\boldsymbol{\beta}\| < c\}$. The consistency of $\hat{\boldsymbol{\beta}}_n$ follows from the fact that c can be arbitrarily small. We state the result formally as follows.

THEOREM 2.1. *Under conditions* (D1)–(D4), *the maximal depth regression estimate* $\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}^*$, *almost surely*.

Conditions (D1)–(D4) are sufficient but not necessary. It helps to note that the maximal depth regression estimator is consistent for the conditional median of y given $\mathbf{x}$ whenever the median is linear in $\mathbf{x}$. This is a property shared with L_1 regression but not other M-estimators. The limit of other M-estimators can only be identified with some additional information on the conditional distributions such as symmetry.

3. Limiting distribution of the maximal depth regression. In this section we derive the asymptotic distribution of the maximal depth estimator for the usual regression model

$$y_i = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}_i + e_i, \qquad i = 1, 2, \ldots, n,$$

where $\mathbf{x}_i$ is a random sample from a distribution in R^{p-1} with finite second moments, e_i's are independent of each other and of $\mathbf{x}_i$'s with a common distribution function F and density function f whose median is zero. We continue to use the same notation as in Section 2.

The following Lemma 3.1 is important for finding the limiting distribution of $\hat{\boldsymbol{\beta}}_n$. First, we itemize our assumptions for easy reference.

(C1) $E\|\mathbf{x}\|^2 \le B$ and $\sup_{\mathbf{1}\in S^p} P(|\mathbf{w}'\mathbf{1}| \le a\|\mathbf{w}\|) \le Ba^\delta$ for some $\delta \in (0, 2]$ and $B < \infty$.
(C2) $|F(x + r) - F(x)| \le B|r|^\delta$ for any x and r.
(C3) As $r \to 0$, $F(r) - F(0) = f(0)r + o(r)$ with $f(0) > 0$.
(C4) $E\{\mathrm{sgn}(\boldsymbol{\gamma}_1'\mathbf{w}\mathbf{w}'\boldsymbol{\gamma}_2)\}$ is continuous in $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in S^p$, and $E\{\mathbf{w}\,\mathrm{sgn}(\mathbf{w}'\boldsymbol{\gamma})\}$ is continuously differentiable in $\boldsymbol{\gamma} \in S^p$.

In typical cases, the constant $\delta = 1$ in (C1) and (C2).

REMARK 3.1. It is clear that conditions (D2) and (D4) are implied by (C1) and (C3). For independent and identically distributed errors whose distribution F has no positive mass at its median, condition (D3) is trivial. Condition

(D1) is true if $E\|\mathbf{x}\|^{1/b} < \infty$. Thus, the maximal depth estimator is consistent under conditions (C1)–(C3).

REMARK 3.2. If $\mathbf{x}_i$'s are not random or the e_i's may have different distributions F_i, the results of this section remain true if the above four conditions are replaced by:

(C1′) $n^{-1}\sum_{i=1}^n \mathbf{w}_i\mathbf{w}_i' \to \mathbf{A}$, a positive definite matrix, as $n \to \infty$, and $\sup_{\mathbf{1}\in S^p} n^{-1}\sum_{i=1}^n I(|\mathbf{w}_i'\mathbf{1}| \le a\|\mathbf{w}_i\|) \le Ba^\delta$ for some $\delta \in (0, 2]$ and $B < \infty$.

(C2′) For any x and r, $n^{-1}\sum_{i=1}^n |F_i(x+r) - F_i(x)| \le B|r|^\delta$.

(C3′) As $r \to 0$, $\max_{i\le n} |F_i(r) - F_i(0) - f_i(0)r| = o(r)$, as $r \to 0$, and $\bar{f} = \inf_n \bar{f}_n(0) = \inf_n n^{-1}\sum_{i=1}^n f_i(0) > 0$.

(C4′) The limit of $n^{-1}\sum_{i=1}^n \mathbf{w}_i \operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})$ (as $n \to \infty$) exists and is continuously differentiable in $\boldsymbol{\gamma} \in S^p$, and the limit of $n^{-1}\sum_{i=1}^n \operatorname{sgn}(\boldsymbol{\gamma}_1'\mathbf{w}_i)\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_2)$ exists uniformly and is continuous in $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in S^p$.

The proofs for our results in this section under conditions (C1′)–(C4′) are almost the same as those under (C1)–(C4) with averaging in place of expectations of $\mathbf{w}_i$. Let

$$S_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^n \{(\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}) - \operatorname{sgn}(e_i))\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\} - E[(\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}) - \operatorname{sgn}(e_i))\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})]. \tag{3.1}$$

In this paper, we use $a_n \ll b_n \ll c_n$ to mean $a_n/b_n \to 0$ and $b_n/c_n \to 0$.

LEMMA 3.1. *If* (C1) *and* (C2) *hold, then for any constant* $\nu > 0$ *and any bounded sequence* $\Delta_n \gg n^{-1/(\delta+2v(1+\delta))}$, *we have*

$$\sup_{\|\boldsymbol{\beta}\|\le\Delta_n,\, \boldsymbol{\gamma}\in S^p} |S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})| = O_p(n^{1/2}\Delta_n^{\delta/2-\nu}).$$

If we further assume $\Delta_n \to 0$ *slowly or regularly in the sense that there exist a constant* $\alpha > 0$ *and a function* $L(x)$ *such that* $\Delta_n = n^{-\alpha}L(n)$ *and* $L(bx)/L(x) \to 1$ *as* $x \to \infty$ *for any* $b > 0$, *then*

$$\limsup_{n\to\infty} \sup_{\|\boldsymbol{\beta}\|\le\Delta_n,\, \boldsymbol{\gamma}\in S^p} |S_n(\boldsymbol{\beta}, \boldsymbol{\gamma})|/(2n\Delta_n^\delta \log\log n)^{1/2} \le 1 \quad a.s.$$

In the Appendix, we actually prove a more general lemma in the form of an exponential inequality. This is often useful for asymptotic analyses in statistics. General results of this type may also be found in Pollard [(1984), page 144]. The following lemma allows for nonrandom designs as in He and Shao (1996), but is proved using a different chaining argument.

LEMMA 3.2. *Suppose that* $\Delta_n > 0$ *is a sequence of constants and* D *is a compact set in* R^p. *For each* $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ *with* $\|\boldsymbol{\beta}\| \le \Delta_n$ *and* $\boldsymbol{\gamma} \in D$, $\{W_1(\boldsymbol{\beta}, \boldsymbol{\gamma}), W_2(\boldsymbol{\beta}, \boldsymbol{\gamma}), \ldots, W_n(\boldsymbol{\beta}, \boldsymbol{\gamma})\}$ *is a sequence of independent random variables of mean zero satisfying*:

(L1) *For some constants* $\delta > 0$ *and* $C_1 > 0$,

$$\frac{1}{n}\sum_{i=1}^{n} E|W_i(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1) - W_i(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2)|^2 \le C_1[\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| + \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|]^{\delta}.$$

(L2) *For some constant* $C_2 > 0$,

$$\big|W_i(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1) - W_i(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2)\big| \le C_2 \quad \text{if } \|\boldsymbol{\beta}_j\| \le \Delta_n \text{ and } \boldsymbol{\gamma}_j \in D, \ \ j = 1, 2.$$

(L3) *For some constant* C_3 *and for any* $d > 0$, $\|\boldsymbol{\beta}_1\| \le \Delta_n$ *and* $\boldsymbol{\gamma}_1 \in D$,

$$\frac{1}{n}\sum_{i=1}^{n} E \sup_{\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}\| + \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}\| \le d} \big|W_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) - W_i(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1)\big|^2 \le C_3 d^{\delta}.$$

Then we have the following results:

(i) *If* $\Delta_n \to 0$ *and* $\Delta_n^{\delta}|\log \Delta_n| \ll \varepsilon_n \ll \sqrt{n}\Delta_n^{\delta(1+\nu)}$ *for some* $\nu \in (0, 1)$, *then, for any* $a > 2$, *there exists* $C_a < \infty$, *such that*

$$P\left(\sup_{\|\boldsymbol{\beta}\| \le \Delta_n,\, \boldsymbol{\gamma} \in D} \max_{m \le n} \left|\sum_{i=1}^{m} (W_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) - W_i(0, \boldsymbol{\gamma}))\right| \ge \sqrt{n}\varepsilon_n\right) \le C_a \exp\{-(aC_1\Delta_n^{\delta})^{-1}\varepsilon_n^2\}. \tag{3.2}$$

(ii) *If* $\log(2 + \Delta_n) \ll \varepsilon_n^2 \ll n$, *then for any* $a > 2$, *there exists* $C_a < \infty$ *such that*

$$P\left(\sup_{\|\boldsymbol{\beta}\| \le \Delta_n,\, \boldsymbol{\gamma} \in D} \max_{m \le n} \left|\sum_{i=1}^{m} (W_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) - W_i(0, \boldsymbol{\gamma}))\right| \ge \sqrt{n}\varepsilon_n\right) \le C_a \exp\{-a^{-1}(\varepsilon_n/C_2)^2\}. \tag{3.3}$$

(iii) *If* $\varepsilon_n = c\sqrt{n}$ *for some constant* $c > 0$ *and* $|\log \Delta_n| = o(n)$, *then* (3.3) *continues to hold for some constant* $a \ge 12$ *even when* (L1) *and* (L3) *are replaced by one weaker condition* (L3$'$) *given below.*

(L3$'$) *There is a constant* $B > 0$ *such that*

$$n^{-1}\sum_{i=1}^{n} E \sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_1\| + \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_1\| \le n^{-B}} \big|W_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) - W_i(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1)\big| \to 0. \tag{3.4}$$

Now back to the maximum depth regression. We first show that $\hat{\boldsymbol{\beta}}_n = O_p(n^{-1/2})$; that is, for any sequence $\zeta_n \to \infty$, we shall show that

$$P(\|\hat{\boldsymbol{\beta}}_n\| > \zeta_n/\sqrt{n}) \to 0. \tag{3.5}$$

We only need to consider the case with $\zeta_n/\sqrt{n} \to 0$ given the consistency of $\hat{\boldsymbol{\beta}}_n$. Note that for any $c > 0$,

$$\sup_{\zeta_n n^{-1/2} \le \|\boldsymbol{\beta}\| \le 1} \min_{\boldsymbol{\gamma} \in S^p} n^{-1/2} \sum_{i=1}^n E[\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}) - \operatorname{sgn}(e_i)]\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})$$
$$\le \sup_{\zeta_n n^{-1/2} \le \|\boldsymbol{\beta}\| \le 1} -2n^{-1/2} \sum_{i=1}^n E|F_i(\mathbf{w}_i'\boldsymbol{\beta}) - F_i(0)| \le -cf(0)\zeta_n n^{-1} N_n,$$

where

$$N_n = \inf_{\mathbf{1} \in S^p} \sum_{i=1}^n P(|\mathbf{w}_i'\mathbf{1}| \ge c),$$

and we have used the fact that $|\mathbf{w}_i'\boldsymbol{\beta}|/\|\boldsymbol{\beta}\| \ge c$ implies, by condition (C3), $|F(\mathbf{w}_i'\boldsymbol{\beta}) - F(0)| \ge |F(c\zeta_n/\sqrt{n}) - F(0)| \ge \frac{1}{2}c\zeta_n f(0)/\sqrt{n}$. By condition (C1) and the fact that $\|\mathbf{w}_i\| \ge 1$, we have

$$n - N_n = \sup_{\mathbf{1} \in S^p} \sum_{i=1}^n P(|\mathbf{w}_i'\mathbf{1}| < c)$$
$$\le \sup_{\mathbf{1} \in S^p} \sum_{i=1}^n P(|\mathbf{w}_i'\mathbf{1}| < c\|\mathbf{w}_i\|) \le Bnc^{\delta}.$$

Therefore, by choosing c small enough so that $Bc^{\delta} < 1/2$, we obtain

$$\sup_{\zeta_n n^{-1/2} \le \|\boldsymbol{\beta}\| \le 1} \min_{\boldsymbol{\gamma} \in S^p} n^{-1/2} \sum_{i=1}^n E[\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}) - \operatorname{sgn}(e_i)]\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) < -\eta(c)\zeta_n. \tag{3.6}$$

Lemma 3.1 then implies that

$$\sup_{\zeta_n/\sqrt{n} \le \|\boldsymbol{\beta}\| \le 1} \min_{\boldsymbol{\gamma} \in S^p} n^{-1/2} \sum_{i=1}^n \{\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}) - \operatorname{sgn}(e_i)\}\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})$$
$$< -\eta(c)\zeta_n + o_p(\zeta_n^{1/2}).$$

This, together with Theorem 2.1, proves (3.5).

Now, define $\boldsymbol{\delta} = \sqrt{n}\boldsymbol{\beta}$ and $\hat{\boldsymbol{\delta}}_n = \sqrt{n}\hat{\boldsymbol{\beta}}_n = O_p(1)$. By condition (C1), we have $n^{-1/2}\max_{i \le n} |\hat{\boldsymbol{\delta}}_n'\mathbf{w}_i| = o_p(1)$. Then by condition (C3), we have, for $\|\boldsymbol{\delta}\| \le V$, any large constant,

$$n^{-1/2} \sum_{i=1}^n E\{(\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}) - \operatorname{sgn}(e_i))\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\}$$
$$= -2n^{-1/2} \sum_{i=1}^n E\{(F(n^{-1/2}\mathbf{w}_i'\boldsymbol{\delta}) - F(0))\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\} = \boldsymbol{\mu}'(\boldsymbol{\gamma})\boldsymbol{\delta} + o(1),$$

where $\boldsymbol{\mu}(\boldsymbol{\gamma}) = -2f(0)E\{\mathbf{w}\,\mathrm{sgn}(\mathbf{w}'\boldsymbol{\gamma})\}$. Therefore, by Lemma 3.1, it holds uniformly for $\|\boldsymbol{\delta}\| \le V$ and $\boldsymbol{\gamma} \in S^p$,

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}_n)\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) \\ &\qquad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{sgn}(e_i)\mathrm{sgn}(\mathbf{w}_i\boldsymbol{\gamma}) + \boldsymbol{\mu}'(\boldsymbol{\gamma})\boldsymbol{\delta} + o_p(1). \end{aligned} \tag{3.7}$$

Notice that $n^{-1/2}\sum_{i=1}^{n}\mathrm{sgn}(e_i)\mathrm{sgn}(\mathbf{w}_i\boldsymbol{\gamma})$ converges to a Gaussian process $W(\boldsymbol{\gamma})$ with mean 0 and covariance function $\Lambda(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = E[\mathrm{sgn}(\mathbf{w}'\boldsymbol{\gamma}_1)\mathrm{sgn}(\mathbf{w}'\boldsymbol{\gamma}_2)]$.

Since $\Lambda(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ is continuous in $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$, we may define $W(\boldsymbol{\gamma})$ so that almost all paths are continuous. Also, note that $\Lambda(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ satisfies the Hölder condition of order δ due to conditions (C1) and (C4). It follows from an application of Lemma 3.2 that the sequence of processes $\{n^{-1/2}\sum_{i=1}^{n}\mathrm{sgn}(e_i)\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\}$ in $D(S^p)$-space is tight. Therefore, it converges weakly to $W(\boldsymbol{\gamma})$ with the Skorohod metric in $D(S^p)$-space. Similarly to Theorem 2.7 of Kim and Pollard (1990), it follows that the limiting distribution of $\hat{\boldsymbol{\delta}}_n$ is characterized by the variable $\boldsymbol{\beta}$ that solves

$$\max_{\boldsymbol{\beta}}\min_{\boldsymbol{\gamma}\in S^p}(W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta}), \tag{3.8}$$

where

$$\boldsymbol{\mu}(\boldsymbol{\gamma}) = -2f(0)E\{\mathrm{sgn}(\mathbf{w}'\boldsymbol{\gamma})\mathbf{w}\}, \tag{3.9}$$

provided that the solution $\boldsymbol{\beta}$ to (3.8) is *unique*. Establishing this uniqueness property can be viewed as the most difficult part of the work we are undertaking in the present paper.

The following lemma, stated for each sample path, plays a fundamental role in the paper.

Suppose that $\boldsymbol{\mu}(\boldsymbol{\gamma})$ is a continuously differentiable function defined on S^p. Extend $\boldsymbol{\mu}(\boldsymbol{\gamma})$ to $R^p - \{0\}$ by $\boldsymbol{\mu}(r\boldsymbol{\gamma}) = \boldsymbol{\mu}(\boldsymbol{\gamma})$ for any $r > 0$ and $\boldsymbol{\gamma} \in S^p$. Let

$$D_{\boldsymbol{\gamma}} = \left.\frac{\partial\boldsymbol{\mu}'(\boldsymbol{\gamma}+\mathbf{1})}{\partial\mathbf{1}}\right|_{\mathbf{1}=0},$$

which is a $p \times p$ matrix. Obviously, this matrix cannot be of full rank.

LEMMA 3.3. *Suppose that $W(\boldsymbol{\gamma})$ is continuous and $\boldsymbol{\mu}(\boldsymbol{\gamma})$ is differentiable on S^p. Under the following conditions* (W1)–(W3), *the solution to* (3.8) *is unique.*

(W1) *For any $\mathbf{1} \in S^p$, the minimum of $\mathbf{1}'\boldsymbol{\mu}(\boldsymbol{\gamma})$ is negative and achieved only at $\boldsymbol{\gamma} = \mathbf{1}$.*

(W2) *There exists at most one direction $\pm\boldsymbol{\alpha} \in S^p$ such that $D_{\boldsymbol{\gamma}}$ is well defined with* rank $p-1$ *and* $(\mathbf{D}_{\boldsymbol{\gamma}})\boldsymbol{\gamma} = 0$ *for all $\boldsymbol{\gamma}$ not parallel to α.*

(W3) *There do not exist $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ such that $W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta} = W(-\boldsymbol{\gamma}) + \boldsymbol{\mu}(-\boldsymbol{\gamma})'\boldsymbol{\beta}$.*

The same proof shows that Lemma 3.3 is true if $\mu(\boldsymbol{\gamma})$ is replaced by $-\mu(\boldsymbol{\gamma})$. It will be shown in the Appendix that $\boldsymbol{\mu}(\boldsymbol{\gamma}) = -2f(0)E\{\text{sgn}(\mathbf{w}'\boldsymbol{\gamma})\mathbf{w}\}$ satisfies (W1)–(W3) if conditions (C1)–(C4) hold. Our main purpose in the paper is to establish the following theorem.

THEOREM 3.1. *Under conditions* (C1)–(C4), $n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ *converges in distribution to the random variable as the solution to*

$$\max_{\boldsymbol{\beta}\in R^p} \min_{\boldsymbol{\gamma}\in S^p} \Big(W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta} \Big),$$

where $\boldsymbol{\mu}(\boldsymbol{\gamma})$ *is given in* (3.9), $W(\boldsymbol{\gamma})$ *is the Gaussian process with mean* 0 *and covariance function* $\text{cov}(W(\boldsymbol{\gamma}_1), W(\boldsymbol{\gamma}_2)) = E\{\text{sgn}(\boldsymbol{\gamma}_1'\mathbf{w}\mathbf{w}'\boldsymbol{\gamma}_2)\}$.

In the case of $p = 2$, the limiting distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ simplifies to that derived in He and Portnoy (1998), even though the two forms look somewhat different. Except for the case of the usual median ($p = 1$) problem, the non-Gaussian limiting distributions given in Theorem 3.1 are typical for projection-based estimators but not convenient for inference. However, some properties of the limiting distributions may be understood; see He (1999) for more details. Tyler (1994) gives another example with the same type of limiting distributions.

Similar arguments to those used in Section 2 plus the second part of Lemma 3.1 allow us to get an almost sure bound on the estimator as follows.

THEOREM 3.2. *Under conditions* (C1) *and* (C2), *we have* $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = O((\log\log n/n)^{1/2})$ *almost surely, provided that* $\inf_{\boldsymbol{\gamma}\in S^p} E|\mathbf{w}'\boldsymbol{\gamma}| > 0$. *If we further assume* (C3), *then*

$$\limsup_{n\to\infty} \sqrt{n/\log\log n}\,|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}| \le \Big(\sqrt{2}f(0) \inf_{\boldsymbol{\gamma}\in S^p} E|\mathbf{w}'\boldsymbol{\gamma}|\Big)^{-1}.$$

4. Asymptotics of the deepest point in R^P. The same techniques used in Section 3 apply to the asymptotic analysis of the deepest point for multivariate data. The result stated in this section completes the work of Nolan (1999). Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample of p-dimensions. The deepest point $\mathbf{T}_n$ is defined as the solution to the max–min problem

$$\sup_{\mathbf{t}} \inf_{\mathbf{u}\in S^p} \sum_{i=1}^n I(\mathbf{u}'(\mathbf{X}_i - \mathbf{t}) > 0). \tag{4.1}$$

We assume that there exists $\boldsymbol{\theta}_0$ as the unique deepest point for the population such that $P(\mathbf{u}'(\mathbf{X} - \boldsymbol{\theta}_0) > 0) = 1/2$ for all $\mathbf{u} \in S^p$. Without loss of generality, assume $\boldsymbol{\theta}_0 = 0$. To get the asymptotic linearization results parallel to those in Section 3, let $P_{\mathbf{u}}$ be the one-dimensional marginal distribution of $\mathbf{u}'\mathbf{X}$, and $p_{\mathbf{u}}$ be its corresponding density function. Nolan (1999) showed that if

(N1) $P_{\mathbf{u}}$ has a unique median at 0, for all $\mathbf{u}$, and

(N2) $P_{\mathbf{u}}$ has a bounded positive density, $p_{\mathbf{u}}$, at 0, and $p_{\mathbf{u}}(x)$ is continuous in $\mathbf{u}$ and x at $x = 0$,

then $n^{1/2}\mathbf{T}_n$ converges to the random variable

$$\operatorname{argmax}_{\mathbf{t}} \min_{\mathbf{u}\in S^p}(Z(\mathbf{u}) - \mathbf{u}'\mathbf{t}p_{\mathbf{u}}(0)), \tag{4.2}$$

where $Z(\mathbf{u})$ is a Gaussian process on $\mathbf{u} \in S^p$ with mean zero and $\mathrm{Cov}[Z(\mathbf{u}), Z(\mathbf{v})] = P(\mathbf{u}'\mathbf{X} > 0, \mathbf{v}'\mathbf{X} > 0) - 1/4$, provided that the solution to (4.2) is unique. In the special case of $p = 2$, a proof is given in Nolan (1999) for the desired uniqueness based on some geometric properties in R^2. We now verify that the conditions of Lemma 3.3 hold so that the limiting distribution (4.2) is established for any dimension p. This is done under a mild assumption:

(N3) $\int \|\dot{f}(\mathbf{x})\|\|\mathbf{x}\|\,d\mathbf{x} < \infty$, where $\dot{f}$ is the gradient of f, the density function of $\mathbf{X}$.

THEOREM 4.1. *For any $p \geq 2$, under conditions* (N1)–(N3), $\sqrt{n}\mathbf{T}_n$ *tends in distribution to the random variable defined by the solution to* (4.2).

PROOF. We use Lemma 3.3 to prove the uniqueness of the solution to the max–min problem (4.2). Let $\boldsymbol{\mu}(\mathbf{u}) = -p_{\mathbf{u}}(0)\mathbf{u}$. We show that the derivative of $\boldsymbol{\mu}(\mathbf{u})$ is $\mathbf{D}_{\mathbf{u}} = -p_{\mathbf{u}}(0)(\mathbf{I} - \mathbf{u}\mathbf{u}') - (\mathbf{u}\mathbf{b}_{\mathbf{u}}')$. To get the directional derivative of $\boldsymbol{\mu}$ along any direction $\mathbf{1}$, we use the product rule. The derivative of $\mathbf{u}$ gives $p_{\mathbf{u}}(0)\mathbf{1}$ and the derivative of $p_{\mathbf{u}}(0)$ gives $-\mathbf{u}\mathbf{u}'\mathbf{1}p_{\mathbf{u}}(0) + (\mathbf{u}\mathbf{b}_{\mathbf{u}}')\mathbf{1}$, where $\mathbf{b}_{\mathbf{u}}$ will be calculated below.

Write $\mathbf{u}_t = (\mathbf{u} + t\mathbf{1})/\|\mathbf{u} + t\mathbf{1}\|$, and consider

$$P(\mathbf{u}_t'\mathbf{X} \leq a) = \int_{\mathbf{u}'\mathbf{x}+t\mathbf{1}'\mathbf{x}\leq a\|\mathbf{u}+t\mathbf{1}\|} f(\mathbf{x})\,d\mathbf{x}.$$

Let $\mathbf{B} = (\mathbf{u}, \mathbf{C})$ be an orthonormal matrix with the first column $\mathbf{u}$. Change the variable $\mathbf{x} = \mathbf{B}\mathbf{y}$ and partition $\mathbf{y}' = (v, \mathbf{z}')$ with $v \in R$. Then the above integral can be written as

$$\int \left[\int_{v\leq(a\|\mathbf{u}+t\mathbf{1}\|-t\mathbf{1}'\mathbf{C}\mathbf{z})/(1+t\mathbf{1}'\mathbf{u})} f(\mathbf{B}\mathbf{y})\,dv\right] d\mathbf{z}.$$

Taking derivative wrt a and evaluating it at $a = 0$ yields

$$p_{\mathbf{u}_t}(0) = \frac{1}{\mathbf{u}_t'\mathbf{u}} \int f\left(-\frac{t\mathbf{1}'\mathbf{C}\mathbf{z}\mathbf{u}}{(\mathbf{u}_t'\mathbf{u})\|\mathbf{u} + t\mathbf{1}\|} + \mathbf{C}\mathbf{z}\right) d\mathbf{z}.$$

The derivative of $1/(\mathbf{u}_t'\mathbf{u})$ wrt t at $t = 0$ is $-\mathbf{u}'\mathbf{1}$. Now taking the derivative of the inside under the integral wrt t at $t = 0$ we get

$$\mathbf{b}_{\mathbf{u}} = -\int [\mathbf{u}'\dot{f}(\mathbf{C}\mathbf{z})](\mathbf{C}\mathbf{z})\,d\mathbf{z}.$$

We have completed the proof of $\mathbf{D}_{\mathbf{u}} = -(\mathbf{I} - \mathbf{u}\mathbf{u}')p_{\mathbf{u}}(0) - \mathbf{u}\mathbf{b}_{\mathbf{u}}'$. The definition of $\mathbf{C}$ implies that $\mathbf{b}_{\mathbf{u}}'\mathbf{u} = 0$, and further that $\mathbf{D}_{\mathbf{u}}\mathbf{u} = 0$. Thus, $\{\mathbf{a}'\mathbf{D}_{\mathbf{u}} : \mathbf{a} \in R^p\} = \{\mathbf{a}'\mathbf{D}_{\mathbf{u}} : \mathbf{a}'\mathbf{u} = 0\} = \{p_{\mathbf{u}}(0)\mathbf{a}' : \mathbf{a}'\mathbf{u} = 0\}$, which means that the rank of $\mathbf{D}_u$ is $p-1$.

Here condition (W2) holds without having to exclude an exceptional direction $\boldsymbol{\alpha}$. The other conditions of Lemma 3.3 hold trivially. We then conclude that the asymptotic distribution for the deepest point estimator holds in any dimension and that the proof of Theorem 4.1 is complete. □

APPENDIX

PROOF OF LEMMA 2.1. We apply Lemma 3.2(iii) here. Under (D1)–(D3), we can verify condition (L3′) by taking $B = \max\{b+1, A\}$. It follows from

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} E \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_1\|+\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\|\le n^{-B}} &\big|\mathrm{sgn}(y_i-\mathbf{w}_i'\boldsymbol{\beta})\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) \\ &\quad - \mathrm{sgn}(y_i-\mathbf{w}_i'\boldsymbol{\beta}_1)\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_1)\big| \\ &\le \frac{2}{n}\sum_{i=1}^{n} E[F_i(\mathbf{w}_i'\boldsymbol{\beta}_1+n^{-A}) - F_i(\mathbf{w}_i'\boldsymbol{\beta}_1-n^{-A})] + I(|\mathbf{w}_i'\boldsymbol{\gamma}_1|<n^{-1}) \\ &\le \frac{2}{n}\sum_{i=1}^{n}\sup_x[F_i(x+n^{-A})-F_i(x-n^{-A})] + 2(1-Q_n(1/n)) = o(1),\end{aligned}$$

where the first inequality uses a fact that $\mathbf{w}_i'\boldsymbol{\gamma}$ and $\mathbf{w}_i'\boldsymbol{\gamma}_1$ have the same sign when $|\mathbf{w}_1\boldsymbol{\gamma}_1| \ge 1/n$ and $|\mathbf{w}_i'\boldsymbol{\gamma}-\mathbf{w}_i'\boldsymbol{\gamma}_1| \le n^{-2}$. Condition (L2) holds automatically.

Assume $\boldsymbol{\beta}^* = 0$ without loss of generality and let

$$H_i(\boldsymbol{\beta},\boldsymbol{\gamma}) = (\mathrm{sgn}(y_i-\mathbf{w}_i'\boldsymbol{\beta}) - \mathrm{sgn}(y_i))\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}). \tag{A.1}$$

By Lemma 3.2(iii) with $\varepsilon = \sqrt{n}$ and $b_n = n^{2A}$ for A given in (D3), we have

$$\sup_{\substack{\|\boldsymbol{\beta}\|\le b_n\\ \boldsymbol{\gamma}\in S^p}}\left|\sum_{i=1}^{n}\{H_i(\boldsymbol{\beta},\boldsymbol{\gamma}) - EH_i(\boldsymbol{\beta},\boldsymbol{\gamma})\}\right| = o(n).$$

To complete the proof, it remains to show that

$$\sup_{\substack{\|\boldsymbol{\beta}\|>b_n\\ \boldsymbol{\gamma}\in S^p}}\left|\sum_{i=1}^{n}\{\mathrm{sgn}(y_i-\mathbf{w}_i'\boldsymbol{\beta})\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) - E\,\mathrm{sgn}(y_i-\mathbf{w}_i'\boldsymbol{\beta})\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\}\right| = o(n),$$

which follows from

$$\sup_{\substack{\|\boldsymbol{\beta}\|>b_n\\ \boldsymbol{\gamma}\in S^p}}\left|\sum_{i=1}^{n}\{(\mathrm{sgn}(y_i-\mathbf{w}_i'\boldsymbol{\beta}) - \mathrm{sgn}(-\mathbf{w}_i'\boldsymbol{\beta}))\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\}\right| = o(n).$$

To verify this, write $A_{\boldsymbol{\beta}} = \{i \le n: |\mathbf{w}_i'\boldsymbol{\beta}| > n^{-A}\|\boldsymbol{\beta}\|\}$. By (D2), $\inf_{\boldsymbol{\beta}}$ size$(A_{\boldsymbol{\beta}}) = n - o(n)$. We further notice that for $i \in A_{\boldsymbol{\beta}}$,

$$|(\mathrm{sgn}(y_i-\mathbf{w}_i'\boldsymbol{\beta}) - \mathrm{sgn}(-\mathbf{w}_i'\boldsymbol{\beta}))\mathrm{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})| \le 2I(|y_i| > |\mathbf{w}_i'\boldsymbol{\beta}|) \le 2I(|y_i| > n^A).$$

Therefore,

$$
\begin{aligned}
n^{-1} & \sup_{\|\boldsymbol{\beta}\|>b_n,\, \boldsymbol{\gamma}\in S^p} \left| \sum_{i=1}^{n} \{(\operatorname{sgn}(y_i - \mathbf{w}_i'\boldsymbol{\beta}) - \operatorname{sgn}(-\mathbf{w}_i'\boldsymbol{\beta}))\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\} \right| \\
& \le n^{-1} \sup_{\|\boldsymbol{\beta}\|>b_n,\, \boldsymbol{\gamma}\in S^p} \left[\sum_{i=1}^{n} \{2I(|y_i| > n^A)\} + \operatorname{size}(A_{\boldsymbol{\beta}}^c) \right] \\
& = 2n^{-1} \sum_{i=1}^{n} I(|y_i| > n^A) + o(1) = o_p(1),
\end{aligned}
$$

where the last step is due to (D3). The proof is then complete. □

PROOF OF LEMMA 3.1. The proof of Lemma 3.1 is a direct application of Lemma 3.2 with $W_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta})\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) - E[\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta})\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})]$. Here we first verify that conditions (L1)–(L3) of Lemma 3.2 are satisfied. First, we notice that $|\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_1) - \operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_2)| \neq 0$ ($= 2$ in fact) if and only if $\mathbf{w}_i'\boldsymbol{\gamma}_1$ and $\mathbf{w}_i'\boldsymbol{\gamma}_2$ have different signs. Consequently, $|\mathbf{w}_i'\boldsymbol{\gamma}_1| \le |\mathbf{w}_i'(\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2)| \le \|\mathbf{w}_i\|\|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|$. This proves that $E|\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_1) - \operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_2)| \le 2P(|\mathbf{w}_i'\boldsymbol{\gamma}_1| \le \|\mathbf{w}_i\|\|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|)$. Now, we can verify condition (L1) by

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} E\big|\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}_1)\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_1) - \operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}_2)\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_2)\big|^2 \\
&\quad \le \frac{4}{n}\sum_{i=1}^{n} \big[E\big|\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}_1) - \operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}_2)\big| + E\big|\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_1) - \operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_2)\big|\big] \\
&\quad \le \frac{8}{n}\sum_{i=1}^{n} \big[E\big|F_i(\mathbf{w}_i'\boldsymbol{\beta}_1) - F_i\big(\mathbf{w}_i'\boldsymbol{\beta}_2\big)\big| + P\big(\big|\mathbf{w}_i'\boldsymbol{\gamma}_1\big| \le \|\mathbf{w}_i\|\|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|\big)\big] \\
&\quad \le \frac{8}{n}\sum_{i=1}^{n} \big[BE\big|\mathbf{w}_i'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\big|^{\delta} + \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|^{\delta}\big] \\
&\quad \le 8(B(E\|\mathbf{w}\|^2)^{\delta/2} + 1)[\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|^{\delta} + \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|^{\delta}],
\end{aligned}
$$

where the third inequality here uses (C2) for the first part and (C1) for the second part.

Condition (L2) is trivial, so it remains to verify condition (L3). For this purpose, we note that by conditions (C1) and (C2),

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n} E \sup_{\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}\| + |\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}\| \le d} \big|\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta})\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) \\
& \qquad\qquad\qquad - \operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}_1)\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_1)|^2 \\
&\quad \le \frac{4}{n}\sum_{i=1}^{n} E \sup_{\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}\| + \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}\| \le d} \big[\big|\operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}) - \operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta}_1)\big| \\
& \qquad\qquad\qquad + \big|\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) - \operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}_1)\big|\big]
\end{aligned}
$$

$$\leq \frac{8}{n}\sum_{i=1}^{n}[P(|e_i - \mathbf{w}_i'\boldsymbol{\beta}_1| \leq \|\mathbf{w}_i\|d) + P(|\mathbf{w}_i'\boldsymbol{\gamma}_1| \leq \|\mathbf{w}_i\|d)]$$

$$\leq \frac{8}{n}\sum_{i=1}^{n} E[B(\|\mathbf{w}_2\|d)^{\delta} + I(|\mathbf{w}_i'\boldsymbol{\gamma}_1| \leq \|\mathbf{w}_i\|d)] \leq 8[B(E\|\mathbf{w}\|^2)^{\delta/2} + 1]d^{\delta}.$$

The first conclusion of Lemma 3.1 follows from Lemma 3.2(i) or (ii) by taking $\varepsilon = \Delta_n^{\delta/2-v}$ in the cases of $\Delta_n \to 0$, but $\varepsilon = \zeta_n \to \infty$ and $\zeta_n \ll \sqrt{n}$ otherwise. For both cases, one can verify that $\Delta_n^{\delta}|\ln \Delta_n| \ll \varepsilon \ll \sqrt{n}\Delta_n^{\delta(1+v)}$ and $\log(2+\Delta_n) \ll \varepsilon \ll \sqrt{n}$.

Now we turn to the proof of the second conclusion. For any $t > 1$, choose ρ and a such that $1 < \rho < t^{2/(1+\delta(1+\alpha))}$ and $2 < a < 2t^2/\rho^{1+\delta(1+\alpha)}$, where α is the index of Δ_n given in the assumptions of Lemma 3.1. Also define $\overline{\Delta}(\ell) = \max\{\Delta_n,\ \rho^{\ell} \leq n < \rho^{\ell+1}\}$ and $\underline{\Delta}(\ell) = \min\{\Delta_n,\ \rho^{\ell} \leq n < \rho^{\ell+1}\}$. Note that, for all large ℓ, $\overline{\Delta}(\ell)/\underline{\Delta}(\ell) < \rho^{1+\alpha}$. Then Lemma 3.2(i) implies, for any large integer ℓ,

$$P\biggl(\sup_{\|\boldsymbol{\beta}\|\leq\overline{\Delta}(\ell),\, \boldsymbol{\gamma}\in S^p}\ \max_{\rho^{\ell}\leq n<\rho^{\ell+1}} |S_n(\boldsymbol{\beta},\boldsymbol{\gamma})|/\sqrt{2n\Delta_n^{\delta}\log\log n} > t\biggr)$$

$$\leq P\biggl(\sup_{\|\boldsymbol{\beta}\|\leq\overline{\Delta}(\ell),\, \boldsymbol{\gamma}\in S^p}\ \max_{\rho^{\ell}\leq n<\rho^{\ell+1}} |S_n(\boldsymbol{\beta},\boldsymbol{\gamma})| > \sqrt{2t^2\rho^{\ell}\underline{\Delta}^{\delta}(\ell)[\log\ell + \log\rho]}\biggr)$$

$$\leq M\Bigl[\exp\Bigl(-2(a\rho)^{-1}t^2[\log\ell+\log\rho](\underline{\Delta}/\overline{\Delta})^{\delta}\Bigr) + \exp(-\rho^{\ell(1/2-\nu)})\Bigr],$$

for some $M < \infty$ and $\nu < 1/2$. The above bound is summable in ℓ, so the desired result follows from the Borel–Cantelli lemma.

Before proceeding to the proof of Lemma 3.2, we quote the Lévy inequality from Loève [(1977), page 259].

LÉVY INEQUALITY. If $X_1,\dots,X_n$ are independent random variables and $S_k = \sum_{i=1}^{k} X_i$, then, for every ϵ,

$$P\Bigl\{\max_{k\leq n}|S_k - \mathrm{Median}(S_k - S_n)| \geq \epsilon\Bigr\} \leq 2P\{|S_n| \geq \epsilon\}.$$

PROOF OF LEMMA 3.2. The proof is based on chaining. It requires a sequence $M := M_n$ satisfying

$$\sqrt{n}M^{-3\delta}\varepsilon_n^{-1} \to \infty \quad \textit{and} \quad \log M = o(\varepsilon_n^2\Delta_n^{-\delta}). \tag{A.2}$$

For simplicity, we assume $\delta = 1$. If general, replace Δ_n by Δ_n^{δ} in the proof. We only give a detailed proof for the first case where $\Delta_n \to 0$. For the cases (ii) and (iii), see the remarks at the end of this proof. In the proof, we shall assume that $\Delta_n^{-1}\varepsilon^2$ has a positive lower bound, for, otherwise, the lemma becomes trivially true. In the case of $\Delta_n \to 0$, we have $\varepsilon_n = o(\sqrt{n}\Delta^{1+\nu})$, so we can simply choose $M = \Delta_n^{-\nu/3}$.

Under our conditions on Δ_n and $\varepsilon := \varepsilon_n$, there exists $\rho := \rho_n \to 0$ slowly enough such that $\min\{M\rho^4,\ M^w\rho\} \to \infty$ with $w = 3\nu/(21\nu+6)$ and for some $4 \le k_1 := k_{1n} \le k_2 := k_{2n}$,

$$\frac{\Delta_n\sqrt{n}}{(M\rho)^{k_1-1}\varepsilon} \to \infty, \qquad \frac{\Delta_n\sqrt{n}}{(M\rho)^{k_1}\varepsilon} \to 0, \tag{A.3}$$

$$\max\left\{\frac{\Delta_n\sqrt{n}}{(M\rho)^{k_2-1}\varepsilon}, \frac{\Delta_n^{1/2}\sqrt{n}}{(M^{1/2}\rho)^{k_2-1}\varepsilon}\right\} \to \infty \tag{A.4}$$

and

$$\max\left\{\frac{\Delta_n\sqrt{n}}{(M\rho)^{k_2}\varepsilon}, \frac{\Delta_n^{1/2}\sqrt{n}}{(M^{1/2}\rho)^{k_2}\varepsilon}\right\} \to 0. \tag{A.5}$$

Note that the above choices of k_1 and k_2 are made for general cases. In our case of $\Delta_n \to 0$, the maxima in (A.4) and (A.5) are just equal to the second ratios there.

Our choice of ρ satisfies

$$M^{(k_2/4)-(3k_1/4)-(1/4)-(3/2\nu)} \ll M^{(k_2-1)/2-k_1}\rho^{k_2-1-k_1} = \frac{\Delta_n\sqrt{n}}{(M\rho)^{k_1}\varepsilon}\cdot\frac{(M^{1/2}\rho)^{k_2-1}\varepsilon}{\Delta_n^{1/2}\sqrt{n}} \to 0,$$

which implies

$$k_2 \le 3k_1 + 1 + 6/\nu. \tag{A.6}$$

Without loss of generality, we assume that $W_i(0,\boldsymbol{\gamma}) \equiv 0$. This is equivalent to working with $W_i^*(\boldsymbol{\beta},\boldsymbol{\gamma}) = W_i(\boldsymbol{\beta},\boldsymbol{\gamma}) - W_i(0,\boldsymbol{\gamma})$ rather than $W_i(\boldsymbol{\beta},\boldsymbol{\gamma})$.

We now use expanding collections of points denoted by $\{(\boldsymbol{\beta}_{j_1},\boldsymbol{\gamma}_{\ell_1})\}$, $\{(\boldsymbol{\beta}_{j_1,j_2},\boldsymbol{\gamma}_{\ell_1,\ell_2})\},\dots,\{(\boldsymbol{\beta}_{j_1,\dots,j_{k_2}},\boldsymbol{\gamma}_{\ell_1,\dots,\ell_{k_2}})\}$, with $j_t,\ell_t = 1,2,\dots,J_t$ and $t = 1,2,\dots k_2$, satisfying

$$\|\boldsymbol{\beta}_{j_1,\dots,j_{t-1}} - \boldsymbol{\beta}_{j_1,\dots,j_t}\| + \|\boldsymbol{\gamma}_{\ell_1,\dots,\ell_{t-1}} - \boldsymbol{\gamma}_{\ell_1,\dots,\ell_t}\| \le \Delta_n M^{-t+1}, \qquad 2 \le t \le k_2.$$

Also, for any $\|\boldsymbol{\beta}\| \le \Delta_n$, $\boldsymbol{\gamma} \in D$, there exist integers $j_1,\dots,j_{k_2}$ and $\ell_1,\dots,\ell_{k_2}$ such that

$$\|\boldsymbol{\beta} - \boldsymbol{\beta}_{j_1,\dots,j_{k_2}}\| + \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_{\ell_1,\dots,\ell_{k_2}}\| \le \Delta_n M^{-k_2}.$$

Note that the tth set of points is constructed by adding J_t additional points around every point in the $(t-1)$th set. These expanding sets can be found with $J_1 \le KM^{2p}\Delta_n^{-p}$ and $J_t \le KM^{2p}$ $(t > 1)$ for some constant K. For brevity, we write $W_i(t) = W_i(\boldsymbol{\beta}_{j_1,\dots,j_t},\boldsymbol{\gamma}_{\ell_1,\dots,\ell_t})$ and

$$U_i = U_{i,j_1,\dots,j_{k_2},\ell_1,\dots,\ell_{k_2}} = \sup\left|W_i(\boldsymbol{\beta},\boldsymbol{\gamma}) - W_i(k_2)\right|$$

in the rest of the proof, where sup is taken over the set $\{(\boldsymbol{\beta}, \boldsymbol{\gamma}): \|\boldsymbol{\beta}-\boldsymbol{\beta}_{j_1,\dots,j_{k_2}}\| + \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_{\ell_1,\dots,\ell_{k_2}}\| \le \Delta_n M^{-k_2}\}$. Then, we have

$$
\begin{aligned}
&P\biggl(\sup_{\|\boldsymbol{\beta}\|\le\Delta_n,\,\boldsymbol{\gamma}\in D}\max_{m\le n}\biggl|\sum_{i=1}^m W_i(\boldsymbol{\beta},\boldsymbol{\gamma})\biggr| \ge \sqrt{n}\varepsilon\biggr)\\
&\quad\le \sum_{1\le j,\,\ell\le J_1} P\biggl(\max_{m\le n}\biggl|\sum_{i=1}^m W_i(\boldsymbol{\beta}_j,\boldsymbol{\gamma}_\ell)\biggr| \ge \varepsilon\sqrt{n}(1-\rho)\biggr)\\
&\qquad+\sum_{t=2}^{k_2}\sum_{j_1,\ell_1,\dots,j_t,\ell_t} P\biggl(\max_{m\le n}\biggl|\sum_{i=1}^m (W_i(t-1)-W_i(t))\biggr| \ge \rho^{t-1}\varepsilon\sqrt{n}(1-\rho)\biggr)\\
&\qquad+\sum_{j_1,\ell_1,\dots,j_{k_2},\ell_{k_2}} P\biggl(\biggl|\sum_{i=1}^n U_i\biggr| \ge \rho^{k_2}\varepsilon\sqrt{n}(1-\rho)\biggr).
\end{aligned}
\tag{A.7}
$$

We shall show that in the right-hand side of (A.7) the first term dominates and gives the desired bound for Lemma 3.2.

For the case of $t=1$, we have $|W_i(\boldsymbol{\beta}_j,\boldsymbol{\gamma}_\ell)| \le C_2$, and

$$
E\bigl|W_i(\boldsymbol{\beta}_j,\boldsymbol{\gamma}_\ell)\bigr|^2 \le C_1\Delta_n. \tag{A.8}
$$

Since $\varepsilon^2 \gg \Delta_n$, we have

$$
\biggl|\text{median}\biggl(\sum_{i=m}^n W_i(\boldsymbol{\beta}_j,\boldsymbol{\gamma}_\ell)\biggr)\biggr| \le C_1^{1/2}\sqrt{n}\Delta_n^{1/2} = o(\sqrt{n}\varepsilon).
$$

Now by the Lévy inequality and Bernstein inequality, we obtain, for any $a>2$, $\rho_1>\rho>0$ and for sufficiently large n,

$$
\begin{aligned}
&P\biggl(\max_{m\le n}\biggl|\sum_{i=1}^m W_i(\boldsymbol{\beta}_j,\boldsymbol{\gamma}_\ell)\biggr| \ge \varepsilon\sqrt{n}(1-\rho)\biggr)\\
&\quad\le 4\exp\Bigl(-\tfrac{1}{2}\varepsilon^2 n(1-\rho_1)^2[nC_1\Delta_n + C_2(1-\rho_1)\varepsilon\sqrt{n}]^{-1}\Bigr)\\
&\quad\le 4\exp\Bigl(-(aC_1\Delta_n)^{-1}\varepsilon^2\Bigr).
\end{aligned}
\tag{A.9}
$$

For any $1<t\le k_2$, the same arguments show that for any $\rho_1>\rho>0$ and for sufficiently large n,

$$
\begin{aligned}
&P\biggl(\max_{m\le n}\biggl|\sum_{i=1}^m (W_i(t-1)-W_i(t))\biggr| \ge \rho^{t-1}\varepsilon\sqrt{n}(1-\rho)\biggr)\\
&\quad\le 2\exp\Bigl(-\frac{1}{4}n\rho^{2t-2}(1-\rho_1)^2\varepsilon^2[nC_1\Delta_n M^{-(t-1)}\\
&\qquad\qquad +C_2\rho^{t-1}(1-\rho_1)\varepsilon\sqrt{n}]^{-1}\Bigr)\\
&\quad\le \begin{cases} 4\exp\Bigl(-b\rho^{2t-2}M^{(t-1)}\Delta_n^{-1}\varepsilon^2\Bigr), & \text{if } 1<t\le k_1,\\ 4\exp(-b\sqrt{n}\rho^{t-1}\varepsilon), & \text{if } k_1<t\le k_2,\end{cases}
\end{aligned}
\tag{A.10}
$$

where $b>0$ is a constant.

To bound the last term of (A.7), note that condition (L2) implies $U_i \le 2C_2$ and condition (L3) and (A.5) imply

$$\sum_{i=1}^{n} E(U_i^2) \le C_3 n \Delta_n M^{-k_2} = o(\rho^{k_2}\sqrt{n}\varepsilon).$$

which, together with (A.5), implies that

$$\sum_{i=1}^{n} E(U_i) = O(n\Delta_n^{1/2} M^{-k_2/2}) = o(\rho^{k_2}\sqrt{n}\varepsilon).$$

Then, for any constant $\rho_2 > \rho_1$ and for sufficiently large n,

$$\begin{aligned} P\biggl(\biggl|\sum_{i=1}^{n} U_i\biggr| &\ge \rho^{k_2}\varepsilon\sqrt{n}(1-\rho)\biggr) \\ &\le P\biggl(\biggl|\sum_{i=1}^{n}(U_i - E(U_i))\biggr| \ge \rho^{k_2}\varepsilon\sqrt{n}(1-\rho_1)\biggr) \\ &\le 2\exp(-\tfrac{1}{2}c\rho^{k_2}\varepsilon\sqrt{n}(1-\rho_2)) \le 2\exp(-b\rho^{k_2}\sqrt{n}\varepsilon), \end{aligned}$$

for some $b > 0$.

Therefore, we have a bound for (A.7) as

$$\begin{aligned} P\biggl(\sup_{\|\boldsymbol{\beta}\|\le\Delta_n,\,\boldsymbol{\gamma}\in D} \max_{m\le n} \biggl|\sum_{i=1}^{m} W_i(\boldsymbol{\beta},\boldsymbol{\gamma})\biggr| &\ge \sqrt{n}\varepsilon\biggr) \\ &\le 4J_1^2 \exp\Bigl(-(aC_1\Delta_n)^{-1}\varepsilon^2\Bigr) \\ &\quad +4\sum_{t=2}^{k_1}(J_1\cdots J_t)^2\exp(-b\rho^{2t-2}M^{t-1}\Delta_n^{-1}\varepsilon^2) \\ &\quad +4\sum_{t=k_1+1}^{k_2+1}(J_1\cdots J_t)^2\exp(-b\sqrt{n}\rho^{t-1}\varepsilon), \end{aligned} \tag{A.11}$$

where we use the convention $J_{k_2+1} = 1$.

Our choices of J_t imply that the first term on the right-hand side of (A.11) is bounded by $\exp(-(a_1C_1\Delta_n)^{-1}\varepsilon^2)$ for any $a_1 > a > 2$.

Since $\rho^2 M \to \infty$, the second term on the right-hand side of (A.11) is of smaller order than $\exp(-(a_1C_1\Delta_n)^{-1}\varepsilon^2)$ as $n \to \infty$.

For the last term in (A.11), we use (A.6) and (A.3). For any $k_1 \le t \le k_2$, we have

$$\sqrt{n}\varepsilon\rho^t \gg \varepsilon^2\Delta_n^{-1}M^{k_1-1}\rho^{t+k_1-1} \gg \varepsilon^2\Delta_n^{-1}\Bigl(M^w\rho\Bigr)^{t+k_1-1},$$

where

$$w = \frac{3\nu}{21\nu+6} \le \min_{k_1\ge 4}\frac{k_1-1}{4k_1+5+6/\nu}.$$

Therefore, we bound the last sum of (A.11) by

$$4\sum_{t=k_1+1}^{k_2+1}(J_1\cdots J_t)^2\exp(-b\sqrt{n}\rho^{t-1}\varepsilon)$$

$$\le 4\sum_{t=k_1}^{k_2}(J_1\cdots J_{t+1})^2\exp\Big(-b\varepsilon^2\Delta_n^{-1}\Big(M^w\rho\Big)^{t+k_1-1}\Big)$$

$$\ll \exp\Big(-(a_1C_1\Delta_n)^{-1}\varepsilon^2\Big).$$

Putting things together, we have proved that (A.11) is bounded by $C\exp(-(a_1C_1\Delta_n)^{-1}\varepsilon^2)$ for any $a_1 > 2$, where C is a constant that may depend on a_1.

Finally, we add some remarks on the proofs for the other two cases. In case (ii), without loss of generality, we may assume that $\varepsilon \to \infty$, for otherwise the result becomes trivial if we choose a large constant C_a. As for case (iii), we only need one chain in the proof, that is, we only need to select $\{\boldsymbol{\beta}_j, \boldsymbol{\gamma}_\ell;\ j, \ell = 1, 2, \dots, J, \}$ such that for any $\|\boldsymbol{\beta}\| \le \Delta_n$ and $\boldsymbol{\gamma} \in D$, there are j and ℓ satisfying

$$\|\boldsymbol{\beta} - \boldsymbol{\beta}_j\| + \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_\ell\| \le n^{-B}.$$

By our assumptions, we can do so with $J \le K\Delta^p n^{2pB}$ and thus $\log J = o(n)$. Also, we have

$$\sum_{i=1}^{n} E \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_j\|+\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_\ell\|\le n^{-B}} \big|W_i(\boldsymbol{\beta},\boldsymbol{\gamma}) - W_i(\boldsymbol{\beta}_j,\boldsymbol{\gamma}_\ell)\big| = o(n).$$

The rest of the proof is similar to that for case (i). The proof of Lemma 3.2 ends here. □

We now prove Lemma 3.3. First, it is clear that the set of solutions to (3.8) is a nonempty convex set in R^p. Let $\boldsymbol{\beta}_0$ be one solution. Suppose that $\max_{\boldsymbol{\beta}}\min_{\boldsymbol{\gamma}\in S^p}(W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta}) = d_0$, and $G = \{\boldsymbol{\gamma} \in S^p\colon W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta}_0 = d_0\}$. By condition (W1), d_0 is finite almost surely. We now have the following lemma.

LEMMA A.1. *There does not exist* $\mathbf{1} \in S^p$ *such that* $\mathbf{1}'\boldsymbol{\mu}(\boldsymbol{\gamma}) > 0$ *for all* $\boldsymbol{\gamma} \in G$.

PROOF. Here $\boldsymbol{\mu}(\boldsymbol{\gamma})$ is continuous and G is a closed set. If the conclusion of Lemma A.1 is not true, then there is a vector $\mathbf{1}$ such that $\delta = \inf_{\boldsymbol{\gamma}\in G}\mathbf{1}'\boldsymbol{\mu}(\boldsymbol{\gamma}) > 0$ as G is obviously a compact set. Set $H = \{\boldsymbol{\gamma} \in S^p\colon \mathbf{1}'\boldsymbol{\mu}(\boldsymbol{\gamma}) > \delta/2\}$. Clearly, $H^c \cap S^p$ is a closed set and $H^c \cap G$ is empty.

Let $d_1 = \max_{\boldsymbol{\gamma}\in H^c\cap S^p}|\mathbf{1}'\boldsymbol{\mu}(\boldsymbol{\gamma})| \in (0,\infty)$ and $d_2 = \min_{\boldsymbol{\gamma}\in H^c\cap S^p}(W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta}_0) > d_0$. Consider the function

$$Q(\boldsymbol{\gamma}) = W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta}_0 + t\boldsymbol{\mu}(\boldsymbol{\gamma})'\mathbf{1},$$

with $t = (d_2 - d_0)/(2d_1)$. If $\boldsymbol{\gamma} \in H^c \cap S^p$, then $Q(\boldsymbol{\gamma}) \ge d_2 - td_1 = (d_0 + d_2)/2 > d_0$. If $\boldsymbol{\gamma} \in H$, then $Q(\boldsymbol{\gamma}) \ge d_0 + t\delta/2 > d_0$. These inequalities show that the solution should not be $\boldsymbol{\beta}_0$. The contradiction proves the lemma. □

Now we shall show that the solution to (3.8) is unique by establishing a set of linear equations that any solution to (3.8) must satisfy.

PROOF OF LEMMA 3.3. As in the proof of Lemma A.1, let $\boldsymbol{\beta}_0$ be a solution, and the minimum over $\boldsymbol{\gamma}$ in (3.8) is achieved at some $\boldsymbol{\gamma}^* \in S^p$ so that $W(\boldsymbol{\gamma}^*) + \boldsymbol{\mu}(\boldsymbol{\gamma}^*)'\boldsymbol{\beta}_0 = d_0$. By Lemma A.1, there are at least three different $\boldsymbol{\gamma}^* \in S^p$ in the set G. For otherwise, Lemma A.1 fails, because of (W1), by choosing $\mathbf{1} = -\boldsymbol{\gamma}^*$ if G contains only one vector or $\mathbf{1} = -(\boldsymbol{\gamma}_1^* + \boldsymbol{\gamma}_2^*)/\|\boldsymbol{\gamma}_1^* + \boldsymbol{\gamma}_2^*\|$ if G contains two vectors. This vector $\mathbf{1}$ is well defined since no pair of vectors in G can be in the opposite directions thanks to (W3). Also implied by (W3) is that we can always choose $\boldsymbol{\gamma}^* \in G$ such that it is not parallel to $\boldsymbol{\alpha}$.

At this $\boldsymbol{\gamma}^*$, consider the arc $\boldsymbol{\gamma} = (\boldsymbol{\gamma}^* + t\mathbf{1})/\|\boldsymbol{\gamma}^* + t\mathbf{1}\|$ as t varies for any direction $\mathbf{1}$ with $\mathbf{1}'\boldsymbol{\gamma}^* = 0$. Since $\boldsymbol{\gamma}^*$ is a minimizing point for $W(\boldsymbol{\gamma}) + \boldsymbol{\mu}(\boldsymbol{\gamma})'\boldsymbol{\beta}_0$ and the function is continuous, there must exist sequences $t_k \uparrow 0$ and $s_k \downarrow 0$ such that at least one sequence is strictly monotone and

$$W(\boldsymbol{\gamma}^* + t_k\mathbf{1}) + \boldsymbol{\mu}(\boldsymbol{\gamma}^* + t_k\mathbf{1})'\boldsymbol{\beta}_0 = W(\boldsymbol{\gamma}^* + s_k\mathbf{1}) + \boldsymbol{\mu}(\boldsymbol{\gamma}^* + s_k\mathbf{1})'\boldsymbol{\beta}_0.$$

Since $\boldsymbol{\mu}(\boldsymbol{\gamma})$ is differentiable, we know that along the sequence $k \to \infty$,

$$\lim_{k\to\infty} \frac{W(\boldsymbol{\gamma}^* + t_k\mathbf{1}) - W(\boldsymbol{\gamma}^* + s_k\mathbf{1})}{t_k - s_k} = \xi_l(\boldsymbol{\gamma}^*)$$

exists and is equal to $[\lim_{k\to\infty}(\boldsymbol{\mu}(\boldsymbol{\gamma}^* + t_k\mathbf{1}) - \boldsymbol{\mu}(\boldsymbol{\gamma}^* + s_k\mathbf{1}))/(t_k - s_k)]'\boldsymbol{\beta}_0$. That is,

$$\xi_l(\boldsymbol{\gamma}^*) = (\mathbf{1}'\mathbf{D}_{\boldsymbol{\gamma}^*})\boldsymbol{\beta}_0 \tag{A.12}$$

for any direction $\mathbf{1}$ orthogonal to $\boldsymbol{\gamma}^*$. By (W2) and the fact that $\boldsymbol{\gamma}^{*\prime}\mathbf{D}_{\boldsymbol{\gamma}^*} = 0$, $\{\mathbf{1}'\mathbf{D}_{\boldsymbol{\gamma}^*}\}$ spans the $(p-1)$-dimensional subspace orthogonal to $\boldsymbol{\gamma}^*$.

Lemma A.1 implies that there exists another $\widetilde{\boldsymbol{\gamma}} \in G$ not parallel to $\boldsymbol{\gamma}^*$ such that $W(\widetilde{\boldsymbol{\gamma}}) + \boldsymbol{\mu}(\widetilde{\boldsymbol{\gamma}})'\boldsymbol{\beta}_0 = W(\boldsymbol{\gamma}^*) + \boldsymbol{\mu}(\boldsymbol{\gamma}^*)'\boldsymbol{\beta}_0$. This gives another equation,

$$W(\widetilde{\boldsymbol{\gamma}}) - W(\boldsymbol{\gamma}^*) = (\boldsymbol{\mu}(\widetilde{\boldsymbol{\gamma}}) - \boldsymbol{\mu}(\boldsymbol{\gamma}^*))'\boldsymbol{\beta}_0. \tag{A.13}$$

By (W1), $(\boldsymbol{\mu}(\widetilde{\boldsymbol{\gamma}}) - \boldsymbol{\mu}(\boldsymbol{\gamma}^*))'\boldsymbol{\gamma}^* \neq 0$, so $\boldsymbol{\mu}(\widetilde{\boldsymbol{\gamma}}) - \boldsymbol{\mu}(\boldsymbol{\gamma}^*)$ is not in the space of $\{\mathbf{1}\colon \mathbf{1}'\boldsymbol{\gamma}^* = 0\}$. This means that (A.12) and (A.13) put together include p linearly independent equations and they uniquely determine $\boldsymbol{\beta}_0$.

Conditions (W1) and (W3) are trivial for the above defined

$$\boldsymbol{\mu}(\boldsymbol{\gamma}) = -2f(0)E\{\operatorname{sgn}(\mathbf{w}'\boldsymbol{\gamma})\mathbf{w}\}.$$

Thus, to complete the proof of Theorem 3.1, we only need to verify that condition (W2) is satisfied, which is shown in the following lemma.

LEMMA A.2. *Let* $\boldsymbol{\gamma}' = (\gamma_0, \boldsymbol{\gamma}_1') \in S^p$ *with* $\|\boldsymbol{\gamma}_1\| \neq 0$ *and let* $\mathbf{B}$ *be a* $(p-1)\times(p-1)$ *orthonormal matrix with* $\boldsymbol{\gamma}_1/\|\boldsymbol{\gamma}_1\|$ *as its first column. Assume that* $F_{\mathbf{B}}(y_0, \mathbf{y}_1)$, *the c.d.f. of* $\mathbf{B}'\mathbf{x} := (y_0, \mathbf{y}_1')'$, *is continuously differentiable with* *respect to y_0 with derivative $\dot{F}_{\mathbf{B}}(y_0, \mathbf{y}_1)$. Then, the derivative matrix of $\boldsymbol{\mu}(\boldsymbol{\gamma})$ is given by*

$$\mathbf{D}_{\gamma} = -4f(0)\|\boldsymbol{\gamma}_1\|^{-1}\int(1, \mathbf{y}'\mathbf{B}')'(1, \mathbf{y}'\mathbf{B}')\dot{F}_{\mathbf{B}}(-\gamma_0/\|\boldsymbol{\gamma}_1\|, d\mathbf{y}_1),$$

where $\mathbf{y}_1 \in R^{p-2}$ and $\mathbf{y} = (-\gamma_0/\|\boldsymbol{\gamma}_1\|, \mathbf{y}_1')'$. Consequently, the directional derivative of $\boldsymbol{\mu}(\boldsymbol{\gamma})$ along the direction $\mathbf{1} \in S^p$ is equal to $\mathbf{1}'\mathbf{D}_{\gamma}$.

It is seen from Lemma A.2 that $\mathbf{D}_{\gamma}$ is well defined if $\boldsymbol{\gamma}$ is not parallel to $\boldsymbol{\alpha} = (1, 0, \dots, 0)'$. Now, we verify that $\mathbf{D}_{\gamma}$ satisfies condition (W2) of Lemma 3.3. First, we note that $\mathbf{D}_{\gamma}\boldsymbol{\gamma} = 0$ holds for any $\boldsymbol{\gamma}$ not parallel to $\boldsymbol{\alpha}$ since $\boldsymbol{\gamma}_1'\mathbf{B} = (\|\boldsymbol{\gamma}_1\|, 0, \dots, 0)$ implies that $(1, \mathbf{y}'\mathbf{B}')\boldsymbol{\gamma} = 0$. Conversely, if $\mathbf{D}_{\gamma}\mathbf{1} = 0$ for some $\mathbf{1} \in S^p$, then $\mathbf{1}'\mathbf{D}_{\gamma}\mathbf{1} = 0$, which, together with the expression of $\mathbf{D}_{\gamma}$, implies that $(1, \mathbf{y}'\mathbf{B}')\mathbf{1} = 0$ for almost all $\mathbf{y} = (-\gamma_0/\|\boldsymbol{\gamma}_1\|, \mathbf{y}_1')$ with $\mathbf{y}_1 \in R^{p-2}$. Partitioning $\mathbf{1}' = (l_0, \mathbf{1}_1')$ and $\mathbf{B} = (\gamma_1/\|\boldsymbol{\gamma}_1\| \vdots \mathbf{B}_1)$, we get

$$l_0 - (\mathbf{1}_1'\boldsymbol{\gamma}_1/\|\boldsymbol{\gamma}_1\|)\gamma_0 + \mathbf{y}_1'\mathbf{B}_1'\mathbf{1}_1 = 0.$$

Since $\mathbf{y}_1$ runs over $p-2$ linearly independent vectors in R^{p-2}, we obtain $l_0 = (\mathbf{1}_1'\boldsymbol{\gamma}_1/\|\boldsymbol{\gamma}_1\|)\gamma_0$ and $\mathbf{1}_1'\mathbf{B}_1 = 0$. Since $\mathbf{B}$ is orthogonal, we get $\mathbf{1}_1 = c\boldsymbol{\gamma}_1$ for some $c \in R$, and hence $l_0 = c\boldsymbol{\gamma}_0$. Therefore, $\mathbf{1}$ must be parallel to $\boldsymbol{\gamma}$. Putting things together, we see that $\mathbf{D}_{\gamma}$ has rank $p-1$ and the set $\{\mathbf{1}'\mathbf{D}_{\gamma}: \mathbf{1}'\boldsymbol{\gamma} = 0\}$ forms a $(p-1)$-dimensional linear space orthogonal to $\boldsymbol{\gamma}$. This shows condition (W2) and completes the proof of Theorem 3.1. □

Now, let us prove Lemma A.2.

PROOF OF LEMMA A.2. For brevity, we suppress the constant factor $2f(0)$ from the definition of $\boldsymbol{\mu}$. We take any direction $\mathbf{1}$ such that $\mathbf{1}' = (l_0, \mathbf{1}_1') \in S^p$. Consider

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{\gamma} + tl) - \boldsymbol{\mu}(\boldsymbol{\gamma}) &= \Big(\int_{\mathbf{w}'\boldsymbol{\gamma}>0} - \int_{\mathbf{w}'(\boldsymbol{\gamma}+t\mathbf{1})>0}\Big)\mathbf{w}G(d\mathbf{x})\\ &\quad -\Big(\int_{\mathbf{w}'\boldsymbol{\gamma}<0} - \int_{\mathbf{w}'(\boldsymbol{\gamma}+t\mathbf{1})<0}\Big)\mathbf{w}G(d\mathbf{x})\\ &:= \Delta_1(t) - \Delta_2(t),\end{aligned}$$

where $G(\cdot)$ is the distribution of $\mathbf{x}$.

Note that $\mathbf{w}' = (1, \mathbf{x}')$. Use change of variables $\mathbf{x} = \mathbf{B}(y_0, \mathbf{y}_1')'$ with the orthogonal matrix $\mathbf{B}$ whose first column is $\boldsymbol{\gamma}_1/\|\boldsymbol{\gamma}_1\|$. Then $y_0 = \boldsymbol{\gamma}_1'\mathbf{x}/\|\boldsymbol{\gamma}_1\|$. Let $\mathbf{1}_1'\mathbf{B} = (a_0, \mathbf{a}_1') \in R \times R^{p-2}$. We have

$$\begin{aligned}\Delta_1(t) &:= \Big(\int_{\mathbf{w}'\boldsymbol{\gamma}>0} - \int_{\mathbf{w}'(\boldsymbol{\gamma}+t\mathbf{1})>0}\Big)\mathbf{w}G(d\mathbf{x})\\ &= \Big(\int_{\gamma_0+\boldsymbol{\gamma}_1'\mathbf{x}>0} - \int_{\gamma_0+tl_0+\boldsymbol{\gamma}_1'\mathbf{x}+t\mathbf{1}_1'\mathbf{x}>0}\Big)(1, \mathbf{x}')'G(d\mathbf{x})\\ &= -\int_{-(\gamma_0+tl_0+t\mathbf{a}_1'\mathbf{y}_1)/(\|\boldsymbol{\gamma}_1\|+ta_0)}^{-\gamma_0/\|\boldsymbol{\gamma}_1\|}(1, \mathbf{y}'\mathbf{B})F_{\mathbf{B}}(dy_0, d\mathbf{y}_1).\end{aligned}$$

It then follows that

$$\lim_{t\to 0}\frac{\Delta_1(t)}{t} = -\int \frac{-a_0\gamma_0 + l_0\|\boldsymbol{\gamma}_1\| + \mathbf{a}_1'\mathbf{y}_1\|\boldsymbol{\gamma}_1\|}{\|\boldsymbol{\gamma}_1\|^2}(1, \mathbf{y}'\mathbf{B})'\dot{F}_{\mathbf{B}}(-\gamma_0/\|\boldsymbol{\gamma}_1\|, d\mathbf{y}_1)$$

with $\mathbf{y}' = (-\gamma_0/\|\boldsymbol{\gamma}_1\|, \mathbf{y}_1')$. This is equivalent to

$$\lim_{t\to 0}\frac{\Delta_1(t)}{t} = -\|\boldsymbol{\gamma}_1\|^{-1}\Big[\int(1, \mathbf{y}'\mathbf{B})'(1, \mathbf{y}'\mathbf{B})\dot{F}_{\mathbf{B}}(-\gamma_0/\|\boldsymbol{\gamma}_1\|, d\mathbf{y}_1)\Big]\mathbf{1}.$$

Similarly, one can show that

$$\lim_{t\to 0}\frac{\Delta_2(t)}{t} = \|\boldsymbol{\gamma}_1\|^{-1}\Big[\int(1, \mathbf{y}'\mathbf{B})'(1, \mathbf{y}'\mathbf{B})\dot{F}_{\mathbf{B}}(-\gamma_0/\|\boldsymbol{\gamma}_1\|, d\mathbf{y}_1)\Big]\mathbf{1}.$$

The proof of Lemma A.2 is then complete. □

PROOF OF THEOREM 3.2. Since $\hat{\boldsymbol{\beta}}_n \to 0$, there exists a sequence of constants $\Delta_n \to 0$ such that $\limsup_{n\to\infty} ||\hat{\boldsymbol{\beta}}_n||/\Delta_n \le 1$ almost surely. We only need to consider $\|\boldsymbol{\beta}\| \le \Delta_n$. Applications of Lemma 3.2 yield

$$\limsup \sup_{\boldsymbol{\gamma}\in S^p}\left|\frac{1}{\sqrt{2n\log\log n}}\sum_{i=1}^n \operatorname{sgn}(e_i)\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\right| \le 1 \quad \text{a.s.}$$

and

$$\begin{aligned}\frac{1}{\sqrt{n\log logn}}&\sum_{i=1}^n \operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta})\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma})\\ &= \sqrt{\frac{n}{\log\log n}}E\operatorname{sgn}(e - \mathbf{w}'\boldsymbol{\beta})\operatorname{sgn}(\mathbf{w}'\boldsymbol{\gamma})\\ &\quad + \frac{1}{\sqrt{n\log\log n}}\sum_{i=1}^n \operatorname{sgn}(e_i)\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) + o(1) \quad \text{a.s.,}\end{aligned}$$

uniformly in $\boldsymbol{\gamma} \in S^p$. By (3.6), there exist $c > 0$ and $\eta(c) > 0$ such that

$$\inf_{\boldsymbol{\gamma}} E\operatorname{sgn}(e - \mathbf{w}'\boldsymbol{\beta})\operatorname{sgn}(\mathbf{w}'\boldsymbol{\gamma}) \le -\eta(c)\zeta_n$$

whenever $\|\boldsymbol{\beta}\| \ge \zeta_n/\sqrt{n}$. Thus, there exists $K < \infty$ such that $\|\boldsymbol{\beta}\| > \zeta_n = K\sqrt{\log\log n/n}$ implies

$$\inf_{\boldsymbol{\gamma}}\frac{1}{\sqrt{n\log\log n}}\sum_{i=1}^n \operatorname{sgn}(e_i - \mathbf{w}_i'\boldsymbol{\beta})\operatorname{sgn}(\mathbf{w}_i'\boldsymbol{\gamma}) < 0$$

for sufficiently large n. Similarly to the arguments in Section 2, we see that the estimate must satisfy $\|\hat{\boldsymbol{\beta}}_n\| \le K\sqrt{\log\log n/n}$ almost surely.

The second conclusion of Theorem 3.2 follows by noticing that, under (C3),

$$\inf_{\boldsymbol{\gamma}} E\operatorname{sgn}(e - \mathbf{w}'\boldsymbol{\beta})\operatorname{sgn}(\mathbf{w}'\boldsymbol{\gamma}) \le -2f(0)\|\boldsymbol{\beta}\| \inf_{\boldsymbol{\gamma}\in S^p} E|\mathbf{w}'\boldsymbol{\gamma}|(1 + o(1)). \qquad \square$$

Acknowledgments. The authors thank two anonymous referees and one Associate Editor whose valuable comments and suggestions helped improve the paper.

REFERENCES

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

HE, X. (1999). Comment on "Regression depth," by P. J. Rousseeuw and M. Hubert. *J. Amer. Statist. Assoc.* **94** 403–404.

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

HE, X. and PORTNOY, S. (1998). Asymptotics of the deepest line. In *Applied Statistical Science III: Nonparametric Statistics and Related Topics* (S.E. Ahmed, M. Ahsanullah, and B.K. Sinha, eds.) 71–81. Nova Science Publications, New York.

HE, X. and SHAO, Q. M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.* **24** 2608–2630.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.* **27**. 783–840.

LOÈVE, M. (1977). *Probability Theory*, 4th ed. Springer, New York.

NOLAN, D. (1999). On Min-Max Majority and Deepest Points. *Statist. Probab. Lett.* **43** 325–334.

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

ROUSSEEUW, P. J. and HUBERT, M. (1999). Regression depth (with discussions). *J. Amer. Statist. Assoc.* **94** 388–402.

ROUSSEEUW, P. J. and STRUYF, A. (1998). Computing location depth and regression depth in higher dimensions. *Statist. Comput.* **8** 193–203.

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

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

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
REPUBLIC OF SINGAPORE 119260

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