The Annals of Statistics
2015, Vol. 43, No. 6, 2738–2765
DOI: 10.1214/15-AOS1359

BRIDGING CENTRALITY AND EXTREMITY: REFINING EMPIRICAL DATA DEPTH USING EXTREME VALUE STATISTICS

BY JOHN H. J. EINMAHL, JUN LI AND REGINA Y. LIU[1]

Tilburg University, University of California, Riverside and Rutgers University

Statistical depth measures the centrality of a point with respect to a given distribution or data cloud. It provides a natural center-outward ordering of multivariate data points and yields a systematic nonparametric multivariate analysis scheme. In particular, the half-space depth is shown to have many desirable properties and broad applicability. However, the empirical half-space depth is zero outside the convex hull of the data. This property has rendered the empirical half-space depth useless outside the data cloud, and limited its utility in applications where the extreme outlying probability mass is the focal point, such as in classification problems and control charts with very small false alarm rates. To address this issue, we apply extreme value statistics to refine the empirical half-space depth in "the tail." This provides an important linkage between data depth, which is useful for inference on centrality, and extreme value statistics, which is useful for inference on extremity. The refined empirical half-space depth can thus extend all its utilities beyond the data cloud, and hence broaden greatly its applicability. The refined estimator is shown to have substantially improved upon the empirical estimator in theory and simulations. The benefit of this improvement is also demonstrated through the applications in classification and statistical process control.

1. Introduction. Statistical depth generally is a measure of centrality with respect to a multivariate distribution or a data cloud. It is shown to have many useful data-driven features for developing statistical inference methods and applications. For example, among other features, it can also yield a center-outward ordering, and thus order statistics and ranks for multivariate data. With its rapid and broad advances, statistical depth has emerged to be a powerful alternative approach in multivariate analysis.

There exist many different notions of statistical depth; see, for example, Liu, Parelius and Singh (1999) and Zuo and Serfling (2000) and the references therein. But the so-called geometric depths such as the half-space depth Tukey (1975) and the simplicial depth Liu (1990) are often preferred in many nonparametric inference methods and applications for their intrinsic desirable properties, as seen in Donoho and Gasko (1992), Liu and Singh (1993, 1997), Yeh and Singh (1997),

Received September 2014; revised June 2015.
[1]Supported by NSF Grants DMS-07-07053 and DMS-10-07683.
MSC2010 subject classifications. Primary 62G05, 62G20, 62G32; secondary 62H30, 62P30.
Key words and phrases. Depth, extremes, nonparametric classification, nonparametric multivariate SPC, tail.

Rousseeuw and Hubert (1999), Liu, Parelius and Singh (1999), Zuo and Serfling (2000), Li and Liu (2004), Hallin, Paindaveine and Šiman (2010) and many others.

In practice, the empirical versions of the half-space depth and the simplicial depth, however, suffer from the problem of vanishing value outside the convex hull of the data. This problem is inherent in any depth function that uses empirical counts based on the data to compute its value. It renders the empirical version of such a depth useless outside the data cloud, and limits its utility in applications involving extreme outlying probability mass. A successful resolution to this problem can avert such limitations and greatly enhance the utility of depth functions. In investigating this problem, we observe that the half-space depth involves projecting data points onto unit vectors, and thus naturally lends itself in the framework of extreme value theory. Therefore, we propose to refine the empirical half-space depth by applying extreme value statistics to "the tail." The aim of this paper is to present this proposal, and assess and demonstrate the improvement achieved by the proposal, in theory and applications.

To be more precise, let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random vectors taking values in $\mathbb{R}^d, d \geq 1$. Denote the common probability measure with P and the empirical measure with P_n; denote closed half-spaces with H. Then the half-space depth at $\mathbf{x} \in \mathbb{R}^d$ is defined by

$$D(\mathbf{x}) = \inf_{H:\mathbf{x}\in H} P(H).$$

Observe that the infimum can be restricted to half-spaces H with $\mathbf{x}$ on their boundary. We can also write

$$D(\mathbf{x}) = \inf_{\|\mathbf{u}\|=1} \mathbb{P}\big(\mathbf{u}^T\mathbf{X}_1 \geq \mathbf{u}^T\mathbf{x}\big),$$

with $\|\cdot\|$ the radius or L_2-norm of a vector. The classical nonparametric way to estimate $D(\mathbf{x})$ is with the empirical half-space depth:

$$D_n(\mathbf{x}) = \inf_{H:\mathbf{x}\in H} P_n(H) = \frac{1}{n} \inf_{\|\mathbf{u}\|=1} \#\big\{i \in \{1, \ldots, n\} : \mathbf{u}^T\mathbf{X}_i \geq \mathbf{u}^T\mathbf{x}\big\}.$$

It follows that for any $\mathbf{x}$ outside the convex hull of the data $D_n(\mathbf{x}) = 0$. This might seem a minor problem. Indeed, when the data are univariate, the probability that a new observation falls outside the convex hull is at most $2/(n+1)$, but in higher dimensions this probability can be quite sizable. For example, for the multivariate normal distribution and $n = 100$ this probability is 8.8% in dimension 2 and 21.7% in dimension 3. Even when n is as large as 500, these probabilities are still 2.1% ($d = 2$) and 6.5% ($d = 3$); see, for example, Efron (1965). Outside the data hull, D_n makes no distinction between different points and provides hardly information about P. This inability of distinguishing points in a sizable subspace can severely restrict the utility of half-space depth in many of its applications, such as statistical process control and classification (see Section 3). Note that the problem is not restricted to D_n being exactly 0: if $D_n(\mathbf{x})$ is positive but very small,

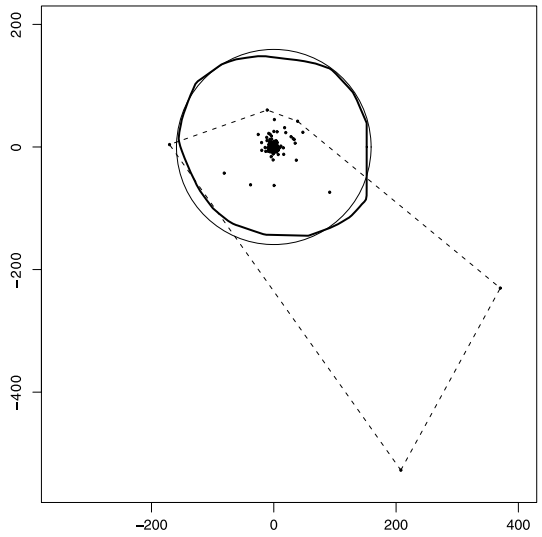

FIG. 1. *Depth contours at level* $1/n$ *based on* D (*circle*), D_n (*dashed*) *and* R_n (*solid*) *for a standard bivariate spherical Cauchy random sample*; $n = 500$.

it might not adequately estimate $D(\mathbf{x})$ due to the scarcity of useful data points. Somewhat related, due to the discrete nature of D_n, ties occur often. For example, $D_n(\mathbf{X}_i) = 1/n$ for *all* the data on the boundary of the data hull, that is, all these data form one tie and cannot be ranked effectively. (For the normal distribution in dimension 3 and $n = 500$ this tie, on average, has a size of about 32.) This phenomenon renders rank procedures based on depth less precise and less efficient.

The goal of this paper is to refine the definition of empirical half-space depth D_n in the tail, that is, for values $\mathbf{x}$ where $D_n(\mathbf{x})$ is zero or quite small. The proposed refined estimator will be called R_n (see Section 2 for the definition) and is based on extreme value theory. The estimator R_n is equal to D_n in the central region, where the depth is relatively high. Outside this region R_n is positive, smooth and it improves substantially on D_n. Therefore, the aforementioned weaknesses of D_n are "repaired."

As an illustration, we consider the estimation of the depth contour at level $1/n$, that is, we want to estimate the set $\{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}) = 1/n\}$, based on a random sample of size n. Using D_n, it is usually estimated with the boundary of the data hull, where indeed $D_n = 1/n$, almost surely. We also estimate it using our refined estimator by $\{\mathbf{x} \in \mathbb{R}^d : R_n(\mathbf{x}) = 1/n\}$. We consider as an example the bivariate spherical Cauchy distribution (see Section 2.3) and simulate one random sample of size $n = 500$; see Figure 1. (The computation of these depth contours is discussed in Remark 6 of Section 2.2.) It clearly shows that R_n greatly improves D_n; D_n fails completely here, whereas R_n performs well. This indicates that our refined estimator can be very useful in practice.

In the next section, we will define R_n and show, under appropriate conditions, its uniform ratio consistency (considering $R_n/D-1$) on a very large region, much larger than the data hull. In contrast, D_n/D is not uniformly close to 1 *on* the data hull. We further show through simulations that these asymptotic differences between R_n and D_n are clearly present for finite samples, that is, that R_n substantially outperforms D_n in the tail. In Section 3, we investigate the impact of these theoretical improvements in real applications of data depth using examples in statistical process control (SPC) and classification. Both applications obtain substantial improvements by using R_n. Finally, we provide some concluding remarks in Section 4. All proofs are deferred to Section 5.

2. Methodology and main results.

2.1. *Dimension one.* We first consider refining D_n in the one-dimensional case, particularly since it serves as a building block for us to refine D_n in higher dimensions. Let $X_1,\dots,X_n$ be i.i.d. random variables with common continuous distribution function F with $0<F(0)<1$. Write $S=1-F$. Let F_n be the (right-continuous) empirical distribution function and define $S_n(x)=1-F_n(x^-)$. The half-space depth and its empirical counterpart in the one-dimensional case are simply $D(x)=\min(F(x),S(x))$ and $D_n(x)=\min(F_n(x),S_n(x))$, respectively. It is clear that the aforementioned shortcomings of D_n are due to the inadequacy of the empirical distribution function as an estimator in the tails. Since extreme value statistics is well suited for inference problems in this setting, we propose applying it to refine D_n in the tails.

In extreme value theory, it is assumed that there exist a location function b and a scale function $a>0$ such that

$$\lim_{t\to\infty} t\bigl(1-F\bigl(a(t)y+b(t)\bigr)\bigr)=-\log G_\gamma(y)=(1+\gamma y)^{-1/\gamma}, \qquad 1+\gamma y>0. \tag{1}$$

Here, G_γ is the limiting extreme value distribution and $\gamma\in\mathbb{R}$ is the extreme value index. If (1) holds, F is said to be in the max domain of attraction of G_γ. See, for example, de Haan and Ferreira (2006). The above assumption guarantees that F has a "regular" tail and makes extrapolation outside the data range possible.

If F is in the max-domain of attraction of G_γ, by setting $t=n/k$ and $x=a(t)y+b(t)$ in (1), we obtain for large n/k and large x

$$\mathbb{P}(X>x)=1-F(x)\approx\frac{k}{n}\biggl(1+\gamma\frac{x-b(n/k)}{a(n/k)}\biggr)^{-1/\gamma}. \tag{2}$$

Let $\hat\gamma$ and $\hat a=\hat a(n/k)$ be estimators for γ and $a=a(n/k)$, respectively. Define $\hat b=\hat b(n/k)=X_{n-k:n}$, where $X_{i:n}$ denotes the ith order statistic of $X_1,\dots,X_n$.

Plugging these estimators into (2), we obtain the following estimator for the right-tail probability $1 - F(x)$:

$$p_n^r(x) = \frac{k}{n}\left(\max\left[0, 1 + \hat{\gamma}\frac{x - \hat{b}}{\hat{a}}\right]\right)^{-1/\hat{\gamma}}. \tag{3}$$

To estimate the left-tail probability, we can define $p_n^l(x)$ similarly as $p_n^r(x)$ by using the $-X_i$.

The general idea of estimating D with our refined estimator is the following. For a given k, we define the central region to be $(X_{k+1:n}, X_{n-k:n})$. For x in this central region, we define $R_n(x) = D_n(x)$, that is, we use the classical empirical half-space depth. In the right tail, that is, when $x \geq X_{n-k:n}$, we refine D_n by defining $R_n(x) = p_n^r(x)$ and similarly, when $x \leq X_{k+1:n}$ (the left tail), we set $R_n(x) = p_n^l(x)$. At the "glue-up" points $X_{n-k:n}$ and $X_{k+1:n}$, we have

$$R_n(X_{n-k:n}) = p_n^r(X_{n-k:n}) = \frac{k}{n} = 1 - F_n(X_{n-k:n}) = D_n\big(X_{n-k:n}^+\big),$$

$$R_n(X_{k+1:n}) = p_n^l(X_{k+1:n}) = \frac{k}{n} = F_n\big(X_{k+1:n}^-\big) = D_n\big(X_{k+1:n}^-\big).$$

In the following, we study the asymptotic properties of our refined empirical half-space depth R_n. Throughout we assume that $k = k_n < n/2$ is an intermediate sequence: a sequence of positive integers satisfying

$$k \to \infty \quad \text{and} \quad k/n \to 0 \qquad \text{as } n \to \infty. \tag{4}$$

We need a second-order condition in both the left tail and the right tail; for simplicity, we will only specify it for the right tail. Let $V(t) = F^{-1}(1 - 1/t)$, $t > 1$, be the tail quantile function. We can and will take the location function $b(t) = V(t)$. We assume that the derivative V' exists and that for some eventually positive or eventually negative function A with $\lim_{t\to\infty} A(t) = 0$ and for some $\rho < 0$ we have

$$\lim_{t\to\infty} \frac{V'(tx)/V'(t) - x^{\gamma-1}}{A(t)} = x^{\gamma-1}\frac{x^\rho - 1}{\rho}, \qquad x > 0. \tag{5}$$

This condition implies (for $\rho < 0$)

$$\lim_{t\to\infty} \sup_{y\geq 1/2, y\neq 1} \left|\frac{((V(ty) - V(t))/(tV'(t)))(\gamma/(y^\gamma - 1)) - 1}{A(t)}\right| < \infty. \tag{6}$$

This limit relation is somewhat similar to Lemma 4.3.5 in de Haan and Ferreira (2006). A proof can be given along the lines of the proof of that lemma; the proof uses in particular Theorem 2.3.9 in de Haan and Ferreira (2006), with U and γ there replaced by V' and $\gamma - 1$, respectively. We can and will take the scale function $a(t) = tV'(t)$. We assume

$$\sqrt{k}A(n/k) \to \lambda \qquad \text{for some } \lambda \in \mathbb{R}. \tag{7}$$

We will also assume that the estimators $\hat{\gamma}$ and $\hat{a}$ are such that

$$\Gamma_n := \sqrt{k}(\hat{\gamma} - \gamma) = O_p(1) \quad \text{and} \quad \sqrt{k}\left(\frac{\hat{a}}{a} - 1\right) = O_p(1). \tag{8}$$

This condition is known to hold for various estimators of γ and a; see de Haan and Ferreira [(2006), Chapters 3 and 4]. Define

$$w_\gamma(t) = t^{-\gamma} \int_1^t s^{\gamma-1} \log s \, ds, \qquad t > 1.$$

Note that, as $t \to \infty$,

$$w_\gamma(t) \sim \begin{cases} \dfrac{1}{\gamma} \log t, & \gamma > 0, \\ \dfrac{1}{2}(\log t)^2, & \gamma = 0, \\ \dfrac{1}{\gamma^2} t^{-\gamma}, & \gamma < 0. \end{cases}$$

THEOREM 1. *Let δ_n be a sequence of numbers in $(0, 1/2)$ such that $n\delta_n \to 0$ as $n \to \infty$. Assume that (1) and its left-tail counterpart hold; also assume $w_\gamma(\frac{k}{n\delta_n})/\sqrt{k} \to 0$ as $n \to \infty$. Then, if (4), (5), (7) and (8) hold, we have*

$$\sup_{x \in \mathbb{R}: D(x) \geq \delta_n} \left| \frac{R_n(x)}{D(x)} - 1 \right| \xrightarrow{p} 0 \qquad \text{as } n \to \infty.$$

The condition on δ_n and k specializes to

$$\frac{\log(n\delta_n)}{\sqrt{k}} \to 0 \qquad \text{for } \gamma > 0 \quad \text{and} \quad \frac{\log^2(n\delta_n)}{\sqrt{k}} \to 0 \qquad \text{for } \gamma = 0.$$

REMARK 1. The main focus of this paper is on the tails where both R_n and D are small, and as such, $R_n - D$ (just like $D_n - D$) is inherently small as well. This implies that the usual consistency statement $\sup_x |R_n(x) - D(x)| \xrightarrow{p} 0$ is not particularly meaningful for assessing the performance of R_n as an estimator of D. Instead, we consider the ratio consistency in terms of $R_n/D - 1$ as stated in Theorem 1. Note that, in addition to the usual consistency $\sup_x |D_n(x) - D(x)| \xrightarrow{p} 0$ [Donoho and Gasko (1992)], Theorem 1 also holds for D_n, when $n\delta_n \to \infty$, but *not* when $n\delta_n$ tends to a nonnegative constant; cf. (19) below. This shows that the region for which R_n/D is close to 1 (for large n and with high probability) is much greater than that for D_n/D.

REMARK 2. It is natural to consider an asymptotic normality result instead of the consistency result in Theorem 1, but note that the convergence rate ($1/r_n$, say, with, $r_n/\sqrt{n} \to 0$) for the process $R_n/D - 1$ in such a result will be determined

by x_n-values with $D(x_n) \to 0$; at a fixed x the weak limit of $r_n(R_n(x)/D(x) - 1) = (r_n/\sqrt{n})\sqrt{n}(R_n(x)/D(x) - 1)$ will be 0. This means that a proper refinement of Theorem 1, specifying the rate of convergence and providing a nondegenerate limit, is not possible. On the other hand, if we consider a single $x = x_n$ in the right tail such that $nD(x_n)/k \to 0$, then it follows from Theorem 4.4.1 in de Haan and Ferreira (2006) (under the assumptions there) that for some μ and $\sigma > 0$

$$\frac{\sqrt{k}}{w_\gamma(k/(nD(x_n)))}\left(\frac{R_n(x_n)}{D(x_n)} - 1\right) \xrightarrow{d} N(\mu, \sigma^2) \qquad \text{as } n \to \infty,$$

since $R_n(x_n) = p_n^r(x_n)$, see (3), with probability tending to one. Indeed, the convergence rate here is slower than for fixed x: $\frac{r_n}{\sqrt{n}} = \frac{\sqrt{k}}{w_\gamma(k/(nD(x_n)))\sqrt{n}} \to 0$.

2.2. *Higher dimensions.* We next consider constructing the refined half-space depth estimator in the more interesting, multivariate case, that is, $d \geq 2$. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random vectors drawn from a common continuous distribution function F. To refine D_n we need now some more structure for F. More precisely, we assume multivariate regular variation for F, that is, there exists a measure ν such that

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{X}_1 \in tB)}{\mathbb{P}(\|\mathbf{X}_1\| \geq t)} = \nu(B) < \infty, \tag{9}$$

for every Borel set B on $\mathbb{R}^d$ that is bounded away from the origin and satisfies $\nu(\partial B) = 0$; see, for example, Jessen and Mikosch (2006). Note that the choice of the "spherical" L_2-norm is not relevant: any other norm can be used instead. This implies that for some $\alpha > 0$

$$\lim_{t\to\infty} \frac{\mathbb{P}(\|\mathbf{X}_1\| \geq tx)}{\mathbb{P}(\|\mathbf{X}_1\| \geq t)} = x^{-\alpha} \qquad \text{for } x > 0.$$

The parameter α is called the tail index and $\gamma = 1/\alpha > 0$ is the extreme value index. Note that, for all $a > 0$, $\nu(aB) = a^{-\alpha}\nu(B)$. We further require

$$\frac{\mathbb{P}(\|\mathbf{X}_1\| > t)}{t^{-\alpha}} \to c \in (0, \infty). \tag{10}$$

This simple condition in effect replaces the second-order condition of the univariate case, although it is a slightly weaker condition; cf. Cai, Einmahl and de Haan (2011), page 1807. We also assume that

(11) $\mathbf{u}^T\mathbf{X}_1$ has a continuous distribution function $F_{\mathbf{u}}$ for every unit vector $\mathbf{u}$,

and that, with $H_{r,\mathbf{u}} := \{\mathbf{x} \in \mathbb{R}^d : \mathbf{u}^T\mathbf{x} \geq r\}$, $r > 0$,

$$\inf_{\|\mathbf{u}\|=1} \nu(H_{1,\mathbf{u}}) > 0. \tag{12}$$

Note that the continuity of the $F_{\mathbf{u}}$ implies the continuity of D. Also, observe that the multivariate regular variation condition (9) implies that for every unit vector $\mathbf{u}$, $F_{\mathbf{u}}$ is in the univariate max domain of attraction with the same $\gamma = 1/\alpha$: as $t \to \infty$,

$$\frac{1 - F_{\mathbf{u}}(tr)}{1 - F_{\mathbf{u}}(t)} = \frac{\mathbb{P}(\mathbf{X}_1 \in tr H_{1,\mathbf{u}})}{\mathbb{P}(\mathbf{X}_1 \in t H_{1,\mathbf{u}})} = \frac{\mathbb{P}(\mathbf{X}_1 \in tr H_{1,\mathbf{u}})}{\mathbb{P}(\|\mathbf{X}_1\| \geq t)} \cdot \frac{\mathbb{P}(\|\mathbf{X}_1\| \geq t)}{\mathbb{P}(\mathbf{X}_1 \in t H_{1,\mathbf{u}})} \to \frac{\nu(r H_{1,\mathbf{u}})}{\nu(H_{1,\mathbf{u}})} = r^{-\alpha}.$$

Recall that the half-space depth, relative to $\mathbb{P}$, is defined as

$$D(\mathbf{x}) = \inf_{\|\mathbf{u}\|=1} \mathbb{P}\big(\mathbf{u}^T \mathbf{X}_1 \geq \mathbf{u}^T \mathbf{x}\big).$$

To estimate $D(\mathbf{x})$, we only need to estimate the one-dimensional tail probabilities $\mathbb{P}(\mathbf{u}^T \mathbf{X}_1 \geq \mathbf{u}^T \mathbf{x})$ along each projection direction $\mathbf{u}$. Since we already know how to construct the refined estimator for a tail probability in the one-dimensional case, we are now ready to define our refined empirical half-space depth R_n in dimension d.

More specifically, fix a direction (a unit vector) $\mathbf{u}$. Consider the univariate data $W_i = \mathbf{u}^T \mathbf{X}_i$, $i = 1, \ldots, n$. We can refine the tail probability estimator of the W_i similarly as in the previous subsection, but since $\gamma > 0$ we can use $a = \gamma b$. This leads, for $w \geq W_{n-k:n}$, to a simplified estimator of the right-tail probabilities:

$$p_{n,\mathbf{u}}(w) = \frac{k}{n} \left(\frac{w}{W_{n-k:n}} \right)^{-\widehat{\alpha}}; \tag{13}$$

cf. (2) and (3). The estimator $\widehat{\alpha} = 1/\hat{\gamma}$ will be based on the $\|\mathbf{X}_i\|$. We will assume that $\hat{\gamma}$ is such that

$$\Gamma_n := \sqrt{k}(\hat{\gamma} - \gamma) = O_p(1). \tag{14}$$

For $w < W_{n-k:n}$ an estimator of $1 - F_{\mathbf{u}}(w)$ is simply $1 - F_{n,\mathbf{u}}(w)$, with $F_{n,\mathbf{u}}$ the empirical distribution function of $W_1, \ldots, W_n$. Denote the thus obtained estimator of $1 - F_{\mathbf{u}}$ with $1 - \widehat{F}_{\mathbf{u}}$. This leads to the refined estimator of $D(\mathbf{x})$:

$$R_n(\mathbf{x}) = \inf_{\|\mathbf{u}\|=1} 1 - \widehat{F}_{\mathbf{u}}\big(\mathbf{u}^T \mathbf{x}-\big).$$

Next, we present the analogue of Theorem 1 for the multivariate R_n. Note that it is much more complicated to analyze R_n here than in dimension one, since for every $\mathbf{x} \in \mathbb{R}^d$ we have infinitely many directions $\mathbf{u}$ instead of only two.

THEOREM 2. *Let δ_n be a sequence of numbers in $(0, 1/2)$ such that $n\delta_n \to 0$ as $n \to \infty$. Also assume $\log(n\delta_n)/\sqrt{k} \to 0$ as $n \to \infty$. Then, if (9), (4), (10), (11), (12) and (14) hold, we have*

$$\sup_{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}) \geq \delta_n} \left| \frac{R_n(\mathbf{x})}{D(\mathbf{x})} - 1 \right| \xrightarrow{p} 0 \qquad \text{as } n \to \infty. \tag{15}$$

REMARK 3. It is known that the half-space depth is affine invariant. This means that the depth value does not change under a linear transformation. Specifically, $D(\mathbf{x}) = D_{\mathbf{A},\mathbf{b}}(\mathbf{A}\mathbf{x} + \mathbf{b})$, where $D_{\mathbf{A},\mathbf{b}}$ indicates the depth value based on the sample $\mathbf{A}\mathbf{X}_i + \mathbf{b}, i = 1, \ldots, n$. Here, $\mathbf{A}$ is a $d \times d$ nonsingular matrix and $\mathbf{b} \in \mathbb{R}^d$. Although this property does not hold for R_n exactly, it holds approximately through (15).

REMARK 4. The class of multivariate regularly varying distributions [see (9)] is quite broad. It contains, for example, all elliptical distributions with a heavy tailed radial distribution (such as multivariate t-distributions) and all distributions in the sum domain of attraction of a multivariate (nonnormal) stable distribution; see, for example, Meerschaert and Scheffler (2001), part III. Some examples are seen in Section 2.3. Note in particular that the extreme density contours of such distributions can have more or less arbitrary shapes, not only spheres or ellipsoids. Two such distributions, with nonconvex or asymmetric extreme density contours, can be found in Cai, Einmahl and de Haan (2011). It is also worth noting that the multivariate regular variation condition can be verified using the test in Einmahl and Krajina (2015).

REMARK 5. For D_n the statement of Theorem 2 holds when $n\delta_n \to \infty$ [see (28) below] but *not* when $n\delta_n$ tends to a nonnegative constant, which again shows that R_n/D is close to 1 (for large n and with high probability) on a much larger region than where D_n/D is.

REMARK 6. (i) *Computation of* R_n: Recall that when D_n or R_n is at least k/n, then they are equal. Let $\mathbf{x}$ be such that $D_n(\mathbf{x}) = R_n(\mathbf{x}) = k/n$ and let $\mathbf{x}^* = c\mathbf{x}$ with $c > 1$. Based on (13), we obtain

$$R_n(\mathbf{x}^*) = c^{-\widehat{\alpha}} R_n(\mathbf{x}). \tag{16}$$

Combination of both properties enables us to calculate R_n readily by utilizing any available algorithm for computing D_n.

(ii) *Computation of depth contour based on* R_n *in Figure* 1: Write $\mathbf{x}^* = (r_\theta \cos\theta, r_\theta \sin\theta)$. We need to find r_θ such that $R_n(\mathbf{x}^*) = 1/n$ for all $\theta \in [0, 2\pi)$. For any fixed θ, similar to the above procedure for computing R_n, we first find $\mathbf{x} = (s_\theta \cos\theta, s_\theta \sin\theta)$ such that $D_n(\mathbf{x}) = k/n$. Then based on (16), $\mathbf{x}^* = k^{1/\widehat{\alpha}}(s_\theta \cos\theta, s_\theta \sin\theta)$ and $R_n(\mathbf{x}^*) = 1/n$. The R_n-depth contour in Figure 1 is drawn using 500 evenly distributed θ's in $[0, 2\pi)$.

REMARK 7. Our estimator R_n involves k and its performance obviously will be affected by the choice of k. The problem of choosing optimal k is an inherent one in extreme value statistics. Various approaches have been proposed in the literature. One commonly used heuristic approach is to plot the relevant estimator versus k, visually identify the first (or earliest) stable (approximately constant) region

in the plot, and then choose the midpoint of this region as k. This approach is the one we used in our numerical studies below. We find more or less the same value of k in the first few samples. For some specific problem settings, procedures for determining the optimal k have been developed. It would be worthwhile developing such a procedure for R_n. Meanwhile, we note that even with the present choice of k, which may well be only suboptimal, R_n already clearly outperforms D_n.

2.3. *Simulation comparison between R_n and D_n.* In this section, we present a simulation study to compare the performance of our refined empirical half-space depth R_n with the performance of the original empirical half-space depth D_n. We consider the following distributions in our simulation study:

- Standard normal distribution. This is a light-tailed distribution with $\gamma = \rho = 0$.
- Cauchy distribution. This is a very heavy-tailed distribution with $\gamma = 1$ and $\rho = -2$.
- t-distribution with 2 degrees of freedom. This is a heavy-tailed distribution with $\gamma = 1/2$ and $\rho = -1$.
- Burr-type distribution, which is a symmetric distribution about 0 with density

$$f(x) = \frac{3|x|^5}{2(1+x^6)^{3/2}}, \qquad x \in \mathbb{R}.$$

  This distribution is less heavy-tailed with $\gamma = 1/3$ and $\rho = -2$.
- Standard bivariate normal distribution. This is a light-tailed distribution with $\gamma = 0$.
- Bivariate spherical Cauchy distribution with density

$$f(x, y) = \frac{1}{2\pi(1+x^2+y^2)^{3/2}}, \qquad (x, y) \in \mathbb{R}^2.$$

  This is a very heavy-tailed distribution with $\gamma = 1$.
- Bivariate elliptical distribution with density ($r_0 \approx 1.2481$)

$$f(x, y) = \begin{cases} \dfrac{3}{4\pi} r_0^4 (1+r_0^6)^{-3/2}, & x^2/4 + y^2 < r_0^2, \\ \dfrac{3(x^2/4+y^2)^2}{4\pi(1+(x^2/4+y^2)^3)^{3/2}}, & x^2/4 + y^2 \ge r_0^2. \end{cases}$$

  This is a less heavy-tailed distribution with $\gamma = 1/3$.
- Bivariate "clover" distribution with density ($r_0 \approx 1.2481$)

$$f(x, y) = \begin{cases} \dfrac{3}{10\pi} r_0^4 (1+r_0^6)^{-3/2} \left(5 + \dfrac{4(x^2+y^2)^2 - 32x^2y^2}{r_0(x^2+y^2)^{3/2}}\right), \\ \qquad x^2 + y^2 < r_0^2, \\ \dfrac{3(9(x^2+y^2)^2 - 32x^2y^2)}{10\pi(1+(x^2+y^2)^3)^{3/2}}, \\ \qquad x^2 + y^2 \ge r_0^2. \end{cases} \tag{17}$$

This is again a less heavy-tailed distribution with $\gamma = 1/3$, however, it is not an elliptical distribution; see Cai, Einmahl and de Haan (2011).

- Trivariate spherical Cauchy distribution with density

$$f(x, y, z) = \frac{1}{\pi^2(1 + x^2 + y^2 + z^2)^2}, \qquad (x, y, z) \in \mathbb{R}^3.$$

This is a very heavy-tailed distribution with $\gamma = 1$.

- Quadrivariate spherical Cauchy distribution with density

$$f(x, y, z, w) = \frac{3}{4\pi^2(1 + x^2 + y^2 + z^2 + w^2)^{3/2}}, \qquad (x, y, z, w) \in \mathbb{R}^4.$$

This is again a very heavy-tailed distribution with $\gamma = 1$.

The first four distributions will be used to assess the finite sample performance of Theorem 1 (although for the standard normal distribution $\rho < 0$ does not hold), and the last five distributions are used to assess the finite sample performance of Theorem 2.

For each of the above distributions, we first generate a random sample of size 500. Based on this random sample, R_n and D_n are then calculated for a point $\mathbf{x}$ where the theoretical depth $D(\mathbf{x})$ is $1/100$, $1/500$, $1/1000$ and $1/2000$, respectively. To calculate R_n, an estimator of $\gamma = 1/\alpha$ (and a) is needed. For the univariate distributions, we use the moment estimator of Dekkers, Einmahl and de Haan (1989) for estimating γ and for a we use a corresponding estimator; see formula (4.2.4) in de Haan and Ferreira (2006). For the multivariate distributions (except the bivariate normal), since we assume that $\gamma > 0$, we use the Hill (1975) estimator, based on the $\|\mathbf{X}_i\|$. For the bivariate normal distribution, because it does not satisfy the conditions of Theorem 2, we use (3) instead of (13) to estimate the right-tail probability of the W_i. In other words,

$$p_{n,\mathbf{u}}(w) = \frac{k}{n}\left(\max\left[0, 1 + \hat{\gamma}\frac{w - \hat{b}_{\mathbf{u}}}{\hat{a}_{\mathbf{u}}}\right]\right)^{-1/\hat{\gamma}},$$

where $\hat{\gamma}$ is the moment estimator based on the $\|\mathbf{X}_i\|$, $\hat{b}_{\mathbf{u}} = W_{n-k:n}$, and $\hat{a}_{\mathbf{u}}$ is again as in (4.2.4) in de Haan and Ferreira (2006). Since (16) does not hold for this case any more, we follow Cuesta-Albertos and Nieto-Reyes (2008) to approximate R_n using 500 $\mathbf{u}$'s that are uniformly and independently distributed on the unit sphere. For all 10 distributions the value of k is selected by searching visually for the first stable part in the plots, based on 3 to 5 samples, as described in more detail in Remark 7. This leads to values of k ranging from 50 to 100: 6 times 50, twice 75 and twice 100.

We carry out the above simulation 100 times for each of the distributions. The boxplots of $R_n(\mathbf{x})/D(\mathbf{x})$ and $D_n(\mathbf{x})/D(\mathbf{x})$ for each of the four depth levels from the 100 simulations for different distributions are plotted in Figures 2 and 3. As we can see from those boxplots, for all the four depth levels and all the distributions except

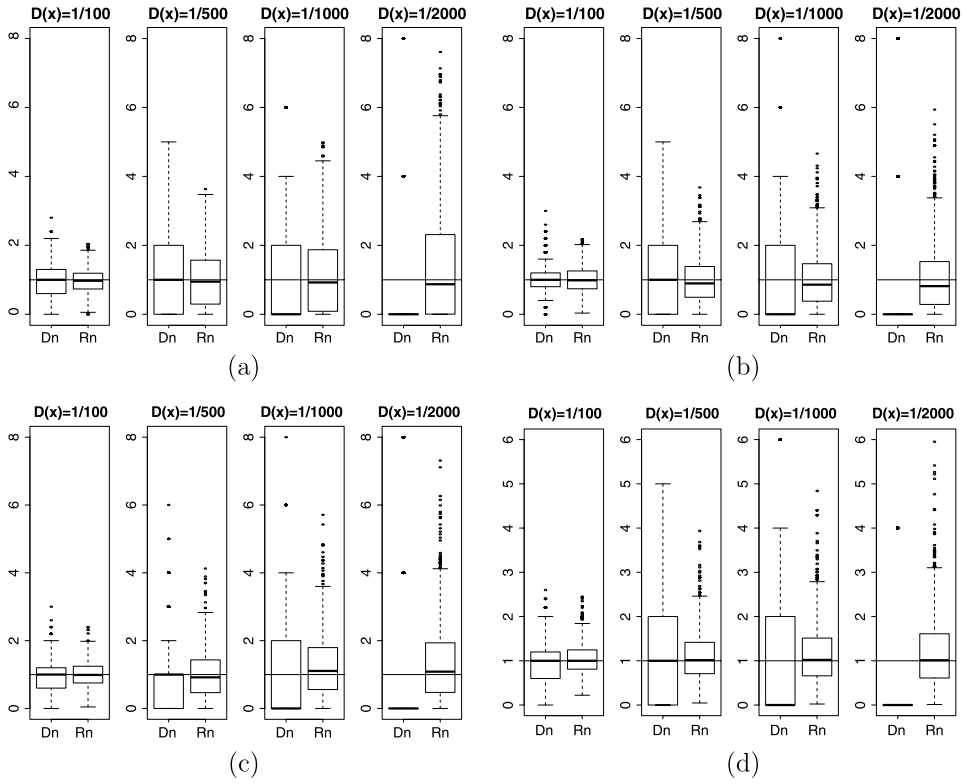

FIG. 2. *Comparison of D_n/D (left) and R_n/D (right) at 4 decreasing levels under* (a) *normal distribution*; (b) *Burr-type distribution*; (c) *t-distribution with 2 degrees of freedom*; (d) *Cauchy distribution*.

the bivariate normal distribution, the $R_n(\mathbf{x})/D(\mathbf{x})$ are all well centered at 1. In contrast, the original empirical halfspace depth D_n can only provide a reasonable estimate of D when D is not too small. When $D(\mathbf{x})$ is small relative to n, most of the $D_n(\mathbf{x})$ are zero. These results support the theoretical findings that R_n is a better estimator than D_n in the tail. For the bivariate normal distribution, although it does not satisfy the assumptions of Theorem 2, the performance of R_n is still much better than the performance of D_n.

3. Impact of the refinement of D_n on applications.

3.1. *Statistical process control.* In this section, we present two applications where R_n significantly improves the performance of the depth based procedures over D_n. The first one is statistical process control (SPC). SPC is the application of statistical methods to the monitoring of a process outcome in order to detect abnormal variations of the process from a specified in-control distribution. It has many applications in manufacturing processes. A typical setup for SPC is the following.

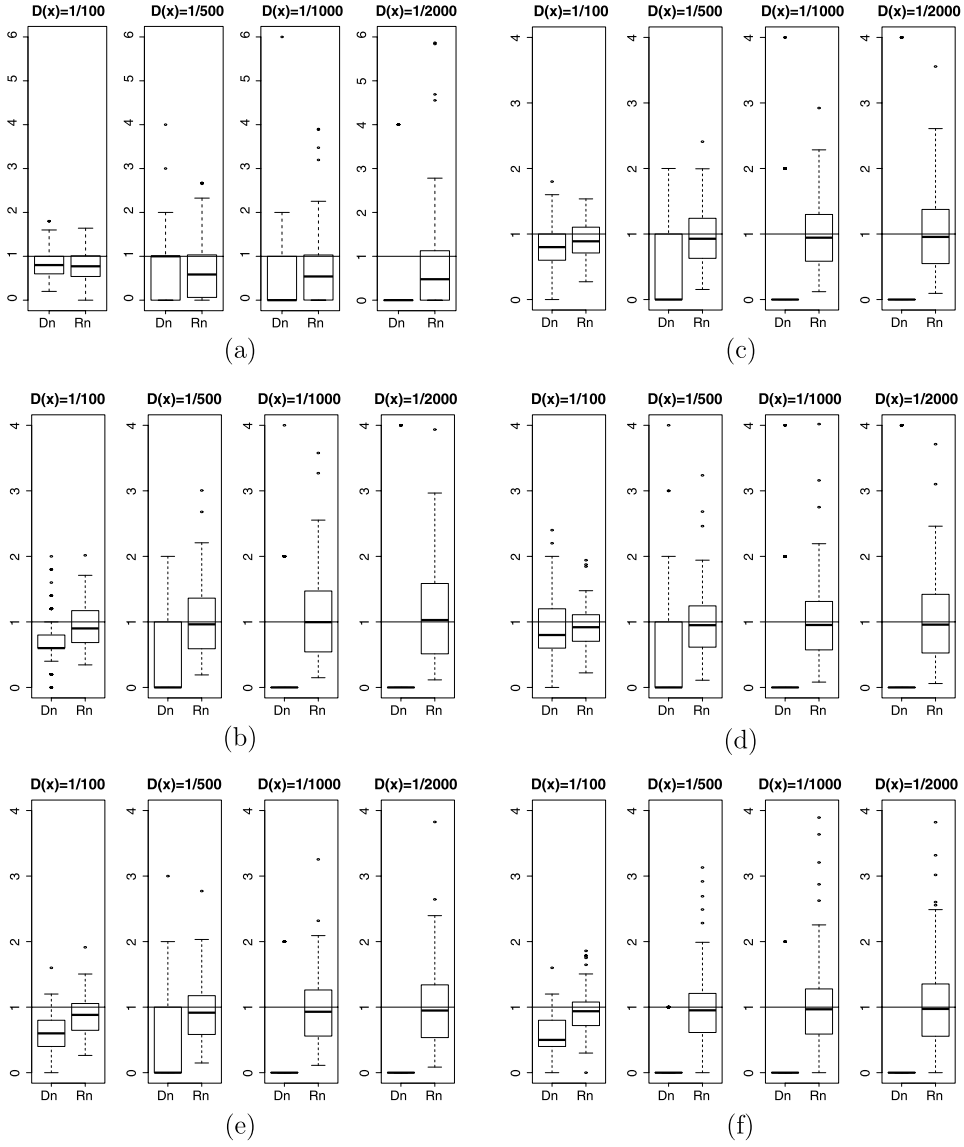

FIG. 3. *Comparison of D_n/D (left) and R_n/D (right) at* 4 *decreasing levels under* (a) *bivariate normal distribution*; (b) *bivariate spherical Cauchy distribution*; (c) *bivariate elliptical distribution*; (d) *bivariate clover distribution*; (e) *trivariate spherical Cauchy distribution*; (f) *quadrivariate spherical Cauchy distribution*.

There are n i.i.d. historical (reference) data for the monitored process outcome, denoted by $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ $(d \geq 1)$, from the in-control process. Let F_0 be the underlying distribution of the $\mathbf{X}_i$, also referred to as the in-control distribution. Let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ be future observations of the process outcome, under the distribution

F_1. The task of SPC is to determine if F_1 is the same as F_0 and if not, to signal when F_1 changes from F_0 as early as possible.

When the process outcome is multivariate and follows a multivariate normal distribution, an SPC procedure with a false alarm rate α can be defined as follows: $\mathbf{Y}_i$ is out of control if $T_i^2 > d(n+1)(n-1)/(n(n-d))F_{d,n-d}(\alpha)$, where $T_i^2 = (\mathbf{Y}_i - \bar{\mathbf{X}})'S^{-1}(\mathbf{Y}_i - \bar{\mathbf{X}})$, $\bar{\mathbf{X}} = \sum_{i=1}^n \mathbf{X}_i/n$, $S = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'/(n-1)$, and $F_{d,n-d}(\alpha)$ is the upper α quantile of an F distribution with d and $n-d$ degrees of freedom.

The above procedure requires that the process outcome follows a multivariate normal distribution. Therefore, we refer to it as the parametric SPC procedure hereafter. In many real world applications, the normality assumption may not hold. Therefore, a nonparametric SPC procedure is more desirable. Following Liu (1995), a nonparametric SPC procedure with a false alarm rate α can be defined as follows: $\mathbf{Y}_i$ is out of control if $\#\{\mathbf{X}_j : D(\mathbf{Y}_i) > D(\mathbf{X}_j), j = 1, \ldots, n\}/n < \alpha$, where D is the depth with respect to F_0. Since the in-control distribution is usually unknown in practice, D in the above procedure is usually replaced by D_n, the empirical depth with respect to the historical data, $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Due to its completely nonparametric nature and its capability of characterizing the geometric structure of the underlying distribution, the half-space depth is a popular choice in the above depth based SPC procedure. Because the future process outcomes $\mathbf{Y}_i$ that lie in the outskirts of the historical data are more of concern in this SPC procedure, how close the achieved false alarm rate to the nominal level α depends on how well the empirical half-space depth D_n estimates the theoretical half-space depth D for those points. As shown in this paper, this estimation is not satisfactory when n is not large enough. Therefore, the achieved false alarm rate can severely deviate from its nominal level α when D_n is used. To overcome this drawback of using D_n, we use our refined halfspace depth R_n in the above SPC procedure instead. Based on the results in Section 2, we expect the above depth based SPC procedure will achieve the nominal false alarm rate if R_n is used.

To demonstrate the performance of the R_n based SPC procedure, we carry out the following simulation. We first generate $n = 500$ historical data $\mathbf{X}_i$ from the standard bivariate normal distribution. We then generate another 5000 future observations $\mathbf{Y}_i$ from the same bivariate normal distribution. We apply to the 5000 $\mathbf{Y}_i$ the following three SPC procedures: the parametric procedure, the D_n based procedure and the R_n based procedure. We calculate R_n for the bivariate normal distribution as described in the previous section. The nominal false alarm rate α for each procedure is set to be at 0.0027 (the false alarm rate for the popular 3-sigma procedure in the univariate normal setting). The achieved false alarm rate for each procedure is then calculated as the proportion of $\mathbf{Y}_i$ being labeled as out-of-control by its SPC procedure. We repeat this simulation 100 times. The boxplots of the achieved false alarm rates from these 100 simulations for different SPC procedures are shown in Figure 4(a).

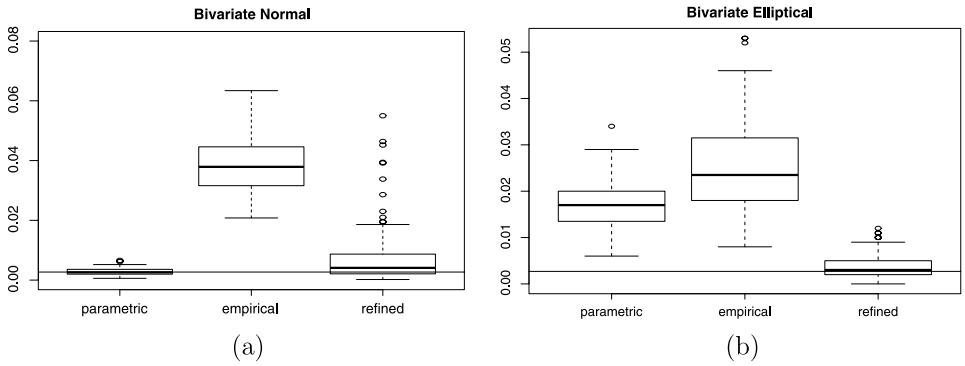

FIG. 4. *The achieved false alarm rates for the parametric procedure, the D_n based procedure and the R_n based procedure under* (a) *bivariate normal distribution*; (b) *bivariate elliptical distribution*.

As we can see from the plot, the parametric procedure can achieve the nominal false alarm rate as expected, since the normality assumption is satisfied in this case. In contrast, the achieved false alarm rate for the D_n based procedure is far higher than the target value 0.0027. It is not surprising since all the $\mathbf{Y}_i$ outside the convex hull of the $\mathbf{X}_i$ will have zero D_n and will be labeled as out-of-control, but some of those $\mathbf{Y}_i$ may have nonzero D and may have been labeled as in-control if D was used. From the plot, we can see that our R_n based procedure can successfully correct the inflated false alarm rate of the D_n based procedure and yields the false alarm rate near the target value 0.0027.

We run the same simulations as above on the data generated from the bivariate elliptical distribution of Section 2.3. Since the bivariate elliptical distribution satisfies the conditions of Theorem 2, here we use R_n based on (13). Figure 4(b) shows the corresponding boxplots of the achieved false alarm rates from 100 simulations for different SPC procedures. As seen from the plot, the parametric procedure can no longer achieve the nominal false alarm rate since the normality assumption does not hold in this case. The D_n based procedure still yields a far higher false alarm rate than the nominal level, while our R_n based procedure can achieve the nominal false alarm rate as expected.

To demonstrate the detection power of our R_n based procedure for process changes, we also carry out the following simulations. Similar to the above false alarm rate study, we first generate $n = 500$ historical data $\mathbf{X}_i$ from the standard bivariate normal distribution. We then generate 5000 future observations $\mathbf{Y}_i$ from another bivariate normal distribution mimicking the following three process changes: (i) location change from $(0, 0)$ to $(2, 2)$; (ii) scale increase from 1 to 2; (iii) both changes in (i) and (ii). Since the D_n based procedure fails to achieve the nominal false alarm rate, we only compare the detection power of the parametric procedure and our R_n based procedure. To benchmark the performance, we also include the procedure based on the theoretical D (D based procedure) in the comparison. In SPC, a common way to measure the detection power of SPC procedures

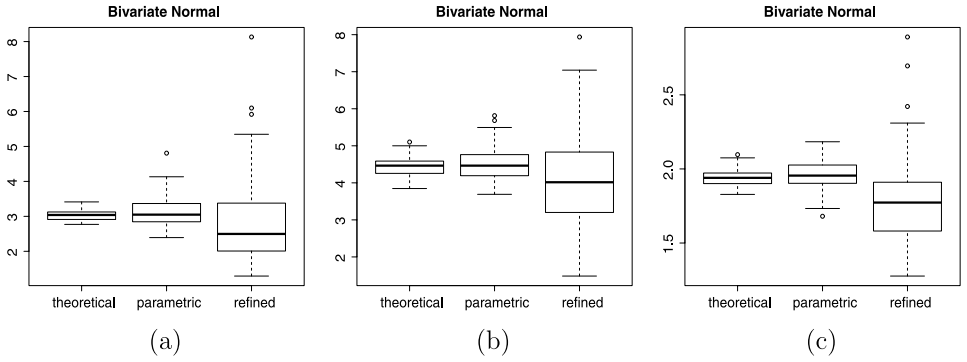

FIG. 5. *The achieved ARLs for the D based procedure, the parametric procedure and the R_n based procedure under bivariate normal distribution for* (a) *location change from* $(0, 0)$ *to* $(2, 2)$; (b) *scale increase from* 1 *to* 2; (c) *both changes in* (a) *and* (b).

is through the average run length (ARL). ARL is the expected number of times a process needs to be sampled until a specified change in the process is detected as out-of-control by the control chart in use. Figure 5 shows the boxplots of the ARLs from 100 simulations for the three procedures under the three process changes. As we can see from the plots, the parametric procedure and the D based procedure perform very similarly. Our R_n based procedure yields slightly smaller ARLs than the D based procedure. This can be explained by R_n's slightly larger false alarm rate than the nominal one in Figure 4(a).

We repeat the above ARL study on the data generated from the bivariate elliptical distribution. Similarly, we consider the following three process changes: (i) location change from $(0, 0)$ to $(4, 4)$; (ii) scale increase from 1 to 2; (iii) both changes in (i) and (ii). Since the parametric procedure does not achieve the nominal false alarm rate in this bivariate elliptical setting, we only compare the ARLs of the D based procedure and our R_n based procedure. Figure 6 shows the boxplots of ARLs of the two procedures under different process changes. As expected, our R_n based procedure performs well compared with the impractical procedure based on the unknown D.

3.2. *Classification.* Another application in which the refined half-space depth R_n helps improve the performance is the classification problem. Classification is one of the most practical subjects in statistics. It has many important applications in different fields. For simplicity, we only focus on two-class classification problem here. In this case, we observe two training samples $\{\mathbf{X}_1, \ldots, \mathbf{X}_m\}$ and $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$ from distributions F and G, respectively. The goal of the classification problem is to assign the future observation Z to either F or G based on some classification rule built on the two training samples. Recently Li, Cuesta-Albertos and Liu (2012) developed a nonparametric classification procedure, called *DD*-classifier, using the *DD*-plot (depth vs. depth plot) introduced in Liu, Parelius and Singh (1999). For

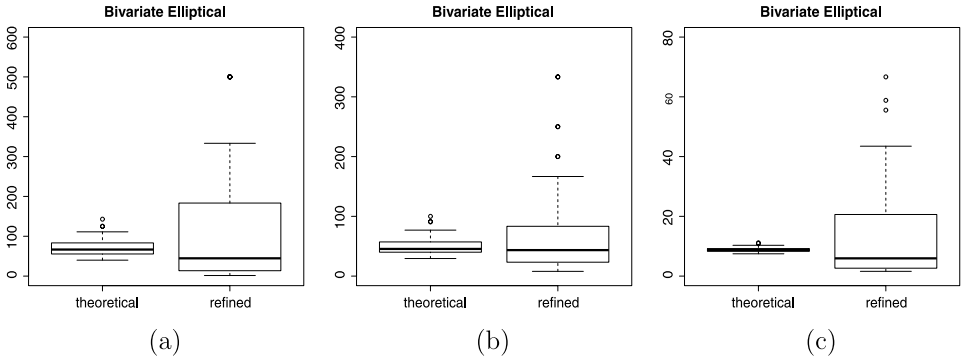

FIG. 6. *The achieved ARLs for the D based procedure and the R_n based procedure under bivariate elliptical distribution for* (a) *location change from* $(0, 0)$ *to* $(4, 4)$; (b) *scale increase from* 1 *to* 2; (c) *both changes in* (a) *and* (b).

any two samples, the *DD*-plot plots the depth values of those pooled sample points with respect to one sample against their depth values with respect to the other sample. The basic idea behind the *DD*-classifier is to look for a curve that best separates the two samples in their *DD*-plot. Since the best separating curve in the *DD*-classifier is required to pass through the origin in the *DD*-plot, any future observations having zero depth values with respect to both samples will be on the separating curve, indicating that they can be from either sample. Therefore, those observations will be randomly assigned to either sample. When the D_n of the half-space depth is used in constructing the *DD*-plot, any point which lies outside of the convex of both samples will have zero half-space depths with respect to both samples. Based on the *DD*-classifier, those points will be randomly assigned to either of the two samples, which will yield roughly a 50% misclassification rate for those points. This simply implies that when using D_n in the *DD*-classifier one loses all the information contained in those points. Next, we present a simulation study showing that the misclassification rate of those points can be improved by using R_n instead of D_n in the *DD*-classifier.

The first simulation setting we consider is when both F and G are bivariate normal distributions. We set F as the standard bivariate normal distribution, and G is another bivariate normal distribution which differs from F in (i) location; (ii) scale; (iii) both location and scale. (The location difference is 2 for both coordinates; the scale difference is also 2 for both coordinates.) For each of the three choices of G, we generate a training set consisting of $m = 500$ and $n = 500$ observations from F and G, respectively. Based on this training set, we obtain the linear *DD*-classifier using R_n to construct the *DD*-plot. Another 5000 test observations (2500 from each group) are then generated. Among those 5000 observations, the misclassification rate for the points which have zero D_n values with respect to both training samples are computed. This experiment is repeated 100 times and the

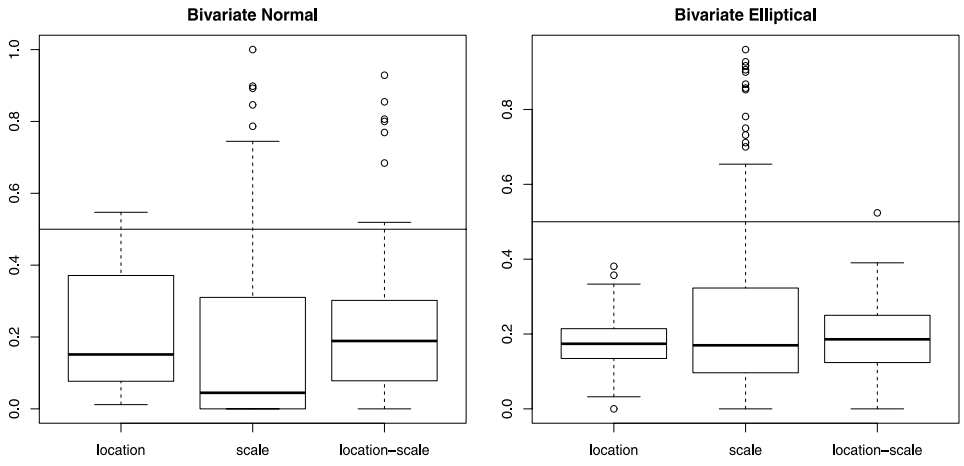

FIG. 7. *The misclassification rate based on R_n under* (a) *bivariate normal distribution*; (b) *bivariate elliptical distribution*.

misclassification rates for those points are then summarized in a boxplot for each choice of G in Figure 7(a).

We repeat this simulation on the data where both F and G are bivariate elliptical distributions; F corresponds to the elliptical density of Section 2.3. Again three kinds of differences are considered: (i) F and G differ in location; (ii) F and G differ in scale; (iii) F and G differ in both location and scale. (The location difference is 4 for both coordinates; the scale difference is 2 for both coordinates.) The boxplots of the misclassification rates for the test observations which have zero D_n values with respect to both training samples are shown in Figure 7(b).

As mentioned earlier, if D_n is used in the DD-classifier, the misclassification rate for the points which lie outside of the convex hull of both samples is roughly 50%. Therefore, as seen from Figure 7, the DD-classifier paired with R_n substantially improves the classification results for those points.

4. Concluding remarks. We have seen that both applications of the halfspace depth in SPC and classification gain substantially from the proposed refinement R_n. In general, we can expect similar gains from using R_n in statistical inference methods involving depth ranks or extreme depth contours, for example, determining p-values using depth in Liu and Singh (1997); constructing multivariate spacings and tolerance regions in Li and Liu (2008).

There are many other well-known depth functions [e.g., the spatial depth Chaudhuri (1996), the Mahalanobis depth Mahalanobis (1936), the projection depth Zuo (2003), etc.] which are not computed from the empirical distribution function, and hence they do not have the said problem in this paper. While these

depths are useful for many applications, they are either parametric in nature or lack of the needed distributional properties to ensure the desired probability masses associated with the central regions formed by the depth ranks or contours. When these properties are essential, the applications may be better served by using the two geometric depths. Case in point are the examples mentioned in the preceding paragraph. This in part explains the importance in refining the empirical half-space depth.

It is easy to see that the problem we faced in this paper stems from the use of the empirical distribution in computing the half-space probabilities. A natural solution then would be to consider instead a smoothed version of the empirical distribution that does not have point masses and is supported on the entire $\mathbb{R}^d$. It is worth noting that our proposed refinement is in fact such a smoothed version of the empirical distribution function in the tail, with the smoothing done by way of extreme value statistics. This extreme-value-theory based smoothing not only has the advantages of both breaking ties in the tail and yielding positive values, but, most importantly, it also produces a statistically much better estimator of the half-space depth in the tail, as shown in our theorems and applications.

It would be worthwhile to investigate whether the extreme-value-theory approach proposed in this paper can be modified to refine the empirical simplicial depth or other depth functions that also use the empirical counts based on the data. The modifications, if any, would seem quite nontrivial, since those depth functions do not have such a clear form of univariate projections as that of the half-space depth.

5. Proofs.

PROOF OF THEOREM 1. Write F^{-1} for the quantile function, the left-continuous inverse of F. We split the region over which the supremum is taken into three regions: $[F^{-1}(\delta_n), X_{k+1:n}]$, $(X_{k+1:n}, X_{n-k:n})$, and $[X_{n-k:n}, F^{-1}((1-\delta_n)^+)]$. Because of symmetry, the first and last region can be dealt with similarly. Therefore, we only consider the latter two regions.

For $x \in (X_{k+1:n}, X_{n-k:n})$, we easily see that

$$\begin{aligned} \min\left(\frac{F_n(x)}{F(x)}, \frac{S_n(x)}{S(x)}\right) &\le \frac{\min(F_n(x), S_n(x))}{\min(F(x), S(x))} = \frac{R_n(x)}{D(x)} \\ &\le \max\left(\frac{F_n(x)}{F(x)}, \frac{S_n(x)}{S(x)}\right). \end{aligned} \tag{18}$$

We have that

$$\sup_{x:F(x)\ge k/(2n)} \left|\frac{F_n(x)}{F(x)} - 1\right| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{x:S(x)\ge k/(2n)} \left|\frac{S_n(x)}{S(x)} - 1\right| \xrightarrow{p} 0; \tag{19}$$

see, for example, Shorack and Wellner [(1986), page 424]. Since $F(X_{k+1:n}) > \frac{k}{2n}$ and $F(X_{n-k:n}) < 1 - \frac{k}{2n}$ with probability tending to one ($n \to \infty$), it follows from (19) and (18) that

$$\sup_{X_{k+1:n} < x < X_{n-k:n}} \left| \frac{R_n(x)}{D(x)} - 1 \right| \xrightarrow{p} 0.$$

Hence, it remains to consider the supremum over the region $[X_{n-k:n}, F^{-1}((1-\delta_n)+)]$. We have with probability tending to one, as $n \to \infty$,

$$\begin{aligned} &\sup_{X_{n-k:n} \le x \le F^{-1}((1-\delta_n)+)} \left| \frac{R_n(x)}{D(x)} - 1 \right| \\ &\quad \le \sup_{\delta_n \le S(x) \le 2k/n, S(x) \ne k/n} \left| \frac{k}{nS(x)} \left(1 + \hat{\gamma} \frac{x - \hat{b}}{\hat{a}}\right)^{-1/\hat{\gamma}} - 1 \right| \\ &\qquad + \left| \left(1 + \hat{\gamma} \frac{b - \hat{b}}{\hat{a}}\right)^{-1/\hat{\gamma}} - 1 \right|. \end{aligned}$$

Write $B_n = \sqrt{k}(\hat{b} - b)/a$. Then we have $B_n = O_p(1)$; see, for example, Theorem 2.4.1 in de Haan and Ferreira (2006). Therefore, to complete the proof of this theorem it suffices to show

$$\sup_{\delta_n \le S(x) \le 2k/n, S(x) \ne k/n} \left| \frac{k}{nS(x)} \left(1 + \hat{\gamma} \frac{x - \hat{b}}{\hat{a}}\right)^{-1/\hat{\gamma}} - 1 \right| \xrightarrow{p} 0. \tag{20}$$

First, we consider the case $\gamma \ne 0$. Write $Y_n = \frac{\hat{\gamma}}{\gamma} \frac{a}{\hat{a}}$. Also, set

$$s = \frac{((x-b)/a)(\gamma/(d_n^{\gamma} - 1)) - 1}{A}$$

$$\text{with } d_n = d_n(x) = \frac{k}{nS(x)} \text{ and } A = A(n/k).$$

We have

$$\begin{aligned} &\frac{k}{nS(x)} \left(1 + \hat{\gamma} \frac{x - \hat{b}}{\hat{a}}\right)^{-1/\hat{\gamma}} \\ &\quad = d_n \left(1 + Y_n \left[\frac{x-b}{a}\gamma - \frac{\hat{b}-b}{a}\gamma\right]\right)^{-1/\hat{\gamma}} \\ &\quad = d_n \left(1 + Y_n \left[(1 + sA)(d_n^{\gamma} - 1) - \frac{\hat{b}-b}{a}\gamma\right]\right)^{-1/\hat{\gamma}} \\ &\quad = \left(d_n^{-\hat{\gamma}} + Y_n d_n^{-\hat{\gamma}} (1 + sA)(d_n^{\gamma} - 1) - \frac{\hat{b}-b}{a}\gamma Y_n d_n^{-\hat{\gamma}}\right)^{-1/\hat{\gamma}} \end{aligned}$$

$$= \left[d_n^{\gamma-\hat{\gamma}} \left(d_n^{-\gamma} \left[1 - Y_n(1+sA) - \frac{\hat{b}-b}{a} \gamma Y_n \right] + Y_n(1+sA) \right) \right]^{-1/\hat{\gamma}}$$
$$=: \left[T_1(T_2+T_3) \right]^{-1/\hat{\gamma}}.$$

We will now prove that $T_1 \xrightarrow{p} 1, T_2 \xrightarrow{p} 0, T_3 \xrightarrow{p} 1$, all *uniformly for x such that* $\delta_n \le S(x) \le 2k/n$ $[S(x) \ne k/n]$. This will yield (20) for $\gamma \ne 0$.

We have

$$T_1 = d_n^{\gamma-\hat{\gamma}} = d_n^{-\Gamma_n/\sqrt{k}} = \exp\left(\frac{-\Gamma_n}{\sqrt{k}} \log \frac{k}{nS(x)} \right).$$

Observe that

$$\left| \frac{-\Gamma_n}{\sqrt{k}} \log \frac{k}{nS(x)} \right| \le \frac{|\Gamma_n|}{\sqrt{k}} \left| \log \frac{k}{nS(x)} \right| \xrightarrow{p} 0.$$

Hence, $T_1 \xrightarrow{p} 1$. Consider $T_3 = Y_n(1+sA)$. We have $Y_n \xrightarrow{p} 1$ and $A(n/k) \to 0$. Hence, (6) yields $T_3 \xrightarrow{p} 1$. Finally,

$$T_2 = \frac{d_n^{-\gamma}}{\sqrt{k}} \sqrt{k} \left(1 - Y_n(1+sA) - \frac{B_n}{\sqrt{k}} \gamma Y_n \right)$$
$$= \frac{d_n^{-\gamma}}{\sqrt{k}} \left(\sqrt{k} \left[1 - \left(1 + O_p\left(\frac{1}{\sqrt{k}} \right) \right) \left(1 + O\left(\frac{1}{\sqrt{k}} \right) \right) \right] - B_n \gamma Y_n \right)$$
$$= \frac{(nS(x))^{\gamma}}{k^{\gamma+1/2}} O_p(1) = o_p(1).$$

Consider now the case $\gamma = 0$. By convention $(d_n^{\gamma} - 1)/\gamma = \log d_n$ now. Write

$$Q := \frac{k}{nS(x)} \left(1 + \hat{\gamma} \frac{x-\hat{b}}{\hat{a}} \right)^{-1/\hat{\gamma}}$$
$$= d_n \left(1 + \frac{a}{\hat{a}} \left[\frac{x-b}{a} \hat{\gamma} - \frac{\hat{b}-b}{a} \hat{\gamma} \right] \right)^{-1/\hat{\gamma}}$$
$$= d_n \left(1 + \hat{\gamma} \frac{a}{\hat{a}} (\log d_n)(1+sA) - \hat{\gamma} \frac{B_n}{\sqrt{k}} \frac{a}{\hat{a}} \right)^{-1/\hat{\gamma}}.$$

Hence,

$$\log Q = \log d_n - \frac{1}{\hat{\gamma}} \log \left(1 + \hat{\gamma} \frac{a}{\hat{a}} (\log d_n)(1+sA) - \hat{\gamma} \frac{B_n}{\sqrt{k}} \frac{a}{\hat{a}} \right).$$

We obtain

$$|\log Q| = \log\left(\frac{k}{n\delta_n} \right) \cdot O_p\left(\frac{1}{\sqrt{k}} \right) + \log^2\left(\frac{k}{n\delta_n} \right) \cdot O_p\left(\frac{1}{\sqrt{k}} \right) \xrightarrow{p} 0.$$

Hence, $Q \xrightarrow{p} 1$, uniformly for x such that $\delta_n \le S(x) \le 2k/n$ $(S(x) \ne k/n)$. This proves (20) for $\gamma = 0$. $\square$

For the proof of Theorem 2, we need two lemmas. In the sequel, we assume that the conditions of Theorem 2 are in force. Write $\Theta = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$ for the unit sphere.

LEMMA 1. *For all* $r > 0$,

$$\lim_{t\to\infty} \sup_{\mathbf{u}\in\Theta} \left| \frac{\mathbb{P}(\mathbf{X}_1 \in t H_{r,\mathbf{u}})}{t^{-\alpha}} - c\nu(H_{r,\mathbf{u}}) \right| = 0.$$

PROOF. Fix $r > 0$. Combining (9) and (10) we have that for all $\mathbf{u} \in \Theta$,

$$\lim_{t\to\infty} \frac{\mathbb{P}(\mathbf{X}_1 \in t H_{r,\mathbf{u}})}{t^{-\alpha}} = c\nu(H_{r,\mathbf{u}}). \tag{21}$$

Assume this convergence does not hold uniformly in $\mathbf{u} \in \Theta$. Then there exist sequences $\mathbf{u}_m \to \mathbf{v}$ and $t_m \to \infty$ such that

$$\mathbb{P}(\mathbf{X}_1 \in t_m H_{r,\mathbf{u}_m})/t_m^{-\alpha} \text{ does } \textit{not} \text{ converge to } c\nu(H_{r,\mathbf{v}}), \text{ as } m \to \infty. \tag{22}$$

W.l.o.g. we assume that $\mathbf{v} = (1, 0, \ldots, 0)$.

We show that (22) cannot hold by showing, for $\mathbf{u} \in \Theta$,

$$\frac{\mathbb{P}(\mathbf{u}^T\mathbf{X}_1 \ge tr)}{\mathbb{P}(X_{1,1} \ge tr)} \to 1 \qquad \text{if } u_1 \to 1, t \to \infty. \tag{23}$$

Because if the latter convergence holds, then if $u_1 \to 1, m \to \infty$,

$$\frac{\mathbb{P}(\mathbf{u}^T\mathbf{X}_1 \ge t_m r)}{t_m^{-\alpha}} = \frac{\mathbb{P}(\mathbf{u}^T\mathbf{X}_1 \ge t_m r)}{\mathbb{P}(X_{1,1} \ge t_m r)} \cdot \frac{\mathbb{P}(X_{1,1} \ge t_m r)}{t_m^{-\alpha}} \to 1 \cdot c\nu(H_{r,\mathbf{v}}). \tag{24}$$

Hence, it remains to show (23). Write $\varepsilon = 1 - u_1$. Then $\varepsilon \to 0$. We have

$$\begin{aligned}
&\mathbb{P}(\mathbf{u}^T\mathbf{X}_1 \ge tr) \\
&\quad = \mathbb{P}(\mathbf{u}^T\mathbf{X}_1 \ge tr, X_{1,1} < (1-\varepsilon^{1/4})tr) + \mathbb{P}(\mathbf{u}^T\mathbf{X}_1 \ge tr, X_{1,1} \ge (1-\varepsilon^{1/4})tr) \\
&\quad \le \mathbb{P}(\mathbf{u}^T\mathbf{X}_1 - (1-\varepsilon)X_{1,1} \ge \varepsilon^{1/4}tr) + \mathbb{P}(X_{1,1} \ge (1-\varepsilon^{1/4})tr) \\
&\quad \le \sum_{j=2}^{d} \mathbb{P}(u_j X_{1,j} \ge \varepsilon^{1/4}tr/(d-1)) + \mathbb{P}(X_{1,1} \ge (1-\varepsilon^{1/4})tr) \\
&\quad \le \sum_{j=2}^{d} \mathbb{P}(|X_{1,j}| \ge \varepsilon^{-1/4}tr/(\sqrt{2}(d-1))) + \mathbb{P}(X_{1,1} \ge (1-\varepsilon^{1/4})tr).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\frac{\mathbb{P}(\mathbf{u}^T\mathbf{X}_1 \ge tr)}{\mathbb{P}(X_{1,1} \ge tr)} \\
&\quad \le \frac{\sum_{j=2}^d \mathbb{P}(|X_{1,j}| \ge \varepsilon^{-1/4} tr/(\sqrt{2}(d-1)))}{t^{-\alpha}} \cdot \frac{t^{-\alpha}}{\mathbb{P}(X_{1,1} \ge tr)} \\
&\qquad + \frac{\mathbb{P}(X_{1,1} \ge (1-\varepsilon^{1/4})tr)}{t^{-\alpha}} \cdot \frac{t^{-\alpha}}{\mathbb{P}(X_{1,1} \ge tr)} \\
&\quad \to 0 \cdot \frac{1}{c\nu(H_{r,\mathbf{v}})} + c\nu(H_{r,\mathbf{v}}) \cdot \frac{1}{c\nu(H_{r,\mathbf{v}})} = 1.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\mathbb{P}\big(\mathbf{u}^T\mathbf{X}_1 \ge tr\big) \\
&\quad \ge \mathbb{P}\big(X_{1,1} \ge \big(1+\varepsilon^{1/4}\big)tr\big) - \mathbb{P}\big(\mathbf{u}^T\mathbf{X}_1 < tr, X_{1,1} \ge \big(1+\varepsilon^{1/4}\big)tr\big) \\
&\quad \ge \mathbb{P}\big(X_{1,1} \ge \big(1+\varepsilon^{1/4}\big)tr\big) - \mathbb{P}\big(\mathbf{u}^T\mathbf{X}_1 - (1-\varepsilon)X_{1,1} \le -\varepsilon^{1/4}tr/2\big)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\mathbb{P}(\mathbf{u}^T\mathbf{X}_1 \ge tr)}{\mathbb{P}(X_{1,1} \ge tr)} \\
&\quad \ge \frac{\mathbb{P}(X_{1,1} \ge (1+\varepsilon^{1/4})tr)}{\mathbb{P}(X_{1,1} \ge tr)} - \frac{\mathbb{P}(\mathbf{u}^T\mathbf{X}_1 - (1-\varepsilon)X_{1,1} \le -\varepsilon^{1/4}tr/2)}{\mathbb{P}(X_{1,1} \ge tr)} \\
&\quad \to c\nu(H_{r,\mathbf{v}}) \cdot \frac{1}{c\nu(H_{r,\mathbf{v}})} - 0 \cdot \frac{1}{c\nu(H_{r,\mathbf{v}})} = 1.
\end{aligned}$$

This completes the proof of (23). □

Define the function g by $g(\mathbf{u}) = c\nu(H_{1,\mathbf{u}})$ and let $V_{\mathbf{u}}(t) = F_{\mathbf{u}}^{-1}(1-1/t)$, $t > 1$, be the tail quantile function corresponding to $F_{\mathbf{u}}$.

LEMMA 2. *We have*

$$\lim_{t\to\infty} \sup_{\mathbf{u}\in\Theta} \left| \frac{V_{\mathbf{u}}(t)}{t^{1/\alpha}} - \big(g(\mathbf{u})\big)^{1/\alpha} \right| = 0. \tag{25}$$

PROOF. Lemma 1, with $r = 1$, yields

$$\lim_{s\to\infty} \sup_{\mathbf{u}\in\Theta} \left| \frac{1 - F_{\mathbf{u}}(s)}{s^{-\alpha}} - g(\mathbf{u}) \right| = 0. \tag{26}$$

Observe that $V_{\mathbf{u}}(t) = s$ implies $F_{\mathbf{u}}(s) = 1 - 1/t$. Hence

$$\frac{V_{\mathbf{u}}(t)}{t^{1/\alpha}} = s\big(1 - F_{\mathbf{u}}(s)\big)^{1/\alpha} = \left(\frac{1 - F_{\mathbf{u}}(s)}{s^{-\alpha}}\right)^{1/\alpha}. \tag{27}$$

Also observe that assumption (12) and $\nu(H_{1,\mathbf{u}}) \le 1$, $\mathbf{u} \in \Theta$, yield

$$0 < \inf_{\mathbf{u}\in\Theta} g(\mathbf{u}) \le \sup_{\mathbf{u}\in\Theta} g(\mathbf{u}) \le c < \infty.$$

Combining this with (27) and (26) easily yields (25). $\square$

PROOF OF THEOREM 2. We will prove that, as $n \to \infty$,

$$\sup_{D(\mathbf{x})\ge k/(2n)} \left| \frac{D_n(\mathbf{x})}{D(\mathbf{x})} - 1 \right| \xrightarrow{p} 0 \quad \text{and} \tag{28}$$

$$\sup_{R_n(\mathbf{x})<k/n, D(\mathbf{x})\ge\delta_n} \left| \frac{R_n(\mathbf{x})}{D(\mathbf{x})} - 1 \right| \xrightarrow{p} 0. \tag{29}$$

To show that (28) and (29) imply (15), it is sufficient to show that (28) implies

$$\sup_{D_n(\mathbf{x})\ge k/n} \left| \frac{D_n(\mathbf{x})}{D(\mathbf{x})} - 1 \right| \xrightarrow{p} 0 \tag{30}$$

and to recall that if $D_n(\mathbf{x}) \ge k/n$ or $R_n(\mathbf{x}) \ge k/n$, then $D_n(\mathbf{x}) = R_n(\mathbf{x})$.

Assume (28) holds. It follows from Donoho and Gasko (1992), that $\sup_{\mathbf{x}} D_n(\mathbf{x}) \ge 1/(d+1)$, with probability 1. Hence, for large n, any point $\hat{\mathbf{x}}$ with maximum depth D_n, satisfies $D_n(\hat{\mathbf{x}}) \ge k/n$ and, with probability tending to one, $D(\hat{\mathbf{x}}) \ge k/n$, because of the uniform consistency of D_n. Now assume for some $\mathbf{x}$, $D_n(\mathbf{x}) \ge k/n$ and $D(\mathbf{x}) < k/(2n)$. Then, with probability tending to one, we can find $\mathbf{x}_0$ on the straight line connecting $\hat{\mathbf{x}}$ and $\mathbf{x}$, such that $D(\mathbf{x}_0) = k/(2n)$ and because of (28), $D_n(\mathbf{x}_0) \le 3k/(4n)$. It is well known that D_n has the "monotonicity relative to deepest point" property [see, e.g., Zuo and Serfling (2000)], and hence $D_n(\mathbf{x}) \le D_n(\mathbf{x}_0) \le 3k/(4n)$. Contradiction. Hence (30).

It remains to prove (28) and (29). We begin with (28). First, we show that

$$P\Big(\bigcup\{H : P(H) \le s\}\Big) = O(s) \qquad \text{as } s \downarrow 0. \tag{31}$$

Define $r_0 = (c \inf_{\mathbf{u}\in\Theta} \nu(H_{1,\mathbf{u}})/2)^{1/\alpha}$. Lemma 1 yields that, uniformly in $\mathbf{u} \in \Theta$,

$$\lim_{s\downarrow 0} \frac{\mathbb{P}(\mathbf{X}_1 \in s^{-1/\alpha} H_{r_0,\mathbf{u}})}{s} = c\nu(H_{r_0,\mathbf{u}}) = c r_0^{-\alpha} \nu(H_{1,\mathbf{u}}) \ge 2.$$

Hence, for small enough s and uniformly in $\mathbf{u} \in \Theta$, $\mathbb{P}(\mathbf{X}_1 \in s^{-1/\alpha} H_{r_0,\mathbf{u}}) > s$. For $\mathbf{u} \in \Theta$, let r_1 be the smallest r such that $P(H_{r,\mathbf{u}}) = s$. Then for small enough s, $H_{r_1,\mathbf{u}} \subset s^{-1/\alpha} H_{r_0,\mathbf{u}}$. Hence, by (9) and (10),

$$\frac{P(\bigcup_{\mathbf{u}\in\Theta} H_{r_1,\mathbf{u}})}{s} \le \frac{P(\bigcup_{\mathbf{u}\in\Theta} s^{-1/\alpha} H_{r_0,\mathbf{u}})}{s} = \frac{P(s^{-1/\alpha} \bigcup_{\mathbf{u}\in\Theta} H_{r_0,\mathbf{u}})}{s}$$

$$\to c\nu\Big(\bigcup_{\mathbf{u}\in\Theta} H_{r_0,\mathbf{u}}\Big) < \infty, \qquad s \downarrow 0,$$

which implies (31).

Using (31), we obtain from Theorem 5.1 in Alexander (1987), with the γ_n there equal to $k/(2n)$, that

$$\sup_{H:P(H)\geq k/(2n)} \left| \frac{P_n(H)}{P(H)} - 1 \right| \xrightarrow{p} 0 \qquad \text{as } n \to \infty. \tag{32}$$

Denote with $H_{\mathbf{x}}$ a half-space with $\mathbf{x}$ on its boundary. We have

$$\frac{D_n(\mathbf{x})}{D(\mathbf{x})} = \frac{\inf_{H_{\mathbf{x}}} P_n(H_{\mathbf{x}})}{\inf_{H_{\mathbf{x}}} P(H_{\mathbf{x}})} \geq \inf_{H_{\mathbf{x}}} \frac{P_n(H_{\mathbf{x}})}{P(H_{\mathbf{x}})}$$

and, with $\varepsilon > 0$, for some $H_{\mathbf{x}}$,

$$\frac{D_n(\mathbf{x})}{D(\mathbf{x})} \leq (1+\varepsilon) \frac{P_n(H_{\mathbf{x}})}{P(H_{\mathbf{x}})}.$$

This, in combination with (32), yields (28).

Finally, we consider (29). Write $p_{\mathbf{u}}(w) = \mathbb{P}(\mathbf{u}^T \mathbf{X}_1 \geq w) = P(H_{w,\mathbf{u}})$. We first show

$$\sup_{w,\mathbf{u}\in\Theta:\delta_n \leq p_{\mathbf{u}}(w) \leq 2k/n} \left| \frac{p_{n,\mathbf{u}}(w)}{p_{\mathbf{u}}(w)} - 1 \right| \xrightarrow{p} 0 \qquad \text{as } n \to \infty. \tag{33}$$

Write $d_{\mathbf{u}}(w) = k/(np_{\mathbf{u}}(w))$. Then

$$\begin{aligned} \frac{p_{n,\mathbf{u}}(w)}{p_{\mathbf{u}}(w)} &= \frac{k}{np_{\mathbf{u}}(w)} \left(\frac{w}{V_{\mathbf{u}}(n/k)} \frac{V_{\mathbf{u}}(n/k)}{W_{n-k:n}} \right)^{-\widehat{\alpha}} \\ &= \left(d_{\mathbf{u}}^{-1/\widehat{\alpha}}(w) \frac{w}{V_{\mathbf{u}}(n/k)} \right)^{-\widehat{\alpha}} \left(\frac{V_{\mathbf{u}}(n/k)}{W_{n-k:n}} \right)^{-\widehat{\alpha}}. \end{aligned} \tag{34}$$

It follows from Lemmas 1 and 2 that

$$\lim_{n\to\infty} \sup_{w,\mathbf{u}\in\Theta:\delta_n \leq p_{\mathbf{u}}(w) \leq 2k/n} \left| d_{\mathbf{u}}^{-1/\alpha}(w) \frac{w}{V_{\mathbf{u}}(n/k)} - 1 \right| = 0.$$

Using this, (14) and $\log(n\delta_n)/\sqrt{k} \to 0$, we obtain

$$\sup_{w,\mathbf{u}\in\Theta:\delta_n \leq p_{\mathbf{u}}(w) \leq 2k/n} \left| \left(d_{\mathbf{u}}^{-1/\widehat{\alpha}}(w) \frac{w}{V_{\mathbf{u}}(n/k)} \right)^{-\widehat{\alpha}} - 1 \right| \xrightarrow{p} 0, \qquad \text{as } n \to \infty. \tag{35}$$

Denote with $G_{\mathbf{u},n}$ the empirical distribution function of the uniform-$(0,1)$ random variables $F_u(\mathbf{u}^T \mathbf{X}_i)$, $i = 1, \ldots, n$, and with $G_{\mathbf{u},n}^{-1}$ the corresponding quantile function. It follows from (32) by routine arguments that

$$\sup_{\mathbf{u}\in\Theta} \left| \frac{1 - G_{\mathbf{u},n}^{-1}(1-k/n)}{k/n} - 1 \right| \xrightarrow{p} 0,$$

and hence, by Lemma 2 and (14), that

$$\sup_{\mathbf{u}\in\Theta}\left|\left(\frac{V_{\mathbf{u}}(n/k)}{W_{n-k:n}}\right)^{-\widehat{\alpha}}-1\right|\xrightarrow{p}0. \tag{36}$$

Combination of (34), (35) and (36), yields (33).

Now we turn to (29). Let $\mathbf{x}$ be such that $R_n(\mathbf{x})<k/n$. Then

$$\begin{aligned}\frac{R_n(\mathbf{x})}{D(\mathbf{x})}&=\frac{\inf_{\mathbf{u}\in\Theta:1-\widehat{F}_{\mathbf{u}}(\mathbf{u}^T\mathbf{x}^-)<k/n}1-\widehat{F}_{\mathbf{u}}(\mathbf{u}^T\mathbf{x}^-)}{\inf_{\mathbf{u}\in\Theta}p_{\mathbf{u}}(\mathbf{u}^T\mathbf{x})}\\&\geq\frac{\inf_{\mathbf{u}\in\Theta:p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})<k/n}p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})}{\inf_{\mathbf{u}\in\Theta:p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})<k/n}p_{\mathbf{u}}(\mathbf{u}^T\mathbf{x})}\geq\inf_{\mathbf{u}\in\Theta:p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})<k/n}\frac{p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})}{p_{\mathbf{u}}(\mathbf{u}^T\mathbf{x})}.\end{aligned} \tag{37}$$

Next, we show that with probability tending to one ($n\to\infty$),

$$\inf_{\mathbf{u}\in\Theta:p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})<k/n}\frac{p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})}{p_{\mathbf{u}}(\mathbf{u}^T\mathbf{x})}\geq\inf_{\mathbf{u}\in\Theta:p_{\mathbf{u}}(\mathbf{u}^T\mathbf{x})\leq2k/n}\frac{p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})}{p_{\mathbf{u}}(\mathbf{u}^T\mathbf{x})}. \tag{38}$$

Assume for some $\mathbf{x}$ and $\mathbf{u}\in\Theta$, $p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})<k/n$ and $p_{\mathbf{u}}(\mathbf{u}^T\mathbf{x})>2k/n$. Then there exists an $\mathbf{x}_0$ of the form $\mathbf{x}+\tilde{c}\mathbf{u}$, for some $\tilde{c}>0$, such that $p_{\mathbf{u}}(\mathbf{u}^T\mathbf{x}_0)=2k/n$. Hence, with probability tending to one because of (33), $p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x}_0)\geq3k/(2n)$ and, therefore, $p_{n,\mathbf{u}}(\mathbf{u}^T\mathbf{x})\geq3k/(2n)$. Contradiction. Hence, we have (38). Combining (38) with (37) and (33), yields

$$\sup_{R_n(\mathbf{x})<k/n,D(\mathbf{x})\geq\delta_n}\left(1-\frac{R_n(\mathbf{x})}{D(\mathbf{x})}\right)\vee0\xrightarrow{p}0. \tag{39}$$

Let $\varepsilon\in(0,1)$ and let $\mathbf{x}$ be such that $R_n(\mathbf{x})<k/n$ and $D(\mathbf{x})\geq\delta_n$. We have for some $\mathbf{u}_0$ that

$$\frac{R_n(\mathbf{x})}{D(\mathbf{x})}\leq(1+\varepsilon/2)\frac{R_n(x)}{p_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x})}\leq(1+\varepsilon/2)\frac{1-\widehat{F}_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x}^-)}{p_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x})},$$

with $p_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x})\geq\delta_n$. If $p_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x})\leq k/(2n)$, then with probability tending to one, (33) yields that $1-\widehat{F}_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x}^-)=p_{n,\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x})$, and hence that $R_n(\mathbf{x})/D(\mathbf{x})\leq1+\varepsilon$. In case $p_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x})>k/(2n)$, we have, using (39), that with probability tending to one that $k/(2n)<p_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x})\leq3D(\mathbf{x})/2\leq2R_n(\mathbf{x})<2k/n$. Hence, combining $1-\widehat{F}_{\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x}^-)\leq(1-F_{n,\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x}^-))\vee p_{n,\mathbf{u}_0}(\mathbf{u}_0^T\mathbf{x})$ with (32) and (33), we obtain that with probability tending to one, $R_n(\mathbf{x})/D(\mathbf{x})\leq1+\varepsilon$. Hence, we have shown

$$\sup_{R_n(\mathbf{x})<k/n,D(\mathbf{x})\geq\delta_n}\left(\frac{R_n(\mathbf{x})}{D(\mathbf{x})}-1\right)\vee0\xrightarrow{p}0.$$

This, in combination with (39), yields (29). □

Acknowledgements. We thank an Associate Editor and three referees for many insightful questions and comments, which helped improve greatly this paper. We also thank Laurens de Haan for his helpful discussions, in particular about the univariate second-order condition.

REFERENCES

ALEXANDER, K. S. (1987). Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probab. Theory Related Fields* **75** 379–423. MR0890285

CAI, J.-J., EINMAHL, J. H. J. and DE HAAN, L. (2011). Estimation of extreme risk regions under multivariate regular variation. *Ann. Statist.* **39** 1803–1826. MR2850221

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872. MR1395753

CUESTA-ALBERTOS, J. A. and NIETO-REYES, A. (2008). The random Tukey depth. *Comput. Statist. Data Anal.* **52** 4979–4988. MR2526207

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An Introduction*. Springer, New York. MR2234156

DEKKERS, A. L. M., EINMAHL, J. H. J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** 1833–1855.

DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. MR1193313

EFRON, B. (1965). The convex hull of a random set of points. *Biometrika* **52** 331–343. MR0207004

EINMAHL, J. H. J. and KRAJINA, A. (2015). Empirical likelihood based testing for multivariate regular variation. Preprint.

HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204

JESSEN, A. H. and MIKOSCH, T. (2006). Regularly varying functions. *Publ. Inst. Math.* (*Beograd*) (*N.S.*) **80** 171–192. MR2281913

LI, J., CUESTA-ALBERTOS, J. A. and LIU, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *J. Amer. Statist. Assoc.* **107** 737–753. MR2980081

LI, J. and LIU, R. Y. (2004). New nonparametric tests of multivariate locations and scales using data depth. *Statist. Sci.* **19** 686–696. MR2185590

LI, J. and LIU, R. Y. (2008). Multivariate spacings based on data depth. I. Construction of nonparametric multivariate tolerance regions. *Ann. Statist.* **36** 1299–1323. MR2418658

LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400

LIU, R. Y. (1995). Control charts for multivariate processes. *J. Amer. Statist. Assoc.* **90** 1380–1387. MR1379481

LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. MR1724033

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. MR1212489

LIU, R. Y. and SINGH, K. (1997). Notions of limiting P values based on data depth and bootstrap. *J. Amer. Statist. Assoc.* **92** 266–277. MR1436115

MAHALANOBIS, P. (1936). On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India* **12** 49–55.

MEERSCHAERT, M. M. and SCHEFFLER, H.-P. (2001). *Limit Distributions for Sums of Independent Random Vectors*: *Heavy Tails in Theory and Practice*. Wiley, New York. MR1840531

ROUSSEEUW, P. J. and HUBERT, M. (1999). Regression depth. *J. Amer. Statist. Assoc.* **94** 388–433. MR1702314

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York. MR0838963

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (*Vancouver, B.C.*, 1974), *Vol.* 2 523–531. Canad. Math. Congress, Montreal, Que. MR0426989

YEH, A. B. and SINGH, K. (1997). Balanced confidence regions based on Tukey's depth and the bootstrap. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **59** 639–652. MR1452031

ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. MR2012822

ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005

J. H. J. EINMAHL
DEPARTMENT OF ECONOMETRICS AND CENTER
TILBURG UNIVERSITY
P.O. BOX 90153
5000 LE TILBURG
THE NETHERLANDS
E-MAIL: j.h.j.einmahl@uvt.nl

J. LI
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, RIVERSIDE
RIVERSIDE, CALIFORNIA 92521
USA
E-MAIL: jun.li@ucr.edu

R. Y. LIU
DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY 08854
USA
E-MAIL: rliu@stat.rutgers.edu