ESTIMATION OF EXTREME RISK REGIONS UNDER MULTIVARIATE REGULAR VARIATION

By Juan-Juan Cai, John H. J. Einmahl and Laurens de Haan[1]

Tilburg University, Tilburg University and Lisbon University

When considering d possibly dependent random variables, one is often interested in extreme risk regions, with very small probability p. We consider risk regions of the form $\{\mathbf{z} \in \mathbb{R}^d : f(\mathbf{z}) \leq \beta\}$, where f is the joint density and β a small number. Estimation of such an extreme risk region is difficult since it contains hardly any or no data. Using extreme value theory, we construct a natural estimator of an extreme risk region and prove a refined form of consistency, given a random sample of multivariate regularly varying random vectors. In a detailed simulation and comparison study, the good performance of the procedure is demonstrated. We also apply our estimator to financial data.

1. Introduction. A two-dimensional normal density or Student t-density is constant on boundaries of certain ellipses. Outside such an ellipse the density is lower than inside. It is straightforward to find such an outer region and its contour (line), for a given small probability. We can consider such contour as a natural multidimensional extension of a (one-dimensional) quantile. Even for extreme sets, that is, very low density levels, the calculations are straightforward.

In this paper we consider, much more general, *multi*variate regularly varying distributions [for a review, see Jessen and Mikosch (2006)]. We consider the latter distributions, since we want to explore in particular extreme sets, that is, sets far removed from the origin. A random vector $\mathbf{X}$ is multivariate regularly varying if there exist a constant $\alpha > 0$, the index and an arbitrary probability measure Ψ on $\Theta = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| = 1\}$, the unit hypersphere, such

that

$$\text{(1)} \qquad \lim_{t \to \infty} \frac{\mathbb{P}(\|\mathbf{X}\| \geq tx, \mathbf{X}/\|\mathbf{X}\| \in A)}{\mathbb{P}(\|\mathbf{X}\| \geq t)} = x^{-\alpha} \Psi(A)$$

for every $x > 0$ and Borel set A in Θ with $\Psi(\partial A) = 0$, with $\|\mathbf{X}\|$ the L_2-norm of $\mathbf{X}$; see Rvačeva (1962). An equivalent statement is

$$\text{(2)} \qquad \lim_{t \to \infty} \frac{\mathbb{P}(\|\mathbf{X}\| \geq tx)}{\mathbb{P}(\|\mathbf{X}\| \geq t)} = x^{-\alpha} \qquad \text{for } x > 0,$$

and there exists a measure ν such that

$$\text{(3)} \qquad \lim_{t \to \infty} \frac{\mathbb{P}(\mathbf{X} \in tB)}{\mathbb{P}(\|\mathbf{X}\| \geq t)} = \nu(B) < \infty$$

for every Borel set B on $\mathbb{R}^d$ that is bounded away from the origin and satisfies $\nu(\partial B) = 0$; here $tB = \{t\mathbf{z} : \mathbf{z} \in B\}$. Note that ν is homogeneous, that is, for all $a > 0$,

$$\text{(4)} \qquad \nu(aB) = a^{-\alpha} \nu(B).$$

Clearly, on $\{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| \geq 1\}$, ν is a *probability* measure. The limit relation in (3) is a multivariate analogue of the "peaks-over-threshold" or "generalized Pareto limit" method in one-dimensional extreme value theory. Particular cases of (1) are distributions in the sum-domain of attraction of α-stable distributions and heavy tailed elliptical distributions such as multivariate t-distributions [see Hashorva (2006)].

We require the convergence in (2) and (3) at the density level:

(a) Suppose that the distribution of $\mathbf{X}$ has a continuous and positive density f and that for some positive function q and some positive function V regularly varying at infinity with negative index $-\alpha$, we have

$$\text{(5)} \qquad \lim_{t \to \infty} \frac{f(t\mathbf{z})}{t^{-d} V(t)} = q(\mathbf{z}) \qquad \text{for all } \mathbf{z} \neq \mathbf{0}$$

and

$$\text{(6)} \qquad \lim_{t \to \infty} \sup_{\mathbf{z} \in \Theta} \left| \frac{f(t\mathbf{z})}{t^{-d} V(t)} - q(\mathbf{z}) \right| = 0.$$

Then q is continuous on $\mathbb{R}^d \setminus \{\mathbf{0}\}$ and $q(a\mathbf{z}) = a^{-d-\alpha} q(\mathbf{z})$ for all $a > 0$ and $\mathbf{z} \neq \mathbf{0}$. Throughout, *we can and will take* $V(t) = \mathbb{P}(\|\mathbf{X}\| > t)$ (see Lemma 1, Section 5). From Lemma 1, it follows that doing so (3) holds with $\nu(B) = \int_B q(\mathbf{z}) \, d\mathbf{z}$.

The extreme region will be of the form

$$Q = \{\mathbf{z} \in \mathbb{R}^d : f(\mathbf{z}) \leq \beta\},$$

where f is the probability density of the random vector $\mathbf{X}$; β is determined in such a way that the probability of Q is equal to a given very small number p, like $1/10{,}000$.

It is the purpose of this paper to estimate Q based on n i.i.d. copies of $\mathbf{X}$. Note that the shape of Q is not predetermined, it depends on the density f. For the estimation of Q, we will use an approximation of f based on the density of Ψ. The values of p we consider are typically of order $1/n$. This means that the number of data points that fall in Q is small and can even be zero, that is, we are extrapolating outside the sample. This lack of relevant data points makes estimation difficult. The estimation of Q is a multivariate analogue of the estimation of extreme quantiles in the univariate setting; see, for example, de Haan and Ferreira (2006), Chapter 4. The multivariate case is much more complicated, however, since we have to estimate a whole set instead of only one value.

Having an estimate of Q can be important in various settings. It can be used as an alarm system in risk management: if a new observation falls in the estimated Q it is a signal of extreme risk. See Einmahl, Li and Liu (2009) for an application to aviation safety along these lines. In a financial or insurance setting, points on the boundary of the estimate of Q can be used for stress testing. The estimate of Q can also be used to rank extreme observations (see Remark 3, Section 2).

For the "central" part of the distribution, that is, β is fixed (and "not too small"), nonparametric estimation of density level sets has been studied in depth in the literature. Two approaches are used, the plug-in approach using density estimation [see Baíllo, Cuesta-Albertos and Cuevas (2001) and Rigollet and Vert (2009)], and the excess mass approach [see Müller and Sawitzki (1991), Polonik (1995) and Tsybakov (1997)]. Our estimation problem and (hence) our approach are quite different from these.

This paper is organized as follows. In Section 2, we derive our estimator and show a refined form of consistency. A simulation and comparison study is presented in Section 3 and a financial application is given in Section 4. Section 5 contains the proof of the main result.

2. Main result. Consider a random sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ with $\mathbf{X}_i \stackrel{d}{=} \mathbf{X}$, for $i = 1, \ldots, n$; their common probability measure on $\mathbb{R}^d$ is denoted with P. Write R_i for the radius $\|\mathbf{X}_i\|$ and $\mathbf{W}_i$ for the direction $\mathbf{X}_i/\|\mathbf{X}_i\|$ of $\mathbf{X}_i$. We wish to estimate an extreme risk region of the form

$$Q = \{\mathbf{z} \in \mathbb{R}^d : f(\mathbf{z}) \le \beta\},$$

where β is such that $PQ = p > 0$, where $p = p_n \to 0$, as $n \to \infty$. This means that Q and β depend on n, that is, $Q = Q_n$ and $\beta = \beta_n$. We shall connect Q_n to a fixed set S not depending on n, defined by

$$S = \{\mathbf{z} : q(\mathbf{z}) \le 1\}.$$

It will turn out that Q_n can be approximated by a properly inflated version of S. In fact, it follows from (6) that the risk regions are asymptotically homothetic as a function of p, for small values of p. Define $H(s) = 1 - V(s) = \mathbb{P}(R \leq s)$ and $U(t) = H^{-1}(1 - \frac{1}{t})$. Note that U is regularly varying at infinity with index $1/\alpha$.

We will approximate Q_n in two steps by a (deterministic) region $\widetilde{Q}_n$. This approximation satisfies

$$(7) \qquad \frac{P(Q_n \triangle \widetilde{Q}_n)}{p} \to 0$$

($\triangle$ denotes "symmetric difference") and is based on the above limit relations. The region $\widetilde{Q}_n$ can therefore be estimated using extreme value theory. The first step is to establish an approximation of $\beta = \beta(p)$. Let

(b) $k = k_n (< n)$ be a sequence of positive integers such that $k \to \infty$ and $k/n \to 0$.

The region Q_n is approximated by

$$\bar{Q}_n = \left\{ \mathbf{z} : f(\mathbf{z}) \leq \left(\frac{np}{k\nu(S)} \right)^{1+d/\alpha} \frac{1}{(n/k)(U(n/k))^d} \right\}.$$

Next, we approximate $\bar{Q}_n$ by a further region $\widetilde{Q}_n$ defined in terms of the limit density q rather than f:

$$(8) \qquad \widetilde{Q}_n = U\left(\frac{n}{k} \right) \left(\frac{k\nu(S)}{np} \right)^{1/\alpha} S.$$

Indeed, S and this approximation of Q_n are homothetic.

Write

$$B_{r,A} = \{ \mathbf{z} : \|\mathbf{z}\| \geq r, \mathbf{z}/\|\mathbf{z}\| \in A \}$$

for a Borel set A on Θ. Clearly, $B_{r,A} = rB_{1,A}$ and hence $\nu(B_{r,A}) = r^{-\alpha}\nu(B_{1,A})$. The relation between the spectral measure Ψ and ν is [cf. (1) and (3)]

$$\Psi(A) = \nu(B_{1,A})$$

for a Borel set $A \subset \Theta$. Recall that the spectral measure is a probability measure. The existence of a density q of ν implies the existence of a density ψ of Ψ, that is,

$$\Psi(A) = \int_A \psi(\mathbf{w}) \, d\lambda(\mathbf{w}),$$

where λ is the Hausdorff measure (surface area) on Θ and

$$q(r\mathbf{w}) = \alpha r^{-\alpha-d}\psi(\mathbf{w}).$$

Next, we write S and $\nu(S)$ in terms of the spectral density:

$$S = \{\mathbf{z} = r\mathbf{w} : r \geq (\alpha\psi(\mathbf{w}))^{1/(\alpha+d)}, \mathbf{w} \in \Theta\}$$

and hence

$$\nu(S) = \alpha^{-\alpha/(\alpha+d)} \int_\Theta (\psi(\mathbf{w}))^{d/(\alpha+d)} \, d\lambda(\mathbf{w}).$$

To estimate $\widetilde{Q}_n$, we need estimators for $U(n/k), \alpha, S$ and $\nu(S)$. From the above expressions for S and $\nu(S)$, we see that this means that we have to estimate $U(n/k)$, α and ψ. First, we define

$$\widehat{U}\left(\frac{n}{k}\right) = R_{n-k:n}$$

[the $(n-k)$th order statistic of the $R_i, \; i = 1, \ldots, n$]. Since the tail of the distribution function of R is regularly varying with index $-\alpha$, we can use one of the well-known estimators of the extreme value index $1/\alpha$, based on the $R_i, i = 1, \ldots, n$; see, for example, Hill (1975), Smith (1987) and Dekkers, Einmahl and de Haan (1989). It remains to estimate ψ. Let $K : [0,1] \to [0,1]$ be a continuous and nonincreasing (kernel) function with $K(0) = 1$ and $K(1) = 0$. For $\mathbf{w} \in \Theta$, define an estimator of $\psi(\mathbf{w})$ by

$$\widehat{\psi}_n(\mathbf{w}) = \frac{c(h,K)}{k} \sum_{i=1}^n K\left(\frac{1 - \mathbf{w}^T\mathbf{W}_i}{h}\right) 1_{[R_i > R_{n-k:n}]}$$

with $0 < h < 1$ and

$$c(h,K) = \left(\int_{C_\mathbf{w}(h)} K\left(\frac{1 - \mathbf{v}^T\mathbf{w}}{h}\right) d\lambda(\mathbf{v})\right)^{-1}, \quad C_\mathbf{w}(h) = \{\mathbf{v} \in \Theta : \mathbf{w}^T\mathbf{v} \geq 1 - h\};$$

cf. Hall, Watson and Cabrera (1987).

For estimating Q_n it suffices to estimate $\widetilde{Q}_n$, see (7). Hence, in view of (8), we define

$$(9) \qquad \widehat{Q}_n = \widehat{U}\left(\frac{n}{k}\right)\left(\frac{k\widehat{\nu(S)}}{np}\right)^{1/\widehat{\alpha}} \widehat{S}$$

with

$$\widehat{S} = \{\mathbf{z} = r\mathbf{w} : r \geq (\widehat{\alpha}\widehat{\psi}_n(\mathbf{w}))^{1/(\widehat{\alpha}+d)}, \mathbf{w} \in \Theta\}$$

and

$$\widehat{\nu(S)} = \widehat{\alpha}^{-\widehat{\alpha}/(\widehat{\alpha}+d)} \int_\Theta (\widehat{\psi}_n(\mathbf{w}))^{d/(\widehat{\alpha}+d)} \, d\lambda(\mathbf{w}).$$

In the definition of the set S, the choice of the value 1 was not motivated. We could have taken any number $c > 0$ instead. Such an alternative definition of

S would lead to exactly the same estimator $\widehat{Q}_n$, which shows that the value 1 plays no role.

Assume

(c) $$\lim_{t \to \infty} \frac{U(t)}{t^{1/\alpha}} = c \qquad \text{for some } c \in (0, \infty).$$

Note that this simple condition is weaker than the usual second order condition with negative second order parameter ρ [see, e.g., Theorem 4.3.8 in de Haan and Ferreira (2006)]; indeed, there exist functions U with $\rho = 0$ that satisfy condition (c).

THEOREM 1.   *Let $p \to 0$ as $n \to \infty$. Assume conditions* (a), (b), (c) *hold and that $\widehat{\alpha}$ is such that $\sqrt{k}(\widehat{\alpha} - \alpha) = O_{\mathbb{P}}(1)$. Also assume that $(\log np)/\sqrt{k} \to 0$, $h \to 0$ and $k/(c(h, K) \log k) \to \infty$, as $n \to \infty$. Then we have*

(10) $$\frac{P(\widehat{Q}_n \triangle Q_n)}{p} \xrightarrow{\mathbb{P}} 0 \qquad as \ n \to \infty,$$

and hence

$$\frac{P(\widehat{Q}_n)}{p} \xrightarrow{\mathbb{P}} 1.$$

REMARK 1.   The tuning parameter k is used in the estimators of $\alpha, U(n/k)$ and ψ. It is important to be able to choose three different values for k, denoted with k_α, k_U and k_ψ, respectively. (Note that "good" values of k_α and k_U are determined by the tail of H—the distribution function of R_1—whereas a good k_ψ is determined by the conditional distribution of $\mathbf{W}_1$, given that $R_1 > r$, for large r.) If we adapt the conditions of the theorem, in particular if (b) holds for k_α, k_U, k_ψ and if $(\log np)/\sqrt{k_\alpha} \to 0$, $k_\psi/(c(h, K) \log k_\psi) \to \infty$ and $(\log k_U)/\sqrt{k_\alpha} \to 0$, then (10) remains true for the generalized estimator that allows for the aforementioned different k-values. We will use this generalized estimator in the simulation study and the real data application.

The actual choice of these k-values is a notorious problem in extreme value theory. A solution of this problem is far beyond the scope of the present paper. We will only give heuristic guidelines here. First, consider the estimation of α. Plot $\widehat{\alpha}$ as a function of k. Now find the first stable, that is, approximately constant, region in the graph of this function. This vertical level is the final estimate of α. It is also possible to use (complicated) *asymptotically* optimal procedures; see, for example, Danielsson et al. (2001). Once the estimate $\widehat{\alpha}$ is fixed, we plot $\widehat{U}(\frac{n}{k})(\frac{k}{n})^{1/\widehat{\alpha}}$ against k and we search for the first stable part in this graph. The vertical level is now the estimate of the constant c in condition (c). Observe that $\widehat{U}(\frac{n}{k})(\frac{k}{n})^{1/\widehat{\alpha}}$ is a building

block of $\widehat{Q}_n$, so we do not need to estimate $U(\frac{n}{k})$ separately. Also observe that we do not (need to) determine k_α and k_U, but only a region of good values. Finally, using again the already fixed $\widehat{\alpha}$, we plot $\widehat{\nu(S)}$ as a function of k and again we search for the first stable region; we take k_ψ to be the midpoint of this region of k-values.

REMARK 2. The class of multivariate regularly varying distributions is quite large. It contains, for example, all elliptical distributions with a heavy tailed radial distribution and all distributions in the domain of a sum-attraction of a multivariate (nonnormal) stable distribution. It seems natural, however, to try to extend the assumption of multivariate regular variation to the case of nonequal tail indices α. It is an important feature of the present model that all directions are equally important: the marginal distributions do not play a special role. An extension to nonequal tail indices would be possible in principle, but it will be of limited value since it only works if *marginal* transformations lead to the present model. Also note that basically all linear combinations of the components inherit the lowest of the marginal tail indices: the tail index is not a smooth function of the direction (if it is not constant). Moreover, the statistical theory that will be needed will be challenging and will lead to a new and different project.

REMARK 3. Note that the estimated extreme risk region $\widehat{Q}_n = \widehat{Q}_n(p)$ depends on p in a continuous way and has the property that $p_1 < p_2$ implies $\widehat{Q}_n(p_1) \subset \widehat{Q}_n(p_2)$. Hence, we can find the smallest p such that an observation is on the boundary of $\widehat{Q}_n(p)$. The corresponding observation can be considered the largest one and we know its "p-value." This is helpful in deciding whether some observation is the most extreme or if it is an outlier. Also, by continuing this procedure we can rank the larger observations.

3. Simulation study. In this section, a detailed simulation study is performed in order to investigate the finite sample performance of our estimator [with $1/\alpha$ estimated using the moment estimator of Dekkers, Einmahl and de Haan (1989) and with $K(u) = 1 - u$]. We consider five multivariate distributions.

- The bivariate Cauchy distribution with density

$$(11) \qquad f(x,y) = \frac{1}{2\pi(1 + x^2 + y^2)^{3/2}}, \qquad (x,y) \in \mathbb{R}^2.$$

  This is a very heavy tailed density, with $\alpha = 1$ and $\psi(\mathbf{w}) = 1/(2\pi)$, for $\mathbf{w} \in \Theta$.
- The trivariate Cauchy distribution with density

$$(12) \qquad f(x,y,z) = \frac{1}{\pi^2(1 + x^2 + y^2 + z^2)^2}, \qquad (x,y,z) \in \mathbb{R}^3.$$

This is also a very heavy tailed density, with $\alpha = 1$ and $\psi(\mathbf{w}) = 1/(4\pi)$, for $\mathbf{w} \in \Theta$.

- A bivariate elliptical distribution with density ($r_0 \approx 1.2481$)

$$(13) \qquad f(x,y) = \begin{cases} \dfrac{3}{4\pi} r_0^4 (1 + r_0^6)^{-3/2}, & x^2/4 + y^2 < r_0^2, \\[2mm] \dfrac{3(x^2/4 + y^2)^2}{4\pi(1 + (x^2/4 + y^2)^3)^{3/2}}, & x^2/4 + y^2 \geq r_0^2. \end{cases}$$

It is less heavy tailed. We have $\alpha = 3$ and $\psi(w_1, w_2) = c(1 + 3w_2^2)^{-5/2}, \mathbf{w} = (w_1, w_2) \in \Theta$, with $c \approx 0.6028$.

- A bivariate "clover" distribution with density ($r_0 \approx 1.2481$)

$$(14) \qquad f(x,y) = \begin{cases} \dfrac{3}{10\pi} r_0^4 (1 + r_0^6)^{-3/2} \left(5 + \dfrac{4(x^2+y^2)^2 - 32x^2y^2}{r_0(x^2+y^2)^{3/2}}\right), \\[2mm] \qquad x^2 + y^2 < r_0^2, \\[2mm] \dfrac{3(9(x^2+y^2)^2 - 32x^2y^2)}{10\pi(1 + (x^2+y^2)^3)^{3/2}}, \\[2mm] \qquad x^2 + y^2 \geq r_0^2. \end{cases}$$

We have $\alpha = 3$, again, and $\psi(w_1, w_2) = (9 - 32w_1^2 w_2^2)/(10\pi), \mathbf{w} = (w_1, w_2) \in \Theta$.

- A bivariate asymmetric shifted distribution with density $[r_0 \approx 1.2331, \widetilde{r}(x,y) := r_0 \vee ((x+5)^2 + y^2)^{1/2}]$

$$(15) \ f(x,y) = \begin{cases} \dfrac{\widetilde{r}^2(x,y)}{6\pi(1 + \widetilde{r}^4(x,y))^{5/4}} \left(3 + \dfrac{x+5}{\widetilde{r}(x,y)}\right), & y \geq 0, \\[3mm] \dfrac{\widetilde{r}^2(x,y)}{6\pi(1 + \widetilde{r}^4(x,y))^{5/4}} \left(3 + \dfrac{(x+5)^3 - 3(x+5)y^2}{\widetilde{r}^3(x,y)}\right), & y < 0. \end{cases}$$

This distribution is not symmetric and the "center" is not the origin, but $(-5, 0)$; $\alpha = 1$ and $\psi(w_1, w_2) = \frac{1}{6\pi}(3 + w_1)$, if $w_2 \geq 0$, and $\psi(w_1, w_2) = \frac{1}{6\pi}(3 + 4w_1^3 - 3w_1)$, if $w_2 < 0$, $\mathbf{w} = (w_1, w_2) \in \Theta$.

First, we simulated single data sets of size 5,000 of the bivariate Cauchy distribution, the elliptical distribution in (13), the clover distribution in (14) and the asymmetric shifted distribution in (15). We computed the true and estimated risk regions for $p = 1/2{,}000$, $1/5{,}000$ or $1/10{,}000$. This is depicted in Figure 1. We see that the estimated regions are relatively close to the true risk regions. It is interesting to note that the p-value (see Remark 3) of the largest observation for the Cauchy sample is 0.000209, which is about $1/n$. This shows that this observation is a typical one. (Looking at the data only, one might want to conclude that this observation is an outlier.) Also note that for the bivariate Cauchy distribution, for, for example, $p = 1/10{,}000$,

Cauchy Density, n=5000, p=1/2000, 1/5000, 1/10000

Elliptical Density, n=5000, p=1/2000, 1/10000

Clover Density, n=5000, p=1/2000, 1/10000

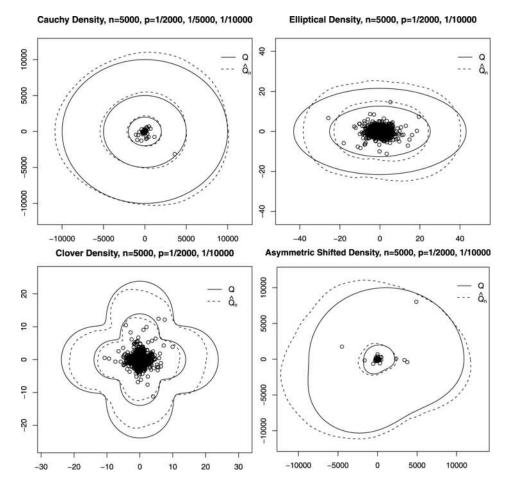**Asymmetric Shifted Density, n=5000, p=1/2000, 1/10000**

FIG. 1. *True and estimated risk regions based on one sample of size 5,000 from the bivariate Cauchy distribution, the elliptical distribution in (13), the clover distribution in (14) and the asymmetric shifted distribution in (15).*

the density f at the boundary of the true risk region is less than 10^{-12}. This emphasizes that we are estimating in an "almost empty" part of the plane and that a fully nonparametric procedure could not work here.

In addition, we simulated one sample of the bivariate distribution with independent t_3-components. This distribution does *not* satisfy condition (a), since the spectral measure is discrete and concentrated on the intersection of the coordinate axes with the unit circle. We also simulated one sample of a bivariate "logarithmic" distribution with $\alpha = 1$ and uniform spectral measure, but where the radial distribution satisfies $U(t)/(t \log t)$ tends to a constant and hence $U(t)/t \to \infty$ as $t \to \infty$, that is, this distribution does *not* satisfy condition (c). Although both distributions do not satisfy our condi-
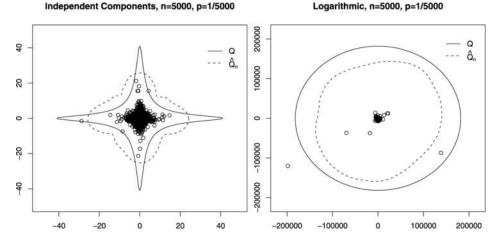
Fig. 2.   *True and estimated risk regions based on one sample of size* 5,000 *from the bivariate distribution with independent* t_3-*components and the "logarithmic" distribution.*

tions, we see nevertheless satisfactory behavior of the estimator in Figure 2. In the left panel, the estimated region has about the right size and the difficult shape is approximated reasonably well; in the right panel, we see that both the shape and the size are approximated quite well.

After this visual assessment of our estimator based on one sample at a time, we now investigate its performance based on 100 simulated samples of size 5,000. We will compare our estimator (denoted EVT) to a nonparametric and to a more parametric estimator. The nonparametric estimator is only defined in case $p = 1/n$ and tries to mimic the largest order statistic as an estimator of the $(1 - 1/n)$th quantile in the univariate case. It aims at elliptical level sets. It is defined as follows. First, calculate the smallest ellipsoid containing half of the data, the so-called MVE. Then inflate this ellipsoid, such that the "largest" observation lies on its boundary. Now the region outside this ellipsoid is the estimator.

For $d = 2$, the more parametric estimator is defined similarly to $\widehat{Q}_n$ in (9), but (only) the estimation of $(\nu(S))^{1/\alpha}S$ is done parametrically. Therefore, this estimator has the same size as $\widehat{Q}_n$, but a different shape. (Note that the fully parametric estimator based on multivariate normality would have a very bad performance.) Take the k observations with radius $R_i > R_{n-k:n}$ and consider the transformed data $(R_i/R_{n-k:n}, \mathbf{W}_i)$. In line with the limit result in (1), assume that these data have a "distribution" $(\cdot)^{-\alpha}\Psi$, where Ψ depends on a parameter ρ. To be precise, we assume for the density

$$\psi_\rho(\theta) = (4\pi)^{-1}(2 + \sin(2(\theta - \rho))), \qquad 0 \le \theta < 2\pi, 0 \le \rho < \pi.$$

TABLE 1
Median of the relative errors $P(\widehat{Q}_n \triangle Q_n)/p$ of the three
estimators, for $p = 1/5,000$ (p1) and $1/10,000$ (p2)

Density	EVT p1	Par p1	NP p1	EVT p2	Par p2
Biv. Cauchy	0.28	0.29	0.72	0.31	0.32
Triv. Cauchy	0.22	–	0.54	0.24	–
Elliptical	0.36	0.51	0.80	0.39	0.54
Clover	0.44	0.57	0.58	0.49	0.61
Asymm. shifted	0.26	0.27	0.61	0.30	0.32

(Here a point on the unit circle is represented by its angle $\theta \in [0, 2\pi)$.) Now α and ρ are estimated by maximum likelihood; observe that this yields the Hill estimator for $1/\alpha$.

Table 1 shows for the three different estimators the median of the 100 relative errors $P(\widehat{Q}_n \triangle Q_n)/p$ for $p = 1/5,000$ (p1) and $1/10,000$ (p2). In Figure 3, boxplots are shown of the relative error $P(\widehat{Q}_n \triangle Q_n)/p$ for $p = 1/5,000$ (p1) and $1/10,000$ (p2). From this table and figure, we see a good performance of our estimator. Its behavior does not change much if p changes from $1/5,000$ to $1/10,000$. The parametric estimator performs reasonably well, but it is outperformed by our estimator, in particular for the elliptical and clover densities. Recall that this estimator can be seen as a modification of our estimator, since it uses the same estimated inflation factor, but the shape is estimated differently. We see a moderate performance of the nonparametric estimator; also, it cannot be adapted to $p = 1/10,000$. Given that the estimation of these extreme risk regions is a statistically difficult problem, we see decent behavior of the three estimation methods. Obviously the parametric and the nonparametric estimator do not perform well if the parametric part of the model is not adequate or if the shape of the region is not elliptical, respectively. The EVT estimator, presented in this paper, does not suffer from these shortcomings and performs well for many multivariate distributions.

4. Application. In this section, an application of our method to foreign exchange rate data is presented. The data are the daily exchange rates of Yen-Dollar and Pound-Dollar from January 4, 1999 to July 31, 2009. Consider the daily log returns given by $X_{i,j} = \log(Y_{i+1,j}/Y_{i,j})$, with $i = 1, \ldots, 2,664$, $j = 1, 2$, and $Y_{i,1}$ is the daily exchange rate of the Yen to the Dollar and $Y_{i,2}$ of the Pound to the Dollar. First, we check the equality of the extreme value indices (the reciprocals of the tail indices) of the right and left tails of both marginal distributions and that of the radius. This yields 5 extreme value indices; the 5 estimates in increasing order are: 0.141, 0.191, 0.223, 0.242, 0.256. Hence, the maximal difference is 0.115. Based on the asymptotic normality of the moment estimator of the extreme value index,
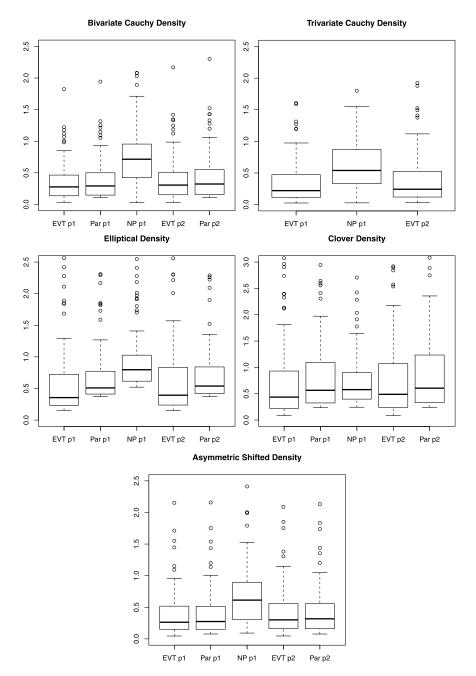
FIG. 3.   *Boxplots of $P(\widehat{Q}_n \triangle Q_n)/p$ for the here proposed estimator and for the parametric and the nonparametric estimator, based on 100 simulated data sets of size 5,000 from the five presented densities for $p = 1/5,000$ (p1) and $1/10,000$ (p2).*
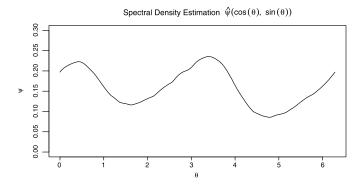
Spectral Density Estimation $\hat{\psi}(\cos(\theta), \sin(\theta))$

FIG. 4. *Estimator of ψ of bivariate exchange rate returns.*

we compute an approximate upper bound for the maximal difference of the 5 estimators under the null hypothesis of equality: 0.264. Hence, there is no evidence that the 5 extreme value indices are different. Other exchange rate data sets share this property. There are also economic arguments supporting this claim. Therefore, we estimate α based on the radius and find $\widehat{\alpha} = 3.90$. As a next step, we estimate the density ψ of the spectral measure. The estimate $\widehat{\psi}_n$ is depicted in Figure 4; it is almost periodic with period π. This yields that the boundary of the estimated extreme risk region is not like a circle, but more like an ellipse. The location of the maxima of $\widehat{\psi}_n$ correspond to the major axis of the region. We estimate the extreme risk regions for $p = 1/2{,}000, 1/5{,}000$ and $1/10{,}000$; see Figure 5. For risk management

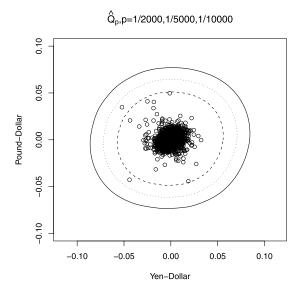$\widehat{Q}_{p}$,p=1/2000,1/5000,1/10000

FIG. 5. *Estimated extreme risk regions of exchange rate returns.*

of financial institutions in the U.S., it is important to know which extreme exchange rate returns w.r.t. the Pound and the Yen can occur and which returns essentially never occur. Our estimate answers this question. More specifically, points on the boundary of the estimated extreme risk region can be used as multivariate stress test scenarios. A scenario on the intersection of the major axis of the ellipse-like boundary of the extreme risk region and the boundary itself corresponds to a larger shock than a scenario on the intersection of the minor axis of the ellipse-like boundary and the boundary itself, but our method shows that their "extremeness" is about the same.

5. Proofs. For the proof of the theorem, we need several lemmas and propositions. We assume throughout that the conditions of the theorem are in force. We start with a lemma on regular variation in $\mathbb{R}^d$.

LEMMA 1. *Write* $l = 1/\int_{\{\|\mathbf{z}\| \geq 1\}} q(\mathbf{z}) \, d\mathbf{z}$. *For any* $\varepsilon > 0$,

$$(16) \qquad \lim_{t \to \infty} \sup_{\|\mathbf{z}\| \geq \varepsilon} \left| \frac{f(t\mathbf{z})}{t^{-d}V(t)} - q(\mathbf{z}) \right| = 0.$$

Moreover

$$(17) \qquad \lim_{t \to \infty} \frac{\mathbb{P}(\mathbf{X} \in tB)}{V(t)} = \int_B q(\mathbf{z}) \, d\mathbf{z}$$

for any Borel set B *bounded away from the origin. Define* $q_t(\mathbf{z}) = t(U(t))^d \times f(U(t)\mathbf{z})$. *Then*

$$(18) \qquad \lim_{t \to \infty} \sup_{\|\mathbf{z}\| \geq \varepsilon} |q_t(\mathbf{z}) - lq(\mathbf{z})| = 0.$$

Let $\widetilde{h}$ *be the density of* H, *then*

$$(19) \qquad \lim_{t \to \infty} \frac{\widetilde{h}(t)}{t^{-1}V(t)} = \frac{\alpha}{l}.$$

PROOF. For any $\|\mathbf{z}\| \geq \varepsilon > 0$ [cf. Theorem 2.1 in de Haan and Resnick (1987)],

$$\left| \frac{f(t\mathbf{z})}{t^{-d}V(t)} - q(\mathbf{z}) \right|$$

$$= \left| \frac{f(t\|\mathbf{z}\|(\mathbf{z}/\|\mathbf{z}\|))}{(t\|z\|)^{-d}V(t\|z\|)} \cdot \frac{(t\|z\|)^{-d}V(t\|z\|)}{t^{-d}V(t)} - q(\mathbf{z}) \right|$$

$$\leq \|\mathbf{z}\|^{-d-\alpha} \left| \frac{f(t\|\mathbf{z}\|(\mathbf{z}/\|\mathbf{z}\|))}{(t\|z\|)^{-d}V(t\|z\|)} - q\left(\frac{\mathbf{z}}{\|\mathbf{z}\|}\right) \right|$$

$$+ \frac{f(t\|\mathbf{z}\|(\mathbf{z}/\|\mathbf{z}\|))}{(t\|z\|)^{-d}V(t\|z\|)} \left| \frac{(t\|z\|)^{-d}V(t\|z\|)}{t^{-d}V(t)} - \|\mathbf{z}\|^{-d-\alpha} \right|$$

Then (16) follows from condition (a).

Let a Borel set B be such that $B \subset \{\|\mathbf{z}\| \geq \gamma\}$, for some $\gamma > 0$. Then for $\mathbf{z} \in B$ and sufficiently large t, $f(t\mathbf{z})/t^{-d}V(t)$ is bounded by $q(\|\mathbf{z}\|^{-1}\mathbf{z})\|\mathbf{z}\|^{-a/2-d}$. Hence, (17) holds by Lebesgue's dominated convergence theorem.

We have from (17), as $t \to \infty$,

$$tV(U(t)) = \frac{V(U(t))}{\mathbb{P}(R \geq U(t))} \to l.$$

Hence (16) implies, uniformly for $\|\mathbf{z}\| \geq \varepsilon$,

$$q_t(\mathbf{z}) = tV(U(t)) \frac{f(U(t)\mathbf{z})}{(U(t))^{-d}V(U(t))} \to lq(\mathbf{z}).$$

Note that

$$1 - H(t) = \mathbb{P}(R > t) = \int_t^\infty \int_\Theta f(r\mathbf{w}) \, d\lambda(\mathbf{w}) r^{d-1} \, dr.$$

By taking derivatives, (16) and the homogeneity of q, we obtain

$$\lim_{t \to \infty} \frac{\widetilde{h}(t)}{t^{-1}V(t)} = \int_\Theta q(\mathbf{w}) \, d\lambda(\mathbf{w}) = \alpha \int_{\{\|\mathbf{z}\| \geq 1\}} q(\mathbf{z}) \, d\mathbf{z} = \alpha/l. \qquad \square$$

We now see that (5) and (6) hold with $V = 1 - H$. From now on, we will make the choice $V = 1 - H$ and hence $l = 1$. Note that with this choice the relations (3) [with $\nu(B) = \int_B q(\mathbf{z}) \, d\mathbf{z}$] and (4) readily follow from (17).

COROLLARY 1. *For all Borel sets B with positive distance from the origin,*

(20) $$\lim_{t \to \infty} tP(U(t)B) = \nu(B)$$

and

(21) $$\lim_{n \to \infty} \frac{\nu(S)}{p} P\left(U\left(\frac{n}{k}\right)\left(\frac{k\nu(S)}{np}\right)^{1/\alpha} B\right) = \nu(B).$$

PROOF. From $\mathbb{P}(R \geq U(t)) = 1/t$ and (3), we obtain (20). It follows from (c) that

(22) $$\frac{U(\nu(S)/p)}{U(n/k)(k\nu(S)/(np))^{1/\alpha}} \to 1.$$

This yields (21). $\square$

LEMMA 2. *For each $\varepsilon > 0$, there exists a $\delta > 0$ and $t_0 > 0$ such that for $t > t_0$*

$$\left\{\mathbf{z} : \frac{f(t\mathbf{z})}{t^{-d}V(t)} \leq \varepsilon\right\} \subset \{\mathbf{z} : \|\mathbf{z}\| > \delta\}.$$

PROOF. It is sufficient to prove $\{\mathbf{z} : \|\mathbf{z}\| \leq \delta\} \subset \{\mathbf{z} : f(t\mathbf{z})/(t^{-d}V(t)) > \varepsilon\}$. First, by (6) and the continuity of q, for some $c_1 > 0$, there exists $s_0 > 0$ such that for $s > s_0$

$$\inf_{\mathbf{w} \in \Theta} \frac{f(s\mathbf{w})}{s^{-d}V(s)} \geq c_1$$

and also for $s_1, s_2 > s_0$ [cf. Proposition B.1.9.5 in de Haan and Ferreira (2006)]

$$\frac{V(s_1)}{V(s_2)} > \frac{1}{2} \left(\frac{s_1}{s_2} \right)^{-\alpha/2}.$$

Now for $t > s_0$ and any $\mathbf{z} \in \{\mathbf{z} : \|\mathbf{z}\| \leq \delta\}$, there are two possibilities.

(i) $t\|\mathbf{z}\| > s_0$, then

$$\frac{f(t\mathbf{z})}{t^{-d}V(t)} = \frac{f(t\|\mathbf{z}\|(\mathbf{z}/\|\mathbf{z}\|))}{(t\|\mathbf{z}\|)^{-d}V(t\|\mathbf{z}\|)} \cdot \frac{(t\|\mathbf{z}\|)^{-d}V(t\|\mathbf{z}\|)}{t^{-d}V(t)} > \frac{1}{2}c_1\delta^{-\alpha/2-d} > \varepsilon;$$

(ii) $t\|\mathbf{z}\| \leq s_0$, then by continuity of f and $f > 0$, we have for some $c_2 > 0$, $f(t\mathbf{z}) \geq c_2$, and hence, since $\lim_{t\to\infty} t^{-d}V(t) = 0$, we obtain for $t > t_0 (\geq s_0)$

$$\frac{f(t\mathbf{z})}{t^{-d}V(t)} > \varepsilon. \qquad \square$$

LEMMA 3. *For $\varepsilon > 0$ and large n,*

$$\bar{Q}_n \subset U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z} : q(\mathbf{z}) \leq 1 + \varepsilon\}$$

and

$$\bar{Q}_n \supset U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z} : q(\mathbf{z}) \leq 1 - \varepsilon\}.$$

PROOF. Recall that $\bar{Q}_n = \{\mathbf{z} : f(\mathbf{z}) \leq (\frac{np}{k\nu(S)})^{1+(d/\alpha)} \frac{1}{(n/k)(U(n/k))^d}\}$. It follows from (22) that for n large enough and $\varepsilon_1 > 0$

$$\bar{Q}_n = U\left(\frac{\nu(S)}{p}\right)\left\{\mathbf{z} : f\left(U\left(\frac{\nu(S)}{p}\right)\mathbf{z}\right) \leq \left(\frac{np}{k\nu(S)}\right)^{1+(d/\alpha)} \frac{1}{(n/k)(U(n/k))^d}\right\}$$

$$= U\left(\frac{\nu(S)}{p}\right)\left\{\mathbf{z} : q_{\nu(S)/p}(\mathbf{z}) \leq \left(\frac{np}{k\nu(S)}\right)^{d/\alpha} \left(U\left(\frac{n}{k}\right)\right)^{-d} \left(U\left(\frac{\nu(S)}{p}\right)\right)^d\right\}$$

$$\subset U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z} : q_{\nu(S)/p}(\mathbf{z}) \leq 1 + \varepsilon_1\}.$$

Now Lemma 2 implies $\{\mathbf{z}: q_{\nu(S)/p}(\mathbf{z}) \leq 1 + \varepsilon_1\} \subset \{\mathbf{z}: \|\mathbf{z}\| > \delta\}$, hence we have by (18)

$$\bar{Q}_n \subset U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z}: q(\mathbf{z}) \leq 1 + \varepsilon\}.$$

The other inclusion follows in the same way (but Lemma 2 is not needed).
□

LEMMA 4. *For $\varepsilon > 0$ and large n,*

$$\widetilde{Q}_n \subset U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z}: q(\mathbf{z}) \leq 1 + \varepsilon\}$$

and

$$\widetilde{Q}_n \supset U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z}: q(\mathbf{z}) \leq 1 - \varepsilon\}.$$

PROOF. Recall that $\widetilde{Q}_n = U(\frac{n}{k})(\frac{k\nu(S)}{np})^{1/\alpha}\{\mathbf{z}: q(z) \leq 1\}$.
Put $T_n = (U(\frac{\nu(S)}{p}))^{-1} U(\frac{n}{k})(\frac{k\nu(S)}{np})^{1/\alpha}$, then

$$\widetilde{Q}_n = U\left(\frac{\nu(S)}{p}\right)\{T_n\mathbf{z}: q(\mathbf{z}) \leq 1\} = U\left(\frac{\nu(S)}{p}\right)\{T_n\mathbf{z}: q(T_n\mathbf{z}) \leq T_n^{-d-\alpha}\}$$

$$= U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z}: q(\mathbf{z}) \leq T_n^{-d-\alpha}\}.$$

Since $T_n \to 1$ as $n \to \infty$ by (22), the result follows. □

PROPOSITION 1. *We have*

$$\lim_{n\to\infty} \frac{P(Q_n \triangle \widetilde{Q}_n)}{p} = 0.$$

PROOF. Note that $P(Q_n \triangle \widetilde{Q}_n) \leq P(Q_n \triangle \bar{Q}_n) + P(\bar{Q}_n \triangle \widetilde{Q}_n)$. Observe that $Q_n \subset \bar{Q}_n$ or $\bar{Q}_n \subset Q_n$, hence $P(Q_n \triangle \bar{Q}_n) \leq |p - P(\bar{Q}_n)|$. By Lemma 3 and Corollary 1, for any $\varepsilon > 0$ and large n

$$\frac{\nu(S)}{p}P(\bar{Q}_n) \leq \frac{\nu(S)}{p}P\left(U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z}: q(\mathbf{z}) \leq 1 + \varepsilon\}\right)$$

$$\to \nu(\{\mathbf{z}: q(\mathbf{z}) \leq 1 + \varepsilon\})$$

$$= \nu(\{\mathbf{z}: q(\mathbf{z}(1 + \varepsilon)^{1/(d+\alpha)}) \leq 1\})$$

$$= \nu(\{(1 + \varepsilon)^{-1/(d+\alpha)}\mathbf{z}: q(\mathbf{z}) \leq 1\})$$

$$= (1 + \varepsilon)^{\alpha/(d+\alpha)}\nu(S).$$

Thus, $\limsup_{n\to\infty} \frac{P(\bar{Q}_n)}{p} \leq (1 + \varepsilon)^{\alpha/(2+\alpha)}$.

Similarly, we have $\liminf_{n\to\infty} \frac{P(\bar{Q}_n)}{p} \geq (1 - \varepsilon)^{\alpha/(2+\alpha)}$. Hence, $\lim_{n\to\infty} \frac{P(\bar{Q}_n)}{p} = 1$, that is, $\lim_{n\to\infty} \frac{P(Q_n \triangle \bar{Q}_n)}{p} = 0$.

In the same way, it follows from Lemmas 3 and 4 that

$$\frac{\nu(S)}{p} P(\bar{Q}_n \triangle \widetilde{Q}_n) \leq \frac{\nu(S)}{p} P\left(U\left(\frac{\nu(S)}{p}\right)\{\mathbf{z}: 1 - \varepsilon \leq q(\mathbf{z}) \leq 1 + \varepsilon\}\right)$$

$$\to \nu(\{\mathbf{z}: 1 - \varepsilon \leq q(\mathbf{z}) \leq 1 + \varepsilon\})$$

$$= \nu(S)((1+\varepsilon)^{\alpha/(d+\alpha)} - (1-\varepsilon)^{\alpha/(d+\alpha)}).$$

Hence, $\lim_{n\to\infty} \frac{P(\bar{Q}_n \triangle \widetilde{Q}_n)}{p} = 0$. $\square$

The following proposition shows uniform consistency of $\widehat{\psi}_n$ and might be of independent interest. There is an abundant literature on density estimation for directional data. In particular, uniform consistency of density estimators for directional data has been established in Bai, Rao and Zhao (1988). Here, however, the data do not have a *fixed* probability density on Θ: the density ψ is defined via a limit relation. Hence, ψ is only an approximate model for the directional data. As a consequence, a more general result is required.

PROPOSITION 2.   *As* $n \to \infty$,

$$\sup_{\mathbf{w}\in\Theta} |\widehat{\psi}_n(\mathbf{w}) - \psi(\mathbf{w})| \xrightarrow{\mathbb{P}} 0.$$

PROOF.   It is easy to see that, for any $\eta > 0$, there exists a function

$$K^* = \sum_{j=1}^{m} \alpha_j 1_{[r_{j-1}, r_j)}$$

with $1 \geq \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m \geq 0$ and $0 = r_0 < r_1 < \cdots < r_m = 1$, such that

$$\sup_{u\in[0,1]} |K(u) - K^*(u)| \leq \eta.$$

Write $U_i = 1 - H(R_i)$, $i = 1, \ldots, n$, and denote the corresponding order statistics with $U_{i:n}$. Let $\widetilde{P}$ be the probability measure on $\Theta \times (0,1)$ corresponding to $(\mathbf{W}_1, U_1)$ and let $\widetilde{P}_n$ be the empirical measure of the $(\mathbf{W}_i, U_i)$ $i = 1, \ldots, n$. Define

$$\psi_n^*(\mathbf{w}) = \frac{c(h,K)}{k} \sum_{i=1}^{n} K^*\left(\frac{1 - \mathbf{w}^T \mathbf{W}_i}{h}\right) 1_{[R_i > R_{n-k:n}]}$$

and

$$\psi_{n,j}^*(\mathbf{w}) = \frac{nc(h,K)}{k} \widetilde{P}_n(D_{\mathbf{w},j} \times (0, U_{k:n}])$$

with $D_{\mathbf{w},j} = \{\mathbf{v} \in \Theta : 1 - hr_j < \mathbf{w}^T\mathbf{v} \le 1 - hr_{j-1}\}$. Observe that $\psi_n^*(\mathbf{w}) = \sum_{j=1}^{m} \alpha_j \psi_{n,j}^*(\mathbf{w})$. Also write

$$\psi_{n,j}(\mathbf{w}) = \frac{nc(h,K)}{k}\widetilde{P}(D_{\mathbf{w},j} \times (0, U_{k:n}]).$$

Let $\varepsilon > 0$. It is sufficient to show that for large n

(23)
$$\mathbb{P}\left(\sup_{\mathbf{w}\in\Theta}\left| \widehat{\psi}_n(\mathbf{w}) - \sum_{j=1}^{m} \alpha_j\psi_{n,j}(\mathbf{w}) \right| \ge 2\varepsilon \right) \le 2\varepsilon,$$

(24)
$$\mathbb{P}\left(\sup_{\mathbf{w}\in\Theta}\left| \sum_{j=1}^{m} \alpha_j(\psi_{n,j}(\mathbf{w}) - c(h,K)\Psi(D_{\mathbf{w},j})) \right| \ge 2\varepsilon \right) \le \varepsilon,$$

(25)
$$\sup_{\mathbf{w}\in\Theta}\left| c(h,K)\sum_{j=1}^{m} \alpha_j\Psi(D_{\mathbf{w},j}) - \psi(\mathbf{w}) \right| \le \varepsilon.$$

For $\mathbf{w} \in \Theta$ and $\delta \in (0,1)$, write $\mathcal{C}_\delta = \{C_{\mathbf{w}}(a) : \mathbf{w} \in \Theta, a \le \delta\}$. Note that, as $n \to \infty$,

(26)
$$\sup_{C\in\mathcal{C}_1, 0<s\le 2} \frac{1}{\lambda(C)}\left| \frac{n}{k}\widetilde{P}(C \times (0, sk/n]) - s\Psi(C) \right| \to 0.$$

This readily follows from

$$\frac{n}{k}\widetilde{P}(C \times (0, sk/n]) = \frac{n}{k}\mathbb{P}\left(\mathbf{W} \in C, R \ge U\left(\frac{n}{sk}\right) \right)$$

$$= \frac{n}{k}\int_{U(n/(sk))}^{\infty} \int_C \frac{f(r\mathbf{w})}{r^{-d}V(r)}\, d\lambda(\mathbf{w})\, r^{-1}V(r)\, dr$$

and (16) and (19).

Now we prove (23). It is easy to show that

$$c(h,K) = \left(\frac{2\pi^{(d-1)/2}}{\Gamma((d-1)/2)}\int_{1-h}^{1} K\left(\frac{1-t}{h}\right)(1-t^2)^{(d-3)/2}\, dt \right)^{-1}$$

and hence

(27)
$$\limsup_{h\downarrow 0} c(h,K)\lambda(C_{\mathbf{w}}(h)) < \infty.$$

We have

$$|\widehat{\psi}_n(\mathbf{w}) - \psi_n^*(\mathbf{w})|$$

$$= \frac{c(h,K)}{k}\left| \sum_{i=1}^{n}\left(K\left(\frac{1-\mathbf{w}^T\mathbf{W}_i}{h}\right) - K^*\left(\frac{1-\mathbf{w}^T\mathbf{W}_i}{h}\right) \right)\mathbf{1}_{[R_i>R_{n-k:n}]} \right|$$

$$(28) \quad \leq \frac{c(h,K)}{k} \sum_{i=1}^{n} \eta 1_{[\mathbf{W}_i \in C_{\mathbf{w}}(h), R_i > R_{n-k:n}]}$$

$$\leq \eta \frac{nc(h,K)}{k} \widetilde{P}(C_{\mathbf{w}}(h) \times (0, U_{k:n}])$$

$$+ \eta \frac{nc(h,K)}{k} |(\widetilde{P}_n - \widetilde{P})(C_{\mathbf{w}}(h) \times (0, U_{k:n}])|.$$

By (26), for η small enough the first term is less than ε, with probability tending to one, uniformly in $\mathbf{w} \in \Theta$. Also,

$$\left| \psi_n^*(\mathbf{w}) - \sum_{j=1}^{m} \alpha_j \psi_{n,j}(\mathbf{w}) \right|$$

$$(29) \qquad \leq \sum_{j=1}^{m} \alpha_j |\psi_{n,j}^*(\mathbf{w}) - \psi_{n,j}(\mathbf{w})|$$

$$\leq \sum_{j=1}^{m} \alpha_j \frac{nc(h,K)}{k} |(\widetilde{P}_n - \widetilde{P})(D_{\mathbf{w},j} \times (0, U_{k:n}])|.$$

From (29), (28) and (27), we see that for a proof of (23) it remains to show that

$$\frac{n}{k\lambda(C_{\mathbf{w}}(h))} \sup_{\mathbf{w} \in \Theta} \sup_{0 < a \leq 1} |(\widetilde{P}_n - \widetilde{P})(C_{\mathbf{w}}(ah) \times (0, U_{k:n}])| \overset{\mathbb{P}}{\to} 0.$$

It can be shown that there exists a constant $c = c(d)$ and finitely many $\mathbf{w}_l$, $l = 1, \ldots, l_h$ such that $l_h = O(c(h,K))$ as $h \downarrow 0$, and for every $\mathbf{w} \in \Theta$ and $0 < a \leq 1$

$$C_{\mathbf{w}}(ah) \in C_{\mathbf{w}_l}(ch) \qquad \text{for some } l.$$

Hence for $\varepsilon_1 > 0$,

$$\mathbb{P} \left(\frac{n}{k\lambda(C_{\mathbf{w}}(h))} \sup_{\mathbf{w} \in \Theta} \sup_{0 < a \leq 1} |(\widetilde{P}_n - \widetilde{P})(C_{\mathbf{w}}(ah) \times (0, U_{k:n}])| \geq \varepsilon_1 \right)$$

$$\leq \mathbb{P} \left(\max_{1 \leq l \leq l_h} \sup_{\substack{C \subset C_{\mathbf{w}_l}(ch) \\ C \in \mathcal{C}_h}} \sup_{0 < s \leq 2} |(\widetilde{P}_n - \widetilde{P})(C \times (0, sk/n])| \geq \varepsilon_1 k/n\lambda(C_{\mathbf{w}}(h)) \right)$$

$$+ \mathbb{P}(U_{k:n} > 2k/n)$$

$$\leq \sum_{l=1}^{l_h} \mathbb{P} \left(\sup_{\substack{C \subset C_{\mathbf{w}_l}(ch) \\ C \in \mathcal{C}_h}} \sup_{0 < s \leq 2} |(\widetilde{P}_n - \widetilde{P})(C \times (0, sk/n])| \geq \varepsilon_1 k/n\lambda(C_{\mathbf{w}}(h)) \right)$$

$$+ \mathbb{P}(U_{k:n} > 2k/n).$$

The latter probability tends to 0, so it suffices to consider the sum of the l_h probabilities. Write $b = \varepsilon_1 k \lambda(C_{\mathbf{w}}(h))$. Fix l and define $N = n\widetilde{P}_n(C_{\mathbf{w}_l}(ch) \times (0, 2k/n])$, $\mu = n\widetilde{P}(C_{\mathbf{w}_l}(ch) \times (0, 2k/n])$. Define the conditional probability measure $\widetilde{P}_c = \frac{n\widetilde{P}}{\mu}$ on $C_{\mathbf{w}_l}(ch) \times (0, 2k/n]$ and let $\widetilde{P}_{c,r}$ be the corresponding empirical measure, based on r observations. We have

$$\mathbb{P}\left(\sup_{\substack{C \subset C_{\mathbf{w}_l}(ch)\, 0<s\leq 2 \\ C \in \mathcal{C}_h}} \sup n|(\widetilde{P}_n - \widetilde{P})(C \times (0, sk/n])| \geq b\right)$$

$$\leq \sum_{r=\lceil \mu-b/3 \rceil}^{r=\lfloor \mu+b/3 \rfloor} \mathbb{P}\left(\sup_{\substack{C \subset C_{\mathbf{w}_l}(ch)\, 0<s\leq 2 \\ C \in \mathcal{C}_h}} \sup n|(\widetilde{P}_n - \widetilde{P})(C \times (0, sk/n])| \geq b|N=r\right)$$

$$\times \mathbb{P}(N=r) + \mathbb{P}(|N-\mu| \geq b/3)$$

$$\leq \sum_{r=\lceil \mu-b/3 \rceil}^{r=\lfloor \mu+b/3 \rfloor} \mathbb{P}\left(\sup_{\substack{C \subset C_{\mathbf{w}_l}(ch)\, 0<s\leq 2 \\ C \in \mathcal{C}_h}} \sup n\left|\left(\widetilde{P}_n - \frac{N}{\mu}\widetilde{P}\right)(C \times (0, sk/n])\right|\right.$$

$$\left. \geq \frac{b}{2}\bigg|N=r\right)\mathbb{P}(N=r)$$

$$+ \sum_{r=\lceil \mu-b/3 \rceil}^{r=\lfloor \mu+b/3 \rfloor} \mathbb{P}\left(\sup_{\substack{C \subset C_{\mathbf{w}_l}(ch)\, 0<s\leq 2 \\ C \in \mathcal{C}_h}} \sup n\left|\frac{(N-\mu)}{\mu}\widetilde{P}(C \times (0, sk/n])\right|\right.$$

$$\left. \geq \frac{b}{2}\bigg|N=r\right)\mathbb{P}(N=r)$$

$$(30) \quad + \mathbb{P}(|N-\mu| \geq b/3)$$

$$\leq \sum_{r=\lceil \mu-b/3 \rceil}^{r=\lfloor \mu+b/3 \rfloor} \mathbb{P}\left(\sup_{\substack{C \subset C_{\mathbf{w}_l}(ch)\, 0<s\leq 2 \\ C \in \mathcal{C}_h}} \sup r|(\widetilde{P}_{c,r} - \widetilde{P}_c)(C \times (0, sk/n])| \geq \frac{b}{2}\right)$$

$$\times \mathbb{P}(N=r)$$

$$+ \sum_{r=\lceil \mu-b/3 \rceil}^{r=\lfloor \mu+b/3 \rfloor} \mathbb{P}\left(|r-\mu| \geq \frac{b}{2}\right)\mathbb{P}(N=r) + \mathbb{P}(|N-\mu| \geq b/3).$$

Note that the first probability of the second sum in the right side of (30) is equal to 0. From Bennett's inequality [cf. Shorack and Wellner (1986), page 851], it follows that for some constant c_1

$$\mathbb{P}(|N-\mu| \geq b/3) \leq 2\exp\left(-\varepsilon_1^2 c_1 \frac{k}{c(h,K)}\right).$$

Hence, since $l_h = O(c(h, K))$,

$$\sum_{l=1}^{l_h} \mathbb{P}(|N - \mu| \geq b/3) = O\left(c(h, K) \exp\left(-\varepsilon_1^2 c_1 \frac{k}{c(h, K)}\right)\right) = o(1).$$

To complete the proof of (23), we need to consider the first sum in the right side of (30). For the first probability in there, we use Corollary 2.9 in Alexander (1984), a good probability bound for empirical processes on VC classes. We obtain as an upper bound

$$16 \exp\left(-\frac{b^2}{4r}\right).$$

Using $r \leq \mu + b/3$, we find for some constant c_2

$$\sum_{l=1}^{l_h} \sum_{r=\lceil \mu - b/3 \rceil}^{r=\lfloor \mu + b/3 \rfloor} \mathbb{P}\left(\sup_{\substack{C \subset C_{\mathbf{w}_l}(ch) \\ C \in \mathcal{C}_h}} \sup_{0 < s \leq 2} r|(\widetilde{P}_{c,r} - \widetilde{P}_c)(C \times (0, sk/n])| \geq \frac{b}{2}\right)$$

$$\times \mathbb{P}(N = r)$$

$$\leq 16 \sum_{l=1}^{l_h} \sum_{r=\lceil \mu - b/3 \rceil}^{r=\lfloor \mu + b/3 \rfloor} \exp\left(-\varepsilon_1^2 c_2 \frac{k}{c(h, K)}\right) \mathbb{P}(N = r)$$

$$\leq 16 \sum_{l=1}^{l_h} \exp\left(-\varepsilon_1^2 c_2 \frac{k}{c(h, K)}\right)$$

$$= o(1).$$

Next, we show (24). From (27) and (26), we obtain for $\varepsilon_2 > 0$ small enough,

$$\sup_{\mathbf{w} \in \Theta} \left| \sum_{j=1}^{m} \alpha_j (\psi_{n,j}(\mathbf{w}) - c(h, K) \Psi(D_{\mathbf{w},j})) \right|$$

$$= \sup_{\mathbf{w} \in \Theta} \left| \sum_{j=1}^{m} \alpha_j c(h, K)(n/k \widetilde{P}(D_{\mathbf{w},j} \times (0, U_{k:n}]) - \Psi(D_{\mathbf{w},j})) \right|$$

$$\leq \varepsilon_2 \sum_{j=1}^{m} \alpha_j c(h, K) \lambda(C_{\mathbf{w}}(h)) + \sup_{\mathbf{w} \in \Theta} \left| \sum_{j=1}^{m} \alpha_j c(h, K)(nU_{k:n}/k - 1) \Psi(D_{\mathbf{w},j}) \right|$$

$$\leq \varepsilon + \left| \frac{n}{k} U_{k:n} - 1 \right| \sum_{j=1}^{m} \alpha_j c(h, K) \lambda(C_{\mathbf{w}}(h)) \sup_{\mathbf{w} \in \Theta} \psi(\mathbf{w}) < 2\varepsilon$$

with probability tending to one.

It remains to prove (25). It is readily seen that $\int_{C_\mathbf{w}(h)} K^*(\frac{1-\mathbf{w}^T\mathbf{v}}{h})\,d\lambda(\mathbf{v}) = \sum_{j=1}^m \alpha_j\lambda(D_{\mathbf{w},j})$. Hence, for $\varepsilon_3 > 0$ small enough

$$\sup_{\mathbf{w}\in\Theta}\left|c(h,K)\sum_{j=1}^m \alpha_j\Psi(D_{\mathbf{w},j}) - \psi(\mathbf{w})\right|$$

$$\leq \sup_{\mathbf{w}\in\Theta}\psi(\mathbf{w})\left|c(h,K)\sum_{j=1}^m \alpha_j\lambda(D_{\mathbf{w},j}) - 1\right| + \varepsilon_3 c(h,K)\sum_{j=1}^m \alpha_j\lambda(D_{\mathbf{w},j})$$

$$\leq \sup_{\mathbf{w}\in\Theta}\psi(\mathbf{w})\left|\frac{\int_{C_\mathbf{w}(h)} K^*((1-\mathbf{w}^T\mathbf{v})/h)\,d\lambda(\mathbf{v})}{\int_{C_\mathbf{w}(h)} K((1-\mathbf{w}^T\mathbf{v})/h)\,d\lambda(\mathbf{v})} - 1\right|$$

$$\quad + \varepsilon_3 c(h,K)\lambda(C_\mathbf{w}(h))\sum_{j=1}^m \alpha_j$$

$$\leq \eta c(h,K)\lambda(C_\mathbf{w}(h))\sup_{\mathbf{w}\in\Theta}\psi(\mathbf{w}) + \varepsilon_3 c(h,K)\lambda(C_\mathbf{w}(h))\sum_{j=1}^m \alpha_j$$

$$\leq \varepsilon. \qquad\qquad \square$$

From Proposition 2 and the consistency of $\widehat\alpha$, we obtain immediately, as $n \to \infty$,

$$\widehat{\nu(S)} \xrightarrow{\mathbb{P}} \nu(S)$$

and, for $\varepsilon > 0$,

(31) $$\mathbb{P}((1+\varepsilon)S \subset \widehat S \subset (1-\varepsilon)S) \to 1.$$

PROPOSITION 3. *As $n \to \infty$,*

$$\frac{P(\widetilde Q_n \triangle \widehat Q_n)}{p} \xrightarrow{\mathbb{P}} 0.$$

PROOF. Note that as $n \to \infty$, we have

$$\widehat U\left(\frac{n}{k}\right)\Big/ U\left(\frac{n}{k}\right) \xrightarrow{\mathbb{P}} 1,$$

$$(\widehat{\nu(S)})^{1/\widehat\alpha} \xrightarrow{\mathbb{P}} (\nu(S))^{1/\alpha},$$

$$\left(\frac{k}{np}\right)^{1/\widehat\alpha - 1/\alpha} = \exp\left(\frac{\sqrt k(\alpha - \widehat\alpha)}{\widehat\alpha\alpha}\left(\frac{\log k}{\sqrt k} - \frac{\log(np)}{\sqrt k}\right)\right) \xrightarrow{\mathbb{P}} 1.$$

Combining these three limit relations, we obtain

$$\frac{\widehat U(n/k)(k\widehat{\nu(S)}/(np))^{1/\widehat\alpha}}{U(n/k)(k\nu(S)/(np))^{1/\alpha}} \xrightarrow{\mathbb{P}} 1.$$

This and (31) yields that with probability tending to one, as $n \to \infty$,

$$(1 + \varepsilon)^2 \widetilde{Q}_n \subset \widehat{Q}_n \subset (1 - \varepsilon)^2 \widetilde{Q}_n.$$

Then,

$$\frac{P(\widetilde{Q}_n \triangle \widehat{Q}_n)}{p} \leq \frac{1}{p} P\left(U\left(\frac{n}{k}\right) \left(\frac{k\nu(S)}{np}\right)^{1/\alpha} ((1 - \varepsilon)^2 S \setminus (1 + \varepsilon)^2 S) \right),$$

and, by (21), the latter expression tends to

$$\begin{aligned}
\nu((1 - \varepsilon)^2 S &\setminus (1 + \varepsilon)^2 S)/\nu(S) \\
&= \nu((1 - \varepsilon)^2 S)/\nu(S) - \nu((1 + \varepsilon)^2 S)/\nu(S) \\
&= (1 - \varepsilon)^{-2\alpha} - (1 + \varepsilon)^{-2\alpha},
\end{aligned}$$

which in turn tends to 0, as $\varepsilon \downarrow 0$. $\quad\square$

PROOF OF THEOREM 1.   The result follows from Propositions 1 and 3. $\quad\square$

REFERENCES

ALEXANDER, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067. [Corr.: **15** (1987) 428–430.] MR0757769

BAI, Z. D., RAO, C. R. and ZHAO, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.* **27** 24–39. MR0971170

BAÍLLO, A., CUESTA-ALBERTOS, J. A. and CUEVAS, A. (2001). Convergence rates in nonparametric estimation of level sets. *Statist. Probab. Lett.* **53** 27–35. MR1843338

DANIELSSON, J., DE HAAN, L., PENG, L. and DE VRIES, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.* **76** 226–248. MR1821820

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

DE HAAN, L. and RESNICK, S. (1987). On regular variation of probability densities. *Stochastic Process. Appl.* **25** 83–93. MR0904266

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. MR1026315

EINMAHL, J. H. J., LI, J. and LIU, R. Y. (2009). Thresholding events of extreme in simultaneous monitoring of multiple risks. *J. Amer. Statist. Assoc.* **104** 982–992. MR2562001

HALL, P., WATSON, G. S. and CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–762. MR0919843

HASHORVA, E. (2006). On the regular variation of elliptical random vectors. *Statist. Probab. Lett.* **76** 1427–1434. MR2245561

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(94)** 171–192. MR2281913

MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multi-modality. *J. Amer. Statist. Assoc.* **86** 738–746. MR1147099

POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204

RIGOLLET, P. and VERT, R. (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15** 1154–1178. MR2597587

RVAČEVA, E. L. (1962). On domains of attraction of multi-dimensional distributions. *Select. Transl. Math. Statist. Probab.* **2** 183–205. MR0150795

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

SMITH, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15** 1174–1207. MR0902252

TSYBAKOV, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. MR1447735

DEPARTMENT OF ECONOMETRICS & OR AND CENTER
TILBURG UNIVERSITY
P.O. BOX 90153
5000 LE TILBURG
THE NETHERLANDS
E-MAIL: j.cai@tilburguniversity.edu
        j.h.j.einmahl@tilburguniversity.edu
        ldehaan@ese.eur.nl