CentER Discussion Paper

No. 2014-035

ESTIMATION OF EXTREME DEPTH-BASED QUANTILE REGIONS

By

Yi He, John H.J. Einmahl

28 May, 2014

ISSN 0924-7815
ISSN 2213-9532

Tilburg University

Estimation of Extreme Depth-based Quantile Regions

Yi He*　　John H.J. Einmahl

Tilburg University　　*Tilburg University*

May 27, 2014

Abstract

Consider the extreme quantile region, induced by the halfspace depth function HD, of the form $\mathcal{Q} = \{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) \leq \beta\}$, such that $P\mathcal{Q} = p$ for a given, very small $p > 0$. This region can hardly be estimated through a fully nonparametric procedure since the sample halfspace depth is 0 outside the convex hull of the data. Using Extreme Value Theory, we construct a natural, semiparametric estimator of this quantile region and prove a refined consistency result. A simulation study clearly demonstrates the good performance of our estimator. We use the procedure for risk management by applying it to stock market returns.

Key words. Extreme value statistics, halfspace depth, multivariate quantile, outlier detection, rare event, tail dependence.
JEL codes. C13, C14.

1. Introduction

The Depth-Outlyingness-Quantiles-Ranks paradigm of Serfling (2010) implies that the concepts of depth and quantile are essentially equivalent under some regularity conditions for a $\mathbb{R}^d$-valued random vector, say $\boldsymbol{X}$. Any statistical depth function can induce a

*Address for correspondence: Dept. of Econometrics & OR and CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: y.he@tilburguniversity.edu

multivariate quantile function and vice versa, see, e.g., Serfling (2006). A multivariate quantile is not defined uniquely since there are many different choices of depth functions. Here we consider a leading one called the halfspace depth function $HD : \mathbb{R}^d \to [0, \infty)$ defined by

$$HD(\boldsymbol{x}; P) = \inf\{P(H) : \boldsymbol{x} \in H \in \mathcal{H}\}, \quad \boldsymbol{x} \in \mathbb{R}^d,$$

where P is the probability measure of $\boldsymbol{X}$ and $\mathcal{H}$ is the class of closed halfspaces. The extreme depth-based quantile region is of the form

$$\mathcal{Q} = \mathcal{Q}(\boldsymbol{X}, \beta) = \{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) \leq \beta\} \tag{1}$$

with $p = P\mathcal{Q} > 0$ a given, very small number. From now on, without confusion, we use the notations $\mathcal{Q}$, $\mathcal{Q}_{\boldsymbol{X}}$, $\mathcal{Q}(\boldsymbol{X}, \beta)$ and $\mathcal{Q}(\boldsymbol{X}; p)$ interchangeably. The extreme quantile contour is defined accordingly as $\mathcal{C} = \mathcal{C}_\beta = \{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) = \beta\}$. It is compelling that the halfspace quantile region can be also generated without any depth setting by directional quantiles (see Hallin et al., 2010, and Kong and Mizera, 2012).

The extreme depth-based quantile has strong practical values, particularly in economics and finance studies. A direct application is to detect data outliers, which occur with extremely small probability, e.g. financial data corresponding to irregular market behavior such as erroneous trades and financial crises. A second application is to reveal the jointly extreme behavior of multivariate risks. This is important for the risk/portfolio manager to understand the diversifiability between multiple risks/assets. Last but not the least, the extreme depth-based quantile can define the unlikely scenarios for stress testing (McNeil and Smith, 2012).

The purpose of this paper is to estimate the quantile region $\mathcal{Q}$ (or the quantile contour $\mathcal{C}$) from a random sample from P. A natural estimation of $\mathcal{Q}$ is to, simply, exploit the sample depth function. Here, in the spirit of Cai et al. (2011), p is very small and typically of order $1/n$. This means that the number of data points that lie in $\mathcal{Q}$ is small and can even be zero, leaving little information for estimating it. Indeed, the fully nonparametric estimator based on the sample depth will perform poorly, which is demonstrated clearly in our simulation study, since the sample halfspace depth is 0 outside the convex hull of the data.

We consider multivariate regularly varying distributions since our interest is in extreme quantile regions that are far away from the distribution 'center' and the origin.

Assumption 1. *The random vector* $\boldsymbol{X}$ *is said to be multivariate regularly varying if there exists a measure* ν *such that*

$$\frac{\mathbb{P}(\boldsymbol{X} \in tB)}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \to \nu(B) < \infty \tag{2}$$

for every Borel set $B \subset \mathbb{R}^d$ that is bounded away from the origin and satisfies $\nu(\partial B) = 0$. Here $\|\cdot\|$ denotes the L_2-norm and $tB = \{t\boldsymbol{x} : \boldsymbol{x} \in B\}$.

It follows that ν is homogeneous, that is, there exists a $\gamma > 0$ such that for all $t > 0$

$$\nu(tB) = t^{-1/\gamma}\nu(B). \tag{3}$$

The number γ is called the extreme value index. For discussions and interpretations of condition (2), one may refer to Jessen and Mikosch (2006) and Cai et al. (2011). Exploiting this condition we will construct an estimator of $\mathcal{Q}$ based on the statistics of extremes methodology.

Note that the present approach is different from those in Cai et al. (2011) and Einmahl et al. (2013). These two papers deal with likelihood-based quantile regions. The results in there rely on additional regular variation conditions at the density level, which are not required here. Likelihood measures outlyingness merely in a local perspective. Points with small likelihood do not necessary lie far from the origin or the distribution center. An example is given in the discussion in Serfling and Zuo (2010).

This paper is organized as follows. In Section 2, we construct our estimator and show some of its properties and we establish a refined consistency result. Section 3 demonstrates the excellent performance of our estimator in a simulation study while Section 4 presents a real-life financial application. The proofs are deferred to Section 5.

2. Main Results

Consider a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with $\boldsymbol{X}_i \stackrel{d}{=} \boldsymbol{X}$, for $i = 1, \ldots, n$. Define the radii $R = \|\boldsymbol{X}\|$ and $R_i = \|\boldsymbol{X}_i\|$ for $i = 1, \ldots, n$. We order the R_i's as $R_{1:n} \leq \ldots \leq R_{n:n}$. Define $F_R(t) = \mathbb{P}(R \leq t)$ and $U(t) = F_R^{\leftarrow}\left(1 - \frac{1}{t}\right)$. Some continuity properties are required.

Assumption 2. *For any* $\boldsymbol{u} \in \Theta = \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\| = 1\}$, *the distribution function* $F_{\boldsymbol{u}}$ *of the projection* $X(\boldsymbol{u}) = \boldsymbol{u}^T\boldsymbol{X}$ *is continuous. Both functions* F_R *and* U *are continuous. Moreover, for any* $\beta > 0$, $P(\mathcal{C}_\beta) = 0$.

It follows that the halfspace depth function HD is continuous. We also have that the quantile region $\mathcal{Q}(\boldsymbol{X}; p)$ exists:

Proposition 1. *Under Assumption 2, for any* $0 < p < 1$, *it holds that* $P(\mathcal{Q}(\boldsymbol{X}, \beta)) = p$, *where* $\beta = \sup\{\widetilde{\beta} : P(\mathcal{Q}(\boldsymbol{X}, \widetilde{\beta})) \leq p\}$.

We parametrize the halfspace $H = H_{r,\boldsymbol{u}}$ by a pair of parameters $(r, \boldsymbol{u})$ with $r \in \mathbb{R}$ and $\boldsymbol{u} \in \Theta$. Here $\boldsymbol{u}$ is its unit normal vector and r is the lower bound of the inner product between $\boldsymbol{u}$ and points in H. Precisely, we write

$$H_{r,\boldsymbol{u}} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{u}^T \boldsymbol{x} \geq r\}$$

and its collection $\mathcal{H} = \{H_{r,\boldsymbol{u}} : r \in \mathbb{R}, \boldsymbol{u} \in \Theta\}$. Then the halfspace depth function can be written in a simplified way as

$$HD(\boldsymbol{x}, P) = \inf_{\boldsymbol{u} \in \Theta} P(H_{\boldsymbol{u}^T x, \boldsymbol{u}}).$$

Therefore, the extreme quantile region we wish to estimate can be rewritten as

$$\mathcal{Q} = \{\boldsymbol{x} \in \mathbb{R}^d : \inf_{\boldsymbol{u} \in \Theta} P(H_{\boldsymbol{u}^T \boldsymbol{x}, \boldsymbol{u}}) \leq \beta\} = \cup_{r \in \mathbb{R}, \boldsymbol{u} \in \Theta} \{H_{r,\boldsymbol{u}} : P(H_{r,\boldsymbol{u}}) \leq \beta\}$$

where $P\mathcal{Q} = p \in (0, 1)$ with $p = p_n \to 0$ as $n \to \infty$. This means both $\mathcal{Q}$ and β depend on n, that is $\mathcal{Q} = \mathcal{Q}_n$ and $\beta = \beta_n$.

We state some further assumptions.

Assumption 3.

$$\lim_{t \to \infty} \frac{\mathbb{P}(R \geq t)}{t^{-1/\gamma}} = c \in (0, \infty).$$

Whereas Assumption 1 is usually called the first-order condition in Extreme Value Theory, Assumption 3 can be considered as a second-order condition. It is weaker than the usual second-order condition with a negative second order parameter ρ (see Cai et al. 2011 and Theorem 2.3.9 in de Haan and Ferreira 2006).

Usually $\nu(H)$, of a halfspace H, tends to infinity if the distance between H and the origin approaches zero because of the homogeneity property of ν. This, however, can be violated if the tail probability $P(H_{t,\boldsymbol{u}}) = \mathbb{P}(\boldsymbol{X} \in tH_{1,\boldsymbol{u}})$ vanish at a faster rate than that of the radius as $t \to \infty$ at a particular direction $\boldsymbol{u}$, i.e. $\nu(H_{1,\boldsymbol{u}}) = 0$. We exclude this irregular case by assuming

Assumption 4.

$$\inf_{\boldsymbol{u}\in\Theta} \nu(H_{1,\boldsymbol{u}}) =: \delta_0 > 0.$$

Observe that $\nu(H_{r,\boldsymbol{u}}) = \infty$ for any halfspace $H_{r,\boldsymbol{u}}$ with $r \leq 0$. Accordingly define the extreme halfspace depth function by

$$HD(\boldsymbol{z}, \nu) = \inf\{\nu(H) : \boldsymbol{z} \in H \in \mathcal{H}\} = \inf_{\boldsymbol{u}\in\Theta} \nu(H_{\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}), \quad z \neq 0.$$

Then the shape of the extreme quantile region $\mathcal{Q}$ will be shown to be similar to that of the region, which does not depend on n, given by

$$S = \{\boldsymbol{z} \in \mathbb{R}^d : HD(\boldsymbol{z}, \nu) \leq 1\} = \{\boldsymbol{z} = r\boldsymbol{w} : r \geq (HD(\boldsymbol{w}, \nu))^\gamma, \boldsymbol{w} \in \Theta\}.$$

To determine the appropriate inflation factor, we need an intermediate sequence $k_n \in \{1, \ldots, n\}$, that is,

Assumption 5. *$k = k_n$ satisfies $k \to \infty$ and $k/n \to 0$, as $n \to \infty$.*

Then the inflation of S approximating $\mathcal{Q}_n$ is given by

$$\widetilde{\mathcal{Q}}_n = U\left(\frac{n}{k}\right)\left(\frac{k\nu(S)}{np}\right)^\gamma S.$$

which will be shown (in Proposition 4) to satisfy

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

To estimate $\widetilde{\mathcal{Q}}_n$ and hence $\mathcal{Q}$, we need estimators for $U\left(\frac{n}{k}\right)$, γ, $\nu(S)$ and S. We start from

$$\widehat{U}\left(\frac{n}{k}\right) = R_{n-k+1:n}.$$

the k-th largest radius in the data. The extreme value index γ can be estimated using the univariate data of radii by various methods; see, e.g., Hill (1975), Smith (1987) and Dekkers et al. (1989). The typical convergence rate of the estimator $\widehat{\gamma}$ is of order $k^{-1/2}$. For the rest, it is sufficient to provide an estimator of the measure ν, which fully determines both the set S and $\nu(S)$. A natural estimator of $\nu(B)$ on any Borel set B, which is away from the origin, is to use the sample version

$$\widehat{\nu}(B) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left[\frac{\boldsymbol{X}_i}{R_{n-k+1:n}} \in B\right] = \frac{n}{k} P_n(R_{n-k+1:n} B),$$

where P_n is the empirical probability measure of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. However, to recover the homogeneity of ν in our estimation, we adopt another estimator on halfspaces $H_{r,\boldsymbol{u}}$ given by $\widehat{\nu}^*(H_{r,\boldsymbol{u}}) = r_+^{-1/\widehat{\gamma}} \widehat{\nu}(H_{1,\boldsymbol{u}})$ with $r_+ = d(0, H_{r,\boldsymbol{u}}) = \max\{r, 0\}$. Then we define

$$\widehat{S} = \{\boldsymbol{z} = r\boldsymbol{w} : r \geq (HD(\boldsymbol{w}, \widehat{\nu}^*))^{\widehat{\gamma}}, \boldsymbol{w} \in \Theta\}.$$

Collecting all the estimators above we can estimate $\mathcal{Q}_n$ by

$$\widehat{\mathcal{Q}}_n = \widehat{\mathcal{Q}}_n(\boldsymbol{X}; p) = \widehat{U}\left(\frac{n}{k}\right)\left(\frac{k\widehat{\nu}(\widehat{S})}{np}\right)^{\widehat{\gamma}} \widehat{S}$$

and $\mathcal{C}_n$ by

$$\widehat{\mathcal{C}}_n = \left\{\widehat{U}\left(\frac{n}{k}\right)\left(\frac{k\widehat{\nu}(\widehat{S})}{np}\right)^{\widehat{\gamma}} (HD(\boldsymbol{w}, \widehat{\nu}^*))^{\widehat{\gamma}}\boldsymbol{w} : \boldsymbol{w} \in \Theta\right\}.$$

Proposition 2. *Under Assumption 2, the estimated quantile regions have, almost surely, the following properties:*

(a) The complement of $\widehat{\mathcal{Q}}_n$, denoted as $\widehat{\mathcal{Q}}_n^c$, is bounded. When, in addition, also Assumptions 1 and 3-5 hold, then, with probability tending to 1, as $n \to \infty$, $\widehat{\mathcal{Q}}_n^c$ is convex.

(b) Orthogonal and scale equivariance: for any orthogonal $d \times d$ matrix $\boldsymbol{R}$ and $c > 0$, provided the estimator $\widehat{\gamma}$ is (positive) scale invariant (e.g. Hill, 1975; Smith, 1987; Dekkers et al., 1989), it holds that

$$\widehat{\mathcal{Q}}_n(c\boldsymbol{R}\boldsymbol{X}; p) = c\boldsymbol{R}\widehat{\mathcal{Q}}_n(\boldsymbol{X}; p) := \{c\boldsymbol{R}\boldsymbol{x} : \boldsymbol{x} \in \widehat{\mathcal{Q}}_n(\boldsymbol{X}; p)\}.$$

(c) The $\widehat{\mathcal{Q}}_n$ are nested: for $p_1 < p_2$, $\widehat{\mathcal{Q}}_n(\boldsymbol{X}; p_1) \subset \widehat{\mathcal{Q}}_n(\boldsymbol{X}; p_2)$.

We now present our main result.

Theorem 1. *Suppose Assumptions 1 - 5 hold, and $\widehat{\gamma}$ is an estimator such that $\sqrt{k}(\widehat{\gamma} - \gamma) = O_{\mathbb{P}}(1)$. Given that $(\log np)/\sqrt{k} \to 0$ then,*

$$\sup_{\boldsymbol{x} \in \widehat{\mathcal{C}}_n} |\log HD(\boldsymbol{x}, P) - \log \beta| \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{P(\mathcal{Q}_n \Delta \widehat{\mathcal{Q}}_n)}{p} \xrightarrow{\mathbb{P}} 0.$$

Remark 1. The above approach treats p as explicitly given and solves the implicit β. We consider that it is also natural to, instead, have β explicitly given; see, e.g., Hallin et al. (2010) and Kong and Mizera (2012). Then the procedure can be adapted by simply replacing $\nu(S)/p$ by $1/\beta$. Precisely, the estimated region becomes

$$\widehat{\mathcal{Q}}_n^* = \widehat{U}\left(\frac{n}{k}\right)\left(\frac{k}{n\beta}\right)^{\widehat{\gamma}}\widehat{S}$$

and the modified quantile contour $\widehat{\mathcal{C}}_n^*$ can be defined analogously. Proposition 2 and Theorem 1 still hold with $\widehat{\mathcal{Q}}_n$ replaced by $\widehat{\mathcal{Q}}_n^*$, $\widehat{\mathcal{C}}_n$ by $\widehat{\mathcal{C}}_n^*$ and p by β. Note that, here, we even do not require the estimator for $\nu(S)$.

Remark 2. When p is sufficiently small, we can write $\partial\mathcal{Q} = \{\rho(\boldsymbol{w})\boldsymbol{w} : \boldsymbol{w} \in \Theta\}$ and $\widehat{\mathcal{C}}_n = \{\widehat{\rho}(\boldsymbol{w})\boldsymbol{w} : \boldsymbol{w} \in \Theta\}$ with (unique) positive radius functions ρ and $\widehat{\rho}$. Then, using the intermediate results in Section 5, it can be shown that

$$\sup_{\boldsymbol{w}\in\Theta}\left|\frac{\widehat{\rho}(\boldsymbol{w})}{\rho(\boldsymbol{w})} - 1\right| \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{\lambda(\mathcal{Q}_n \Delta \widehat{\mathcal{Q}}_n)}{\lambda(\mathcal{Q}_n^c)} \xrightarrow{\mathbb{P}} 0,$$

where λ denotes Lebesgue measure.

Remark 3. We can separate the choices of k for estimation of γ and the measure ν, respectively k_γ and k_ν, say. Then Theorem 1 requires both k_γ and k_ν to satisfy Assumption 5 while the rest of the conditions are made solely on k_γ. A heuristic procedure for choosing these k-values in practice is given in Cai et al. (2011).

3. Simulation Study

In this section, a simulation study is carried out to evaluate the finite-sample performance of our extreme quantile estimator, for sample size $n = 5000$. Boxplots are presented based on 100 scenarios. We take $k = 400$ for the whole procedure. We consider four multivariate distributions.

- The bivariate Cauchy distribution ($\gamma = 1$) with density

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

- The trivariate Cauchy distribution ($\gamma = 1$) with density

$$f(x) = \frac{1}{\pi^2(1 + x^2 + y^2 + z^2)^2}\,, \qquad (x, y, z) \in \mathbb{R}^3. \tag{5}$$

- A bivariate elliptical distribution ($\gamma = 1/3$) with density

$$f(x, y) = \frac{3(x^2/4 + y^2)^2}{4\pi(1 + (x^2/4 + y^2)^3)^{3/2}}\,, \qquad (x, y) \in \mathbb{R}^2. \tag{6}$$

  The random vector with this distribution can be obtained by a two-step transform of a bivariate Cauchy random vector. The first step is a linear transformation in $\mathbb{R}^2$: $(x, y) \mapsto (2x, y)$. The second step is a transformation of the radius under polar coordinates: $(r, \theta) \mapsto (r^{1/3}, \theta)$.

- An affine transformation of the bivariate Cauchy ($\gamma = 1$) random vector $\boldsymbol{Y}$:

$$\boldsymbol{X} = \boldsymbol{AY} + \boldsymbol{b}, \quad \boldsymbol{A} = \begin{bmatrix} 2 & 0.3 \\ 0.3 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}. \tag{7}$$

Figure 1 shows the true and estimated extreme quantile regions/contours for $p = 1/2000$, $1/5000$, or $1/10000$. The estimated regions are close to the true ones. It is clear that our (estimated) extreme quantile regions belong to an 'almost empty' space, i.e., a space with few or even no observations.

Next, we compare our extreme estimator to the (fully) nonparametric estimator, which is directly based on the sample depth function. We take the boundary of the sample quantile contour with $\beta = 1/n$, i.e., the convex hull of the observations, as an estimator of the quantile contour. Its outer region is then an estimator for the corresponding quantile region. Since now the β's are explicitly given we use the modified estimator $\widehat{\mathcal{Q}}_n^*$ for the extreme quantile region, see Remark 1. Figure 2 gives an example. Obviously our 'extreme' estimator completely outperforms the nonparametric one.

Tables 1 - 2 and Figures 3 - 4 clearly demonstrate the good performance of the 'extreme' estimator in this simulation study. The fully nonparametric estimator produces relative errors with much larger medians and ranges for all the distributions we consider.

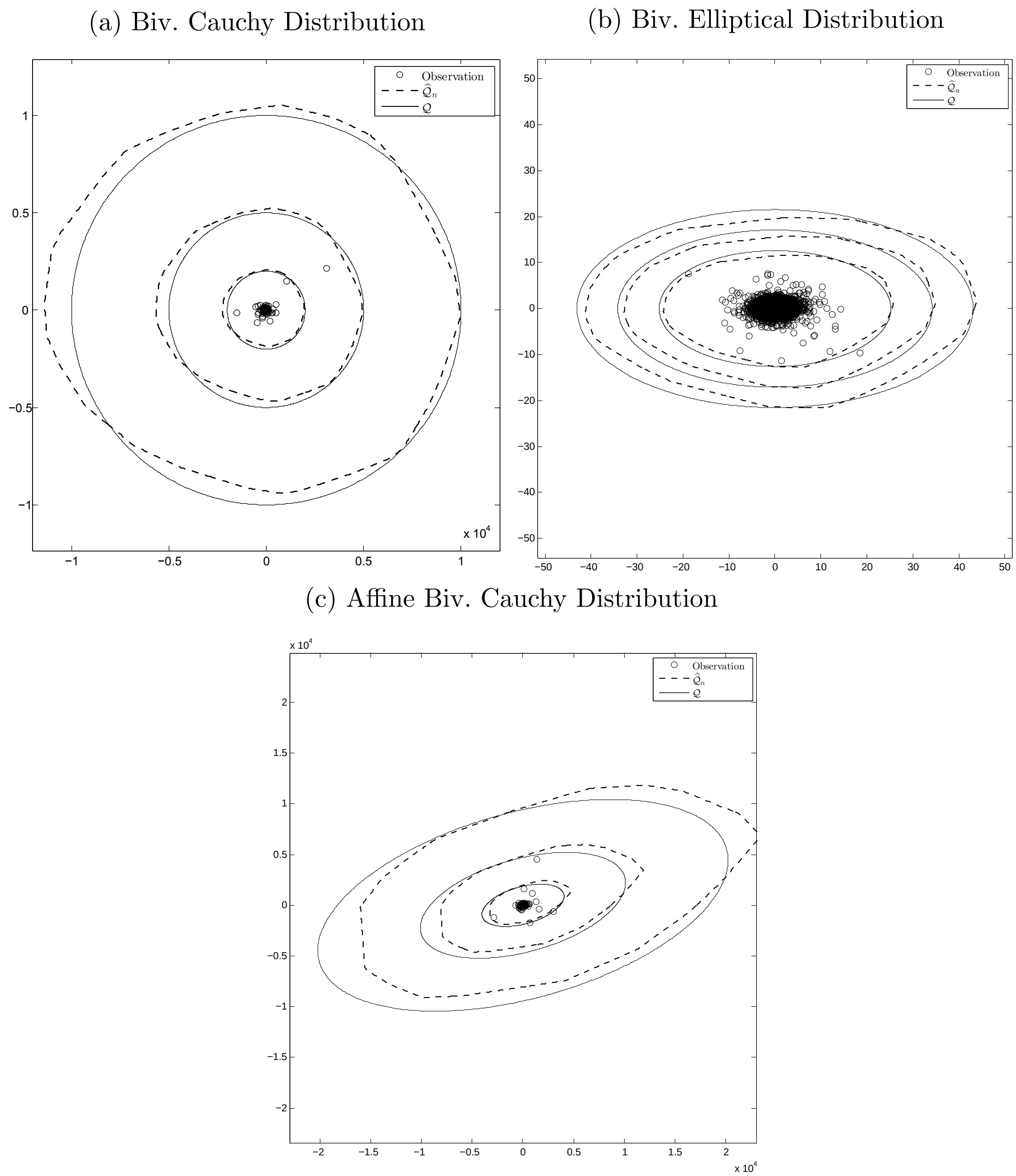


Figure 1: True and estimated quantile regions of the bivariate Cauchy, the bivariate elliptical distribution and the affine transform of the bivariate Cauchy distribution for $p = 1/2000$, $1/5000$ or $1/10000$ based on one sample of size 5000, for $k = 400$.

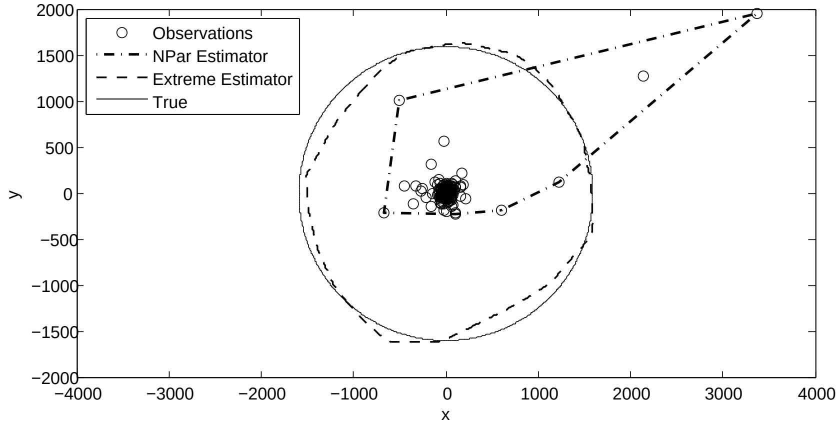
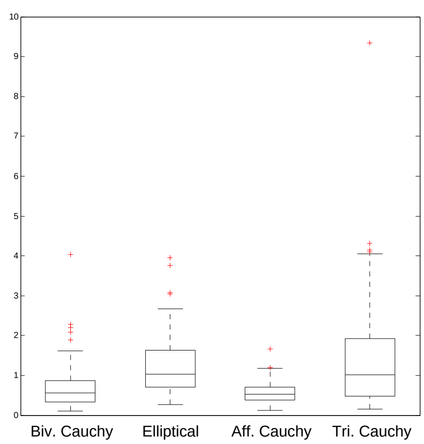


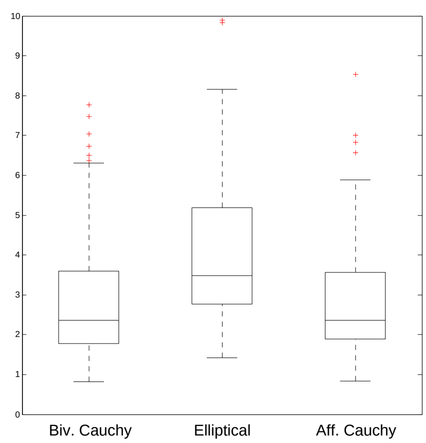
Figure 2: True and estimated quantile regions of bivariate Cauchy distribution for $\beta = 1/5000$ based on one sample of size 5000, for $k = 400$.

Distribution	Biv. Cauchy	Elliptical	Aff. Cauchy	Triv. Cauchy
Extreme	0.556	1.034	0.521	1.019
NPar	2.361	3.487	2.358	-

Table 1: Median of the relative errors $P(\widehat{\mathcal{Q}}_n^* \Delta \mathcal{Q}_n)/\beta$ for $\beta = 1/5000$ based on 100 samples of size 5000 for $k = 400$.



(a) Extreme Estimator

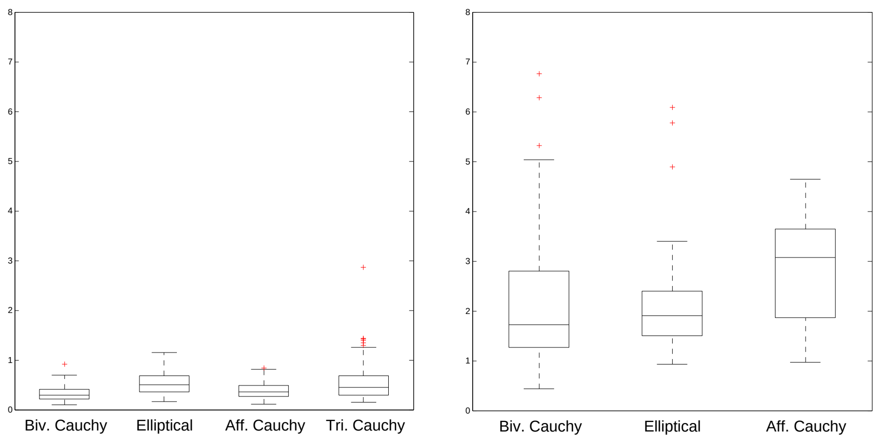


(b) Nonparametric Estimator

Figure 3: Boxplots of $P(\widehat{\mathcal{Q}}_n^* \Delta \mathcal{Q}_n)/\beta$ for the extreme and nonparametric estimators, based on 100 simulated data sets of size 5000 with $\beta = 1/5000$.

Distribution	Biv. Cauchy	Elliptical	Aff. Cauchy	Triv. Cauchy
Extreme	0.293	0.501	0.357	0.447
NPar	1.723	1.906	3.070	-

Table 2: Median of the relative errors $\sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n^*} |\log HD(\boldsymbol{x}, P) - \log \beta|$ for $\beta = 1/5000$ based on 100 samples of size 5000 for $k = 400$.



(a) Extreme Estimation (b) Nonparametric Estimation

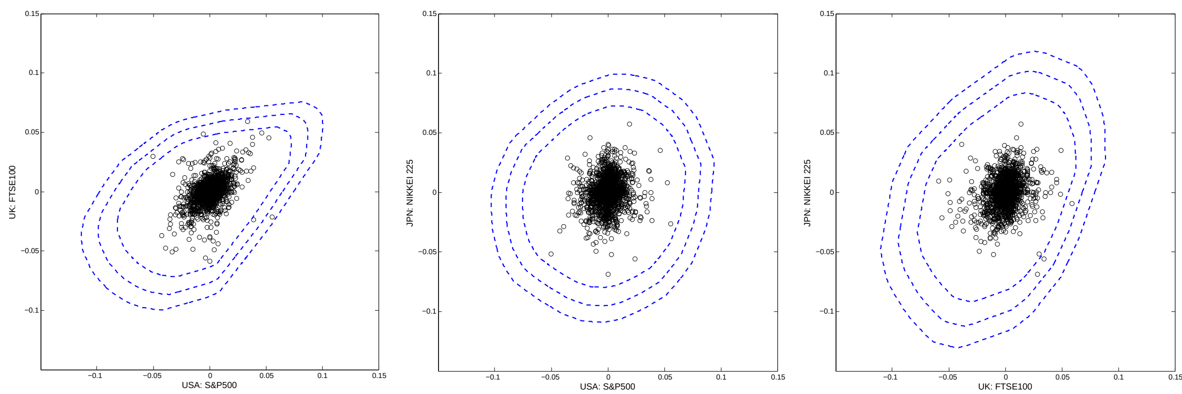
Figure 4: Boxplots of $\sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n^*} |\log HD(\boldsymbol{x}, P) - \log \beta|$ for the extreme and nonparametric estimators, based on 100 simulated data sets of size 5000 with $\beta = 1/5000$.

4. Application

In this section, we present a real-world finance application. The dataset, downloaded from Datastream, consists of the dividend-adjusted daily international market price indexes of S&P 500 from the United States, FTSE 100 from the United Kingdom and Nikkei 225 from Japan. The sample period is from July 2nd 2001 to December 31st 2007. The daily market return is then computed as the logarithm of the ratio of current and one-period ago price, giving rise to 1695 observations for each country.

As usual, the squared stock returns exhibits moderate autocorrelation and the Ljung-Box test rejects the serial independence for all these univariate datasets. Hence, we cannot work with the raw data since the i.i.d. assumption may be inappropriate. A solution is to, instead, work on the 'innovations', which can be obtained by filtering out the clustering and leverage effects from the raw return data. For each return time series, we apply an asymmetric Beta-t-EGARCH(1,1) filtration on the demeaned market returns; see Chapter 4 in Harvey (2013). The Beta-t-EGARCH filter better matches the heavy-tailed data than the Gaussian models. Now the Ljung-Box tests do not reject the serial independence of the original, absolute, nor squared sample innovations. The estimated innovations will also be called the filtered returns.

Next, we check the equality of the extreme value indices for the positive and negative tails of the univariate returns, implied by Assumption 1. The Hill estimates, for $k = 60$, for the right and left tails of the filtered returns in all three markets, in increasing order, are: 0.1566, 0.1642, 0.1792, 0.1948, 0.2263, 0.2495. The maximal difference is 0.0928, which is smaller than the 95%-level approximate upper bound 0.1015 based on the asymptotic normality of the Hill estimator. Hence, there is no evidence for the inequality of these six extreme value indices.



(a) U.S. vs U.K. (b) U.S. vs Japan (c) U.K. vs Japan

Figure 5: Predicted bivariate quantile regions of raw returns on Jan. 1, 2008 for p=1/2000, 1/5000, 1/10000 based on the index price data from July 2, 2001 to Dec. 31, 2007.

Figure 5 shows the predicted bivariate extreme quantile regions/contours of raw returns for $p = 1/2000$, $1/5000$, $1/10000$ for January 1, 2008 (that is, one day ahead) for every pair of markets while Figure 6 shows, using the spherical coordinate system, the trivariate extreme quantile region/contour of filtered returns for $p = 1/10000$ for all three markets. We take $k = 180$ for all the three bivariate cases and $k_\gamma = 200$ and $k_\nu = 300$ for the trivariate case (see Remark 3). The estimate of the (time-varying) extreme quantile region for raw returns is transformed from the one for filtered returns by the affine invariance of the halfspace depth, see, e.g., Zuo and Serfling (2000).

These figures convey crucial information to the risk manager. The (estimated) extreme quantile regions reveal which combinations of extreme returns can occur and which cannot. Neglecting the joint behavior can lead to overestimated diversifiability of risks across international markets and therefore underestimation of the true level of risks.

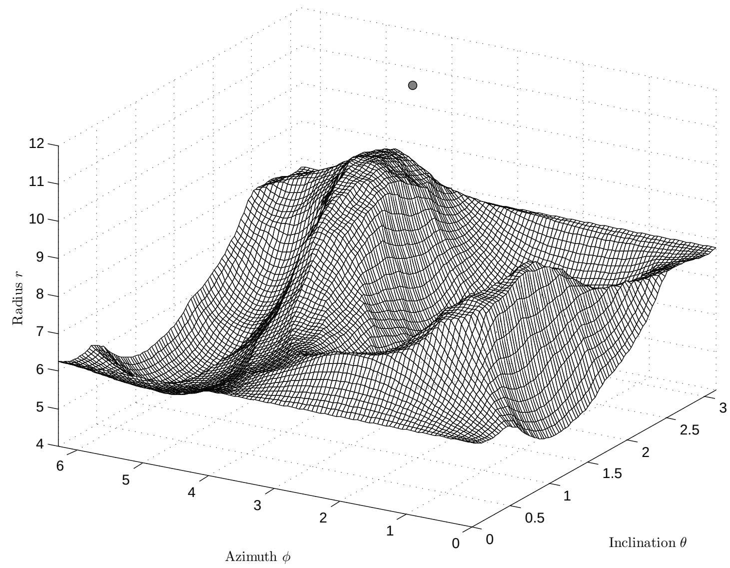


Figure 6: Estimated trivariate quantile regions of filtered returns with p=1/10000.

Another application of the depth-based extreme quantile regions is to detect data outliers. Here, we consider a practical definition of outlier, namely that a data point has a rare joint behavior: more precisely, it lies in the (estimated) quantile region with an extremely small p, say 1/10000. Note that an outlier in a high dimensional space is not necessarily an outlier in its subspaces with reduced dimensions. This means the outcome depends on the choice of the data space. In our sample, we observe the biggest loss in the US market on February 27, 2007. This drop is mainly due to the 'Chinese correction', i.e., the burst of the Chinese stock bubble of 2007. On the same day the SSE Composite index of the Shanghai Stock Exchange dropped by 9%, breaking the 10-year record. We observe, from Figure 6, that this data point is inside the extreme trivariate quantile region, i.e. the upper space of the surface, consisting of all three markets with $p = 1/10000$. We conclude that this point is an outlier in the three-dimensional space.

5. Proofs

Proof of Proposition 1. Let $0 < p < 1$ and $\beta = \sup\{\widetilde{\beta} : P(\mathcal{Q}(\boldsymbol{X}, \widetilde{\beta})) \leq p\} \in (0, 1)$. Notice that $0 < \beta < 1$. Take a sequence of positive numbers $\{\beta_m^-\}_{m=1}^\infty$ such that $\beta_m^- \uparrow \beta$ as $m \to \infty$. It follows that $\{\mathcal{Q}(\boldsymbol{X}, \beta_m^-)\}_{m=1}^\infty$ is an increasing sequence of sets. Therefore

$$P\left(\bigcup_{m=1}^\infty \mathcal{Q}(\boldsymbol{X}, \beta_m^-)\right) = \lim_{m\to\infty} P(\mathcal{Q}(\boldsymbol{X}, \beta_m^-)) \leq p.$$

It is easy to verify that

$$\bigcup_{m=1}^\infty \mathcal{Q}(\boldsymbol{X}, \beta_m^-) = \left\{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) < \beta\right\} = \mathcal{Q}(\boldsymbol{X}, \beta) \setminus \mathcal{C}_\beta.$$

Hence $P(\mathcal{Q}(\boldsymbol{X}, \beta)) = P(\mathcal{Q}(\boldsymbol{X}, \beta) \setminus \mathcal{C}_\beta) \leq p$ by Assumption 2. On the other hand, taking a number sequence $\{\beta_m^+\}_{n=1}^\infty$ such that $\beta_m^+ \downarrow \beta^*$ as $m \to \infty$, analogously, it holds that

$$p \leq \lim_{m\to\infty} P(\mathcal{Q}(\boldsymbol{X}, \beta_m^+)) = P\left(\bigcap_{m=1}^\infty \mathcal{Q}(\boldsymbol{X}, \beta_m^+)\right) = P(\mathcal{Q}(\boldsymbol{X}, \beta)).$$

It follows that $P(\mathcal{Q}(\boldsymbol{X}, \beta)) = p$. □

Proof of Proposition 2. (a) For boundedness and convexity we only need to examine $\widehat{S}^c$. The boundedness holds since, almost surely,

$$HD(\boldsymbol{w}, \widehat{\nu}^*) \leq \widehat{\nu}(H_{1,\boldsymbol{w}}) \leq 1, \quad \boldsymbol{w} \in \Theta.$$

Note that $\inf_{\boldsymbol{u}\in\Theta} \widehat{\nu}(H_{1,\boldsymbol{u}}) > 0$ with probability tending to 1, see (8) below and Assumption 4. Now, if $\inf_{\boldsymbol{u}\in\Theta} \widehat{\nu}(H_{1,\boldsymbol{u}}) > 0$,

$$\widehat{S} = \{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, \widehat{\nu}^*) \leq 1\} = \cup_{r\in\mathbb{R},\boldsymbol{u}\in\Theta}\{H_{r,\boldsymbol{u}} : \widehat{\nu}^*(H_{r,\boldsymbol{u}}) \leq 1\}.$$

Hence,

$$\widehat{S}^c = \{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, \widehat{\nu}^*) > 1\} = \cap_{r\in\mathbb{R},\boldsymbol{u}\in\Theta}\{H_{r,\boldsymbol{u}}^c : \widehat{\nu}^*(H_{r,\boldsymbol{u}}) \leq 1\},$$

where $H_{r,\boldsymbol{u}}^c = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{u}^T\boldsymbol{x} < r\}$, is convex. Then the convexity of $\widehat{Q}_n^c$ follows.

(b) It suffices to prove the orthogonal and scale equivariance separately. The orthogonal transformation has no impact on the radii $R_1, \ldots, R_n$ of the random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. It is easy to verify that the only change is $\widehat{S}_{\boldsymbol{AX}} = \boldsymbol{A}\widehat{S}_{\boldsymbol{X}}$, then the orthogonal equivariance follows. The scale equivariance comes in a similar way by using the facts $\widehat{U}_{c\boldsymbol{X}}(n/k) = c\widehat{U}_{\boldsymbol{X}}(n/k)$ and other components of the estimate remain the same.

(c) Straightforward. □

To prove Theorem 1, we need some lemmas and propositions. In the sequel we will always assume that the conditions of Theorem 1 hold.

Lemma 1.

$$\lim_{t\to\infty}\sup_{\boldsymbol{u}\in\Theta}\left|\frac{\mathbb{P}(\boldsymbol{X}\in tH_{1,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\|\geq t)}-\nu(H_{1,\boldsymbol{u}})\right|=0$$

.

Proof. Lemma 1 in Einmahl et al. (2014) yields

$$\lim_{t\to\infty}\sup_{\boldsymbol{u}\in\Theta}\left|\frac{\mathbb{P}(\boldsymbol{X}\in tH_{1,\boldsymbol{u}})}{ct^{-1/\gamma}}-\nu(H_{1,\boldsymbol{u}})\right|=0.$$

Now the result follows from Assumption 3. □

Lemma 2. *For any $\varepsilon>0$,*

$$\lim_{t\to\infty}\sup_{H\in\mathcal{H}^\varepsilon}\left|\frac{\mathbb{P}(\boldsymbol{X}\in tH)}{\mathbb{P}(\|\boldsymbol{X}\|\geq t)}-\nu(H)\right|=0,$$

where $\mathcal{H}^\varepsilon=\{H_{r,\boldsymbol{u}}\in\mathcal{H}:r\geq\varepsilon\}$.

Proof. For $r\geq\varepsilon>0$,

$$\left|\frac{\mathbb{P}(\boldsymbol{X}\in tH_{r,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\|\geq t)}-\nu(H_{r,\boldsymbol{u}})\right|\leq\frac{\mathbb{P}(\boldsymbol{X}\in trH_{1,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\|\geq tr)}\left|\frac{\mathbb{P}(\|\boldsymbol{X}\|\geq tr)}{\mathbb{P}(\|\boldsymbol{X}\|\geq t)}-r^{-1/\gamma}\right|$$
$$+\,r^{-1/\gamma}\left|\frac{\mathbb{P}(\boldsymbol{X}\in trH_{1,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\|\geq tr)}-\nu(H_{1,\boldsymbol{u}})\right|.$$

Then result follows from Assumption 3 and Lemma 1 [cf. Theorem 2.1 in de Haan and Resnick (1987)]. □

Lemma 3. *Let δ be a constant such that $0<\delta<\delta_0^\gamma$. Take any $\varepsilon>0$. For $\boldsymbol{z}\in\mathbb{R}^d$ with $\|\boldsymbol{z}\|\geq\varepsilon$,*

$$HD(\boldsymbol{z},\nu)=\inf_{\boldsymbol{u}^T\boldsymbol{z}\geq\delta\varepsilon}\nu(H_{\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})$$

and there exists a $M>0$, which only depends on $\delta\varepsilon$, such that for all $t\geq M$

$$HD(t\boldsymbol{z},P)=\inf_{\boldsymbol{u}^T\boldsymbol{z}\geq\delta\varepsilon}P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}).$$

Proof. We only prove the second part. The proof of the first part is similar. For $\|\boldsymbol{z}\| \geq \varepsilon$, by Lemma 2 we have

$$\inf_{\boldsymbol{u}^T\boldsymbol{z}<\delta\varepsilon} \left\{ \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \right\} \geq \inf_{\boldsymbol{u}\in\Theta} \left\{ \frac{P(H_{t\delta\varepsilon,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \right\} \to \inf_{\boldsymbol{u}\in\Theta} \nu\left(H_{\delta\varepsilon,\boldsymbol{u}}\right) = (\delta\varepsilon)^{-1/\gamma}\delta_0$$

and

$$\inf_{\boldsymbol{u}\in\Theta} \left\{ \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \right\} \leq \frac{P(H_{t\|\boldsymbol{z}\|,\boldsymbol{z}/\|\boldsymbol{z}\|})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \to \nu(H_{\|\boldsymbol{z}\|,\frac{\boldsymbol{z}}{\|\boldsymbol{z}\|}}) = \|\boldsymbol{z}\|^{-\alpha}\, \nu(H_{1,\frac{\boldsymbol{z}}{\|\boldsymbol{z}\|}}) \leq \varepsilon^{-1/\gamma},$$

where we use the fact ν is a probability measure on $\{\boldsymbol{z} \in \mathbb{R}^d : \|\boldsymbol{z}\| \geq 1\} \supset H_{1,\boldsymbol{u}}$ for all $\boldsymbol{u} \in \Theta$. Note that $(\delta\varepsilon)^{-1/\gamma}\delta_0 > \varepsilon^{-1/\gamma}$. It then follows from Lemma 2 that there exists a $M = M_{\delta\varepsilon} > 0$, such that for all $t \geq M$

$$\inf_{\boldsymbol{u}^T\boldsymbol{z}<\delta\varepsilon} \left\{ \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},u})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \right\} > \frac{1}{2}((\delta\varepsilon)^{-1/\gamma}\delta_0 + \varepsilon^{-1/\gamma}) > \inf_{\boldsymbol{u}\in\Theta} \left\{ \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \right\}.$$

It implies $\inf_{\boldsymbol{u}^T\boldsymbol{z}<\delta\varepsilon} P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}) > \inf_{\boldsymbol{u}\in\Theta} P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}) = HD(t\boldsymbol{z}, P)$ and consequently the second part of the lemma. □

Proposition 3. *For any $\varepsilon > 0$,*

$$\lim_{t\to\infty} \sup_{\|\boldsymbol{z}\|\geq\varepsilon} \left| \frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}\left(\|\boldsymbol{X}\| \geq t\right)} - HD(\boldsymbol{z}, \nu) \right| = 0.$$

Proof. Take a constant δ such that $0 < \delta < \delta_0^\gamma$. From Lemma 3, we know for sufficiently large t

$$\begin{aligned}\sup_{\|\boldsymbol{z}\|\geq\varepsilon} \left| \frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}\left(\|\boldsymbol{X}\| \geq t\right)} - HD(\boldsymbol{z}, \nu) \right| &= \sup_{\|\boldsymbol{z}\|\geq\varepsilon} \left| \inf_{\boldsymbol{u}^T\boldsymbol{z}\geq\delta\varepsilon} \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}\left(\|\boldsymbol{X}\| \geq t\right)} - \inf_{\boldsymbol{u}^T\boldsymbol{z}\geq\delta\varepsilon} \nu(H_{\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}) \right| \\ &\leq \sup_{\|\boldsymbol{z}\|\geq\varepsilon} \sup_{\boldsymbol{u}^T\boldsymbol{z}\geq\delta\varepsilon} \left| \frac{P(H_{t\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}})}{\mathbb{P}\left(\|\boldsymbol{X}\| \geq t\right)} - \nu(H_{\boldsymbol{u}^T\boldsymbol{z},\boldsymbol{u}}) \right|.\end{aligned}$$

The rest follows by Lemma 2. □

Lemma 4. *For each $\varepsilon > 0$, there exists $t_0 > 0$ such that for $t > t_0$*

$$\left\{ \boldsymbol{z} \in \mathbb{R}^d : \frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} \leq \varepsilon \right\} \subset \left\{ \boldsymbol{z} \in \mathbb{R}^d : \|\boldsymbol{z}\| > \left(\frac{\delta_0}{2\varepsilon}\right)^\gamma \right\}.$$

Proof. It suffices to prove, for large t

$$\left\{ \boldsymbol{z} \in \mathbb{R}^d : \|\boldsymbol{z}\| \leq \left(\frac{\delta_0}{2\varepsilon}\right)^\gamma \right\} \subset \left\{ \boldsymbol{z} \in \mathbb{R}^d : \frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}(\|\boldsymbol{X}\| \geq t)} > \varepsilon \right\}.$$

Write $\delta = (\delta_0/2\varepsilon)^\gamma$. Take any $\boldsymbol{z} \in \mathbb{R}^d$ with $\|\boldsymbol{z}\| \le \delta$. Lemma 2 yields

$$\inf_{\boldsymbol{u}\in\Theta} \frac{P(tH_{\delta,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} \to \inf_{\boldsymbol{u}\in\Theta} \nu(H_{\delta,\boldsymbol{u}}) = \delta^{-1/\gamma} \inf_{\boldsymbol{u}\in\Theta} \nu(H_{1,\boldsymbol{u}}) = \delta^{-1/\gamma}\delta_0 = 2\varepsilon.$$

Hence there exists a $t_0 > 0$ such that for $t > t_0$

$$\frac{HD(t\boldsymbol{z}, P)}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} = \inf_{\boldsymbol{u}\in\Theta} \frac{P(tH_{\boldsymbol{u}^T z,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} \ge \inf_{\boldsymbol{u}\in\Theta} \frac{P(tH_{\delta,\boldsymbol{u}})}{\mathbb{P}(\|\boldsymbol{X}\| \ge t)} > 2\varepsilon - \varepsilon = \varepsilon.$$

□

Lemma 5. *As $n \to \infty$, $p/\beta \to \nu(S)$.*

Proof. Under Assumption 2, $\mathbb{P}(\|\boldsymbol{X}\| \ge U(1/\beta)) = \beta$. Hence,

$$\begin{aligned} p =& P(\mathcal{Q}_n) = P(\{\boldsymbol{x} \in \mathbb{R}^d : HD(\boldsymbol{x}, P) \le \beta\}) \\ =& P(\{U(1/\beta)\boldsymbol{z} : HD(U(1/\beta)\boldsymbol{z}, P) \le \beta\}) \\ =& P\left(U(1/\beta)\left\{\boldsymbol{z} : \frac{HD(U(1/\beta)z, P)}{\mathbb{P}(\|X\| \ge U(1/\beta))} \le 1\right\}\right). \end{aligned}$$

By Lemma 4 we know, when n is large,

$$S_n := \left\{\boldsymbol{z} \in \mathbb{R}^d : \frac{HD(U(1/\beta)\boldsymbol{z}; P)}{\mathbb{P}(\|\boldsymbol{X}\| \ge U(1/\beta))} \le 1\right\} \subset \left\{\boldsymbol{z} \in \mathbb{R}^d : \|\boldsymbol{z}\| > \left(\frac{\delta_0}{2}\right)^\gamma\right\}.$$

Then Proposition 3 yields that for any $\varepsilon > 0$ there exists a $M = M_\varepsilon$ such that when $n > M$,

$$(1+\varepsilon)S \subset S_n \subset (1-\varepsilon)S.$$

Therefore

$$\lim_{n\to\infty} \left|\frac{p}{\beta} - \nu(S)\right| \le \nu((1-\varepsilon)S) - \nu((1+\varepsilon)S) = \nu(S)((1-\varepsilon)^{-1/\gamma} - (1+\varepsilon)^{-1/\gamma}),$$

which immediately implies our result since ε is arbitrary. □

Proposition 4.

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

Proof. Let $\varepsilon > 0$. Note that by Proposition 3, Lemma 4 and Lemma 5, analogously as Lemmas 3 and 4 in Cai et al. (2011), for large n,

$$U\left(\frac{\nu(S)}{p}\right)\{\boldsymbol{z} : HD(\boldsymbol{z}, \nu) \le 1 - \varepsilon\} \subset \mathcal{Q}_n \subset U\left(\frac{\nu(S)}{p}\right)\{\boldsymbol{z} : HD(\boldsymbol{z}, \nu) \le 1 + \varepsilon\}$$

and

$$U\left(\frac{\nu(S)}{p}\right)\{\boldsymbol{z}: HD(\boldsymbol{z},\nu)\le 1-\varepsilon\}\subset \widetilde{\mathcal{Q}}_n \subset U\left(\frac{\nu(S)}{p}\right)\{\boldsymbol{z}: HD(\boldsymbol{z},\nu)\le 1+\varepsilon\}.$$

It follows that

$$\begin{aligned}\frac{\nu(S)}{p}P(\mathcal{Q}_n\Delta\widetilde{\mathcal{Q}}_n) &\le \frac{\nu(S)}{p}P\left(U\left(\frac{\nu(S)}{p}\right)\{\boldsymbol{z}: 1-\varepsilon\le HD(\boldsymbol{z},\nu)\le 1+\varepsilon\}\right)\\ &\to \nu\left(\{\boldsymbol{z}: 1-\varepsilon\le HD(\boldsymbol{z},\nu)\le 1+\varepsilon\}\right)=2\varepsilon\nu(S).\end{aligned}$$

Since ε is arbitrary, $P(\mathcal{Q}_n\Delta\widetilde{\mathcal{Q}}_n)/p\to 0$. □

Proposition 5. *As $n\to\infty$,*

$$\sup_{\boldsymbol{w}\in\Theta}\left|\frac{HD(\boldsymbol{w},\widehat{\nu}^*)}{HD(\boldsymbol{w},\nu)}-1\right|\xrightarrow{\mathbb{P}}0.$$

Proof. First we show

$$\sup_{\boldsymbol{u}\in\Theta}|\widehat{\nu}(H_{1,\boldsymbol{u}})-\nu(H_{1,\boldsymbol{u}})|\xrightarrow{\mathbb{P}}0. \tag{8}$$

Note that

$$\begin{aligned}&\sup_{\boldsymbol{u}\in\Theta}|\widehat{\nu}(H_{1,\boldsymbol{u}})-\nu(H_{1,\boldsymbol{u}})|\\ &\le \sup_{\boldsymbol{u}\in\Theta}\left|\frac{n}{k}P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)-\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)\right|+\sup_{\boldsymbol{u}\in\Theta}\left|\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)-\nu(H_{1,\boldsymbol{u}})\right|\\ &\le \sup_{\boldsymbol{u}\in\Theta}\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|+\sup_{\boldsymbol{u}\in\Theta}\left|\frac{n}{k}P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)-\nu(H_{1,u})\right|.\end{aligned}$$

From Lemma 1 we know for (8) it suffices to show

$$\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|=\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}\frac{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|\xrightarrow{\mathbb{P}}0.$$

In other words, it suffices to show

$$\mathrm{I}:=\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P_n\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|\xrightarrow{\mathbb{P}}0 \text{ and } \mathrm{II}:=\sup_{\boldsymbol{u}\in\Theta}\left|\frac{P\left(\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\right)}-1\right|\xrightarrow{\mathbb{P}}0.$$

For any $0<\eta<1$, define events $\Omega_n=\{(1-\eta)U\left(\frac{n}{k}\right)\le\widehat{U}\left(\frac{n}{k}\right)\le(1+\eta)U\left(\frac{n}{k}\right)\}$. Then it follows that $\mathbb{P}(\Omega_n)\to 1$ as $n\to\infty$ since $\widehat{U}\left(\frac{n}{k}\right)/U\left(\frac{n}{k}\right)\xrightarrow{\mathbb{P}}1$ by Assumption 2. On Ω_n, we have

$$(1-\eta)U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\subset\widehat{U}\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}\subset(1+\eta)U\left(\frac{n}{k}\right)H_{1,\boldsymbol{u}}.$$

Notice that, denoting $\mathcal{H}_{1+\eta} = \{H_{r,\boldsymbol{u}} \in \mathcal{H} : r \leq 1+\eta\}$,

$$\inf_{H \in \mathcal{H}_{1+\eta}} \frac{n}{k} P\left(U\left(\frac{n}{k}\right) H\right) \to \inf_{H \in \mathcal{H}_{1+\eta}} \nu(H) = (1+\eta)^{-1/\gamma} \inf_{\boldsymbol{u} \in \Theta} \nu(H_{1,\boldsymbol{u}}) =: 2\delta.$$

Then from Theorem 5.1 in Alexander (1987) we have, if n is large

$$\mathrm{I} \leq \sup_{H \in \mathcal{H}_{1+\eta}} \left| \frac{P_n\left(U\left(\frac{n}{k}\right) H\right)}{P\left(U\left(\frac{n}{k}\right) H\right)} - 1 \right| \leq \sup_{P(H) \geq \frac{k\delta}{n}} \left| \frac{P_n(H)}{P(H)} - 1 \right| \xrightarrow{\mathbb{P}} 0.$$

On the other hand, on Ω_n, for any $u \in \Theta$

$$\frac{P\left(U\left(\frac{n}{k}\right) H_{1+\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)} \leq \frac{P\left(\widehat{U}\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)} \leq \frac{P\left(U\left(\frac{n}{k}\right) H_{1-\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)}.$$

Then

$$\begin{aligned}
\mathrm{II} &\leq \sup_{u \in \Theta} \left\{ \max\left\{ 1 - \frac{P\left(U\left(\frac{n}{k}\right) H_{1+\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)}, \frac{P\left(U\left(\frac{n}{k}\right) H_{1-\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)} - 1 \right\} \right\} \\
&\leq \max\left\{ 1 - \inf_{\boldsymbol{u} \in \Theta} \frac{P\left(U\left(\frac{n}{k}\right) H_{1+\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)}, \sup_{\boldsymbol{u} \in \Theta} \frac{P\left(U\left(\frac{n}{k}\right) H_{1-\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)} - 1 \right\} \\
&=: \max\{\mathrm{II}_1, \mathrm{II}_2\}.
\end{aligned}$$

Note that

$$\inf_{\boldsymbol{u} \in \Theta} \frac{P\left(U\left(\frac{n}{k}\right) H_{1+\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)} = \inf_{\boldsymbol{u} \in \Theta} \frac{\frac{n}{k} P\left(U\left(\frac{n}{k}\right) H_{1+\eta,\boldsymbol{u}}\right)}{\frac{n}{k} P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)} \to \inf_{\boldsymbol{u} \in \Theta} \frac{\nu(H_{1+\eta,\boldsymbol{u}})}{\nu(H_{1,\boldsymbol{u}})} = (1+\eta)^{-1/\gamma}$$

and similarly

$$\sup_{\boldsymbol{u} \in \Theta} \frac{P\left(U\left(\frac{n}{k}\right) H_{1-\eta,\boldsymbol{u}}\right)}{P\left(U\left(\frac{n}{k}\right) H_{1,\boldsymbol{u}}\right)} \to (1-\eta)^{-1/\gamma}.$$

Since η can be arbitrarily small, it follows from above that both II_1 and II_2 can be arbitrarily small when n is sufficiently large. This implies that $\mathrm{II} \xrightarrow{\mathbb{P}} 0$. Hence (8) holds. For the rest now it is sufficient to show

$$\sup_{\boldsymbol{w} \in \Theta} |HD(\boldsymbol{w}, \widehat{\nu}^*) - HD(\boldsymbol{w}, \nu)| \xrightarrow{\mathbb{P}} 0. \tag{9}$$

Note that, if (9) is true, we are done since $\inf_{\boldsymbol{w} \in \Theta} HD(w, \nu) = \inf_{\boldsymbol{w} \in \Theta} \inf_{\boldsymbol{u} \in \Theta} \{\nu(H_{\boldsymbol{u}^T \boldsymbol{w}, \boldsymbol{u}})\} \geq \inf_{\boldsymbol{u} \in \Theta} \{\nu(H_{1,\boldsymbol{u}})\} > 0$. Take a $\delta > 0$ such that $0 < \delta < \delta_0^\gamma \leq \delta_0$. From Lemma 3 we know for $\boldsymbol{w} \in \Theta$

$$HD(\boldsymbol{w}, \nu) = \inf_{\boldsymbol{u}^T \boldsymbol{w} \geq \delta} \{\nu(H_{\boldsymbol{u}^T \boldsymbol{w}, \boldsymbol{u}})\}.$$

Next, define events $\widetilde{\Omega}_n = \{\inf_{u\in\Theta} \widehat{\nu}(H_{1,u}) > \delta\}$. It holds that $\mathbb{P}(\widetilde{\Omega}_n) \to 1$ by the uniform convergence of $\widehat{\nu}$ from (8). Analogously to Lemma 3 we also have

$$HD(\boldsymbol{w}, \widehat{\nu}^*) = \inf_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \{\widehat{\nu}^*(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}})\}.$$

Then, noting that $\boldsymbol{u}^T\boldsymbol{w} \leq 1$ for $\boldsymbol{u}, \boldsymbol{w} \in \Theta$,

$$\begin{aligned}
&\sup_{\boldsymbol{w}\in\Theta} |HD(\boldsymbol{w}, \widehat{\nu}^*) - HD(\boldsymbol{w}, \nu)| \\
&= \sup_{\boldsymbol{w}\in\Theta} \left| \inf_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \widehat{\nu}^*(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}}) - \inf_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \nu(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}}) \right| \\
&\leq \sup_{\boldsymbol{w}\in\Theta} \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \left| \widehat{\nu}^*(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}}) - \nu(H_{\boldsymbol{u}^T\boldsymbol{w},\boldsymbol{u}}) \right| \\
&= \sup_{\boldsymbol{w}\in\Theta} \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \left| (\boldsymbol{u}^T\boldsymbol{w})^{-1/\widehat{\gamma}} \widehat{\nu}(H_{1,\boldsymbol{u}}) - (\boldsymbol{u}^T\boldsymbol{w})^{-1/\gamma} \nu(H_{1,\boldsymbol{u}}) \right| \\
&\leq \sup_{\boldsymbol{u}\in\Theta} \widehat{\nu}(H_{1,\boldsymbol{u}}) \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} (\boldsymbol{u}^T\boldsymbol{w})^{-1/\gamma} \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} \left| (\boldsymbol{u}^T\boldsymbol{w})^{1/\gamma-1/\widehat{\gamma}} - 1 \right| \\
&+ \sup_{\boldsymbol{u}^T\boldsymbol{w}\geq\delta} (\boldsymbol{u}^T\boldsymbol{w})^{-1/\gamma} \sup_{\boldsymbol{u}\in\Theta} |\widehat{\nu}(H_{1,\boldsymbol{u}}) - \nu(H_{1,\boldsymbol{u}})| \\
&\leq \delta^{-1/\gamma} \left| \delta^{1/\gamma-1/\widehat{\gamma}} - 1 \right| + \delta^{-1/\gamma} \sup_{\boldsymbol{u}\in\Theta} |\widehat{\nu}(H_{1,\boldsymbol{u}}) - \nu(H_{1,\boldsymbol{u}})|.
\end{aligned}$$

By the consistency of $\widehat{\gamma}$ and (8) we can conclude that (9) is true. □

Lemma 6. *As $n \to \infty$,*

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

Proof. Let $\varepsilon > 0$. Under Assumptions 1, 2 and 5, applying Chebyshev's inequality, for any $\delta > 0$ yields

$$\begin{aligned}
&\mathbb{P}\left(\left| \frac{n}{k} P_n\left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) - \frac{n}{k} P\left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) \right| \geq \delta \right) \\
&= \mathbb{P}\left(\left| P_n\left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) - P\left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) \right| \geq \frac{k}{n}\delta \right) \\
&\leq \frac{P\left(U(n/k)(1+\varepsilon)^2 S\right)\left[1 - P\left(U(n/k)(1+\varepsilon)^2 S\right)\right]}{n(k\delta/n)^2} \\
&= \frac{1}{k} \frac{P\left(U(n/k)(1+\varepsilon)^2 S\right)}{\mathbb{P}\left(\|\boldsymbol{X}\| \geq U(n/k)\right)} [1 - P\left(U(n/k)(1+\varepsilon)^2 S\right)] \frac{1}{\delta^2} \to 0
\end{aligned}$$

and therefore

$$\frac{n}{k} P_n\left(U\left(\frac{n}{k}\right)(1+\varepsilon)^2 S \right) \xrightarrow{\mathbb{P}} \nu((1+\varepsilon)^2 S) = (1+\varepsilon)^{-2/\gamma} \nu(S).$$

Similarly,

$$\frac{n}{k}P_n\left(U\left(\frac{n}{k}\right)(1-\varepsilon)^2 S\right) \xrightarrow{\mathbb{P}} \nu((1+\varepsilon)^2 S) = (1-\varepsilon)^{-2/\gamma}\nu(S).$$

Define events

$$\Omega_n = \{(1+\varepsilon)^2 U(n/k)S \subset \widehat{U}(n/k)\widehat{S} \subset (1-\varepsilon)^2 U(n/k)S\}$$

then $P(\Omega_n) \to 1$ because of $\widehat{U}(n/k)/U(n/k) \xrightarrow{P} 1$ under Assumption 2 and Proposition 5. On Ω_n, for large n,

$$\begin{aligned}\left|\widehat{\nu}(\widehat{S}) - \nu(S)\right| &\leq \left|\frac{n}{k}P_n(\widehat{U}(n/k)\widehat{S}) - \frac{n}{k}P(U(n/k)S)\right| + \left|\frac{n}{k}P(U(n/k)S) - \nu(S)\right| \\ &\leq \frac{n}{k}P((1-\varepsilon)^2 U(n/k)S) - \frac{n}{k}P((1+\varepsilon)^2 U(n/k)S) + \left|\frac{n}{k}P(U(n/k)S) - \nu(S)\right| \\ &\to [(1-\varepsilon)^{-2/\gamma} - (1+\varepsilon)^{-2/\gamma}]\nu(S)\end{aligned}$$

where ε can be chosen arbitrarily small. Hence $\widehat{\nu}(\widehat{S}) \xrightarrow{\mathbb{P}} \nu(S)$. □

Proof of Theorem 1. Define

$$\widehat{r}_n^{\boldsymbol{w}} := \frac{\widehat{U}(n/k)(k\widehat{\nu}(\widehat{S})/(np))^{\widehat{\gamma}}}{U(1/\beta)}(HD(\boldsymbol{w},\widehat{\nu}^*))^{\widehat{\gamma}}.$$

Note that, as $n \to \infty$, the continuity of U yields $\widehat{U}(n/k)/U(n/k) \xrightarrow{\mathbb{P}} 1$ while Lemma 6 and the consistency of $\widehat{\gamma}$ implies that $(k\widehat{\nu}(\widehat{S})/np)^{\widehat{\gamma}}/(k\nu(S)/np)^{\gamma} \xrightarrow{\mathbb{P}} 1$. Moreover, Assumption 3 gives that, as $n \to \infty$,

$$U(n/k)(k/n)^{\gamma} \xrightarrow{\mathbb{P}} c^{\gamma} \quad \text{and} \quad U(1/\beta)\beta^{\gamma} \xrightarrow{\mathbb{P}} c^{\gamma}.$$

Hence, by Lemma 5, it holds that

$$\frac{\widehat{U}(n/k)(k\widehat{\nu}(\widehat{S})/np)^{\widehat{\gamma}}}{U(1/\beta)} = \frac{\widehat{U}(n/k)(k\widehat{\nu}(\widehat{S})/np)^{\widehat{\gamma}}}{U(n/k)(k\nu(S)/np)^{\gamma}}\frac{U(n/k)(k/n)^{\gamma}}{U(1/\beta)\beta^{\gamma}}\left(\frac{\beta\nu(S)}{p}\right)^{\gamma} \xrightarrow{\mathbb{P}} 1 \cdot \frac{c^{\gamma}}{c^{\gamma}} \cdot 1^{\gamma} = 1.$$

Combining with Proposition 5 and writing $r^{\boldsymbol{w}} = (HD(\boldsymbol{w},\nu))^{\gamma}$, we obtain

$$\sup_{\boldsymbol{w}\in\Theta}\left|\frac{\widehat{r}_n^{\boldsymbol{w}}}{r^{\boldsymbol{w}}} - 1\right| \xrightarrow{\mathbb{P}} 0.$$

This implies that, for any $\varepsilon > 0$, the probability of the events $\Omega_n = \{(1-\varepsilon)r^{\boldsymbol{w}} \le \widehat{r}_n^{\boldsymbol{w}} \le (1+\varepsilon)r^{\boldsymbol{w}}, \text{ for all } \boldsymbol{w} \in \Theta\}$ converges to 1 as $n \to \infty$. Then, on Ω_n for large n,

$$\begin{aligned}
&\sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n} \left| \frac{HD(\boldsymbol{x},P)}{\beta} - 1 \right| \\
&= \sup_{\boldsymbol{w}\in\Theta} \left| \frac{HD(U(1/\beta)\widehat{r}_n^{\boldsymbol{w}}\boldsymbol{w},P)}{\mathbb{P}(\|\boldsymbol{X}\| \ge U(1/\beta))} - HD(HD(\boldsymbol{w},\nu)^{\gamma}\boldsymbol{w},\nu) \right| \\
&\le \sup_{\boldsymbol{w}\in\Theta} \left| \frac{HD(U(1/\beta)r^{\boldsymbol{w}}\boldsymbol{w},P)}{P(\|\boldsymbol{X}\| \ge U(1/\beta))} - HD(r^{\boldsymbol{w}}\boldsymbol{w},\nu) \right| \\
&\quad + \sup_{\boldsymbol{w}\in\Theta} \left| \frac{HD((1-\varepsilon)U(1/\beta)r^{\boldsymbol{w}}\boldsymbol{w},P)}{P(\|\boldsymbol{X}\| \ge U(1/\beta))} - \frac{HD((1+\varepsilon)U(1/\beta)r^{\boldsymbol{w}}\boldsymbol{w},P)}{P(\|\boldsymbol{X}\| \ge U(1/\beta))} \right| \\
&\quad =: \mathrm{I} + \mathrm{II}.
\end{aligned}$$

By Proposition 3, we know $\mathrm{I} \xrightarrow{\mathbb{P}} 0$ and

$$\mathrm{II} \xrightarrow{\mathbb{P}} [(1-\varepsilon)^{-1/\gamma} - (1+\varepsilon)^{-1/\gamma}]HD(r^{\boldsymbol{w}}\boldsymbol{w},\nu) = (1-\varepsilon)^{-1/\gamma} - (1+\varepsilon)^{-1/\gamma}.$$

Since ε can be arbitrarily small, it holds that

$$\sup_{\boldsymbol{x}\in\widehat{\mathcal{C}}_n} \left| \frac{HD(\boldsymbol{x},P)}{\beta} - 1 \right| \xrightarrow{\mathbb{P}} 0,$$

which immediately implies the first part of the theorem.

The proof of the second part is the same as that of Theorem 1 in Cai et al. (2011) using the propositions and lemmas above. □

References

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

Cai, J.-J., J. H. J. Einmahl, and L. de Haan (2011). Estimation of extreme risk regions under multivariate regular variation. *The Annals of Statistics 39*, 1803–1826.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: an Introduction.* New York: Springer.

de Haan, L. and S. Resnick (1987). On regular variation of probability densities. *Stochastic Processes and their Applications 25*, 83–93.

Dekkers, A. L. M., J. H. J. Einmahl, and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics 17*, 1833–1855.

Einmahl, J. H. J., L. de Haan, and A. Krajina (2013). Estimating extreme bivariate quantile regions. *Extremes 16*, 121–145.

Einmahl, J. H. J., J. Li, and R. Y. Liu (2014). Bridging centrality and extremity: refining empirical data depth using extreme value theory. *Preprint*.

Hallin, M., D. Paindaveine, and M. Siman (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *The Annals of Statistics 38*, 635–669.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails with Applications to Financial and Economic Time Series*. New York: Cambridge University Press.

Hill, B. M. (1975). A simple general approach to inference about the tail of distribution. *The Annals of Statistics 3*, 1163–1174.

Jessen, A. H. and T. Mikosch (2006). Regularly varying functions. *Publications de l'Institut Mathématique (Beograd) 80*, 171–192.

Kong, L. and I. Mizera (2012). Quantile tomography: Using quantiles with multivariate data. *Statistica Sinica 22*, 1580–1610.

McNeil, A. J. and A. D. Smith (2012). Multivariate stress scenarios and solvency. *Insurance: Mathematics and Economics 50*, 299–308.

Serfling, R. (2006). Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science 72*, 1–16.

Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *Journal of Nonparametric Statistics 22*, 915–936.

Serfling, R. and Y. Zuo (2010). Discussion of "Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth" by M. Hallin, D. Paindaveine and M. Šiman. *The Annals of Statistics 38*, 676–684.

Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics 15*, 1174–1207.

Zuo, Y. and R. Serfling (2000). General notions of statistical depth function. *The Annals of Statistics 28*, 461–482.