Pere, Jaakko; Ilmonen, Pauliina; Viitasaari, Lauri

On extreme quantile region estimation under heavy-tailed elliptical distributions

Published in:
Journal of Multivariate Analysis

DOI:
10.1016/j.jmva.2024.105314

Published: 01/07/2024

Document Version
Publisher's PDF, also known as Version of record

Journal of Multivariate Analysis 202 (2024) 105314

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On extreme quantile region estimation under heavy-tailed elliptical distributions

Jaakko Pere [a,*], Pauliina Ilmonen [a], Lauri Viitasaari [b]

[a] Aalto University School of Science, Finland
[b] Aalto University School of Business, Finland

ARTICLE INFO

AMS 2020 subject classifications:
primary 62G32
secondary 62H12

Keywords:
Elliptical distribution
Extreme quantile estimation
Heavy-tailed distribution
Multivariate quantile

ABSTRACT

Consider the estimation of an extreme quantile region corresponding to a very small probability. Estimation of extreme quantile regions is important but difficult since extreme regions contain only a few or no observations. In this article, we propose an affine equivariant extreme quantile region estimator for heavy-tailed elliptical distributions. The estimator is constructed by extending a well-known univariate extreme quantile estimator. Consistency of the estimator is proved under estimated location and scatter. The practicality of the developed estimator is illustrated with simulations and a real data example.

1. Introduction

There are many notions of multivariate quantile, see, for example, [1] for a review. After all, there is no natural ordering between multivariate observations, and thus, it is not straightforward to define the concept of multivariate quantile. However, assuming that an m-variate random variable $\boldsymbol{X}$ has density f, we can define a quantile region Q_p with p probability mass by

$$Q_p = \{\boldsymbol{x} \in \mathbb{R}^m : f(\boldsymbol{x}) \leq \beta\},$$

where β is chosen such that $\mathbb{P}(\boldsymbol{X} \in Q_p) = p$.

We consider the estimation of extreme quantile regions of the aforementioned form. By extreme quantile regions we mean Q_p corresponding to a very small p. For example, often in practice $p < 1/n$, where n denotes the sample size. Thus, in this problem setting it is possible that no observations lie in the extreme quantile region Q_p that we desire to estimate. Consequently, nonparametric approaches such as the plug-in estimators considered in [2] do not suffice. In order to extrapolate outside the sample we use the tools of extreme value theory (EVT).

Estimation of extreme quantile regions has been considered by various authors. For example, [3,4] considered the estimation of extreme quantile regions under multivariate regular variation. The former approach is based on modeling the underlying density and the latter one is based on halfspace depth contours and not the actual density contours. On the other hand, in [5], the estimation of bivariate extreme quantile regions was considered under the multivariate domain of attraction condition, which is a weaker condition than the multivariate regular variation. Namely, the multivariate domain of attraction condition allows for marginals with different extreme value indices. For reviews about the multivariate regular variation and multivariate domain of attraction conditions, we refer to [6,7].

In [3], several applications for the estimation of extreme risk regions are listed. One application is an alarm system or outlier detection. That is, an observation signals a risk or is classified as an outlier if it lies in the estimated extreme quantile region. Also,

* Corresponding author.
E-mail address: jaakko.pere@aalto.fi (J. Pere).

https://doi.org/10.1016/j.jmva.2024.105314
Received 5 May 2023; Received in revised form 18 March 2024; Accepted 18 March 2024
Available online 20 March 2024

the boundaries of extreme quantile regions can be used for stress testing. That is, one can test if extreme scenarios corresponding to the points from the boundary of a chosen Q_p would be sufficiently extreme to break the underlying system. Lastly, extreme quantile regions provide an ordering for outlying observations, see Remark 3 in [3]. The aforementioned applications can be useful in risk management. For examples in the context of insurance and aviation safety, see [5,8]. Examples in financial settings are presented in [3,4,9]. For an application in healthcare, see [10]. This article also includes an empirical illustration in the financial context using the same data set as in [4,9].

In this paper, we restrict to multivariate regularly varying elliptical distributions. Restriction to elliptical distributions means that the developed estimator cannot be applied as broadly as the estimators constructed in [3,4]. However, the ellipticity assumption allows for other useful properties.

Firstly, the constructed estimator is very easy to compute compared to, for example, the computationally demanding estimator provided in [4]. This is a consequence of the fact that, under ellipticity, estimation of the sets Q_p reduces to the estimation of location, scatter and univariate extreme quantile of a univariate random variable called the generating variate. Thus, from the viewpoint of EVT, only univariate theory suffices. This can be motivated by the fact that, in the case of elliptical distributions, multivariate regular variation is equivalent to the heavy-tailedness of the generating variate [11]. A tool that turns out to be crucial in the construction of the estimator is the Mahalanobis distance. That is, in order to apply univariate EVT we have to approximate the generating variate from the available sample, and this can be achieved with the Mahalanobis distance.

Secondly, our estimator is affine equivariant. On the contrary, the more general estimator given in [4] is only equivariant up to scaling and orthogonal transformations. Essentially, the affine equivariance of our estimator follows from the affine invariance of the Mahalanobis distance. As a building block, we use the Mahalanobis distance based separating Hill estimator, which is an affine invariant extreme value index estimator tailored for elliptical distributions. Note that generally, estimators of the extreme value index are not affine invariant. For a review of the separating Hill estimator, see [12] (under known location–scatter pair) and [13] (under estimated location–scatter pair). Affine equivariance is an important property for practical applications. For example, we desire that the outlyingness of a data point does not depend on the coordinate system or on the units of marginal random variables, which holds if the estimator of Q_p is affine equivariant.

It is important to note that there exist other notions of multivariate extreme quantiles than the one based on density or depth contours. For example, [10,14] considered the estimation of extreme quantiles based on the notion of geometric quantiles defined in [15]. Geometric quantiles are defined as a solution to a generalization of an optimization problem that characterizes univariate quantiles. Quantiles based on density or depth contours are sets indexed with probabilities p, however, geometric quantiles are points in $\mathbb{R}^m$ indexed by vectors $\boldsymbol{u}$ on an m-dimensional unit ball. That is, geometric quantiles have both direction and magnitude. In the context of [10,14] extreme quantiles are geometric quantiles with $\|\boldsymbol{u}\|$ close to one. We, however, consider Q_p for a small p. Moreover, [9] developed an estimator for extreme directional multivariate quantiles. Directional multivariate quantiles were introduced in [16,17]. This approach is based on oriented orthants. That is, [9] define oriented orthant with vertex $\boldsymbol{x} \in \mathbb{R}^m$ and direction $\boldsymbol{v}$, where $\boldsymbol{v}$ is a vector on a $(m-1)$-sphere. Then quantile indexed by direction $\boldsymbol{v}$ and probability p is the boundary of the set of vertices $\boldsymbol{x}$ for which the corresponding oriented orthants at direction $-\boldsymbol{v}$ include at least $1-p$ probability mass. For example, see Figures 4 and 11 on [9] for illustrations of directional quantiles. In the context of [9], extreme quantiles at direction $\boldsymbol{v}$ are directional quantiles corresponding to a small p.

The rest of the article is organized as follows. In Section 2, we review some selected results in univariate extreme value theory, focusing on univariate extreme quantile estimation. In Section 3 we give the necessary preliminaries about elliptical distributions, construct the extreme quantile estimator and give the consistency and affine equivariance results for the developed estimator. Section 4 provides a simulation study that assesses finite sample properties of the developed estimator, as well as compares it to a competitor. A real data example in a financial context is provided in Section 5. Lastly, the proofs of the results are given in the Appendix.

2. Univariate extreme quantile estimation

Throughout this section, let $X_1, \dots, X_n$ be i.i.d. univariate random variables with the cumulative distribution function F. We denote the corresponding order statistics by $X_{1,n} \le \cdots \le X_{n,n}$. We define the tail quantile function U corresponding to a distribution F by

$$U(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right), \quad t > 1,$$

where the left-continuous inverse $f^{\leftarrow}$ of a nondecreasing function f is given by $f^{\leftarrow}(y) = \inf\{x : f(x) \ge y\}$. The tail function of a distribution is denoted by $\bar{F} = 1 - F$. If we want to stress that the distribution F corresponds to some specific random variable X, we use the notation $F = F_X$. A similar convention is used for the tail quantile function $U = U_X$, the tail function $\bar{F} = \bar{F}_X$, and the density $f = f_X$ if it exists. Let Y_n and Z_n be sequences of random variables. The notation $Y_n = o_{\mathbb{P}}(Z_n)$ means that Y_n/Z_n converges to 0 in probability as $n \to \infty$. Similarly, we denote $Y_n = O_{\mathbb{P}}(Z_n)$, if Y_n/Z_n is bounded in probability.

In the univariate case, the problem of extreme quantile estimation can be formulated as follows. We want to estimate the $(1-p)$-quantile $x_p = U(1/p)$ for a very small p. The probability p can be so small that, for example, the quantile x_p is on the right-hand side of all the observations. In this case, the left-continuous inverse of the empirical distribution at point $1-p$ is an insufficient estimator and one has to rely on extreme value theory (EVT). Analysis of extreme events usually requires some regularity conditions.

Definition 1 (*Domain of Attraction*). We say that F is in the domain of attraction of a nondegenerate distribution G if there exists sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x),$$

for each continuity point x of G, or equivalently, we have the weak convergence of the normalized sample maxima,

$$\frac{\max(X_1, \dots, X_n) - b_n}{a_n} \xrightarrow{d} G, \quad n \to \infty,$$

where $X_1, \dots, X_n$ are i.i.d. with common distribution F. We denote $F \in \mathcal{D}(G)$ if distribution F is in the domain of attraction of G. The limiting distribution G is called the extreme value distribution.

It has been shown by [18,19] that the limit distribution G has a certain form. Namely, the location-scale family of G is characterized by one parameter γ that is called the extreme value index.

Theorem 1 (*[6], Theorem 1.1.3*). *Let F be a cumulative distribution function and assume that $F \in \mathcal{D}(G)$. Then there exists $\gamma \in \mathbb{R}$ such that*

$$G(x) = G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right),$$

where $1 + \gamma x > 0$. For the case $\gamma = 0$, the right-hand side is interpreted as $\exp(-e^{-x})$.

It is typical to divide the distributions $F \in \mathcal{D}(G_\gamma)$ in three classes, depending on the value of the extreme value index γ:

(i) If $\gamma > 0$, then G_γ is said to be a Fréchet distribution and $F \in \mathcal{D}(G_\gamma)$ is called heavy-tailed;
(ii) If $\gamma = 0$, then G_γ is said to be a Gumbel distribution and $F \in \mathcal{D}(G_\gamma)$ is called light-tailed;
(iii) If $\gamma < 0$, then G_γ is said to be a Weibull distribution and $F \in \mathcal{D}(G_\gamma)$ is called short-tailed.

There are many equivalent characterizations of the domain of attraction condition, see [6] for a review. In Theorem 2 we state one of these characterizations. We start by giving the definitions of extended regular variation and regular variation.

Definition 2 (*Extended Regular Variation*). The tail quantile function U is said to be of extended regular variation if there exists a positive function a such that for some $\gamma \in \mathbb{R}$ and for all $x > 0$,

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma},$$

where for $\gamma = 0$ the right-hand side is interpreted as $\ln x$.

Definition 3 (*Regular Variation*). The tail quantile function U is regularly varying if for some $\gamma \in \mathbb{R}$ and for all $x > 0$,

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma.$$

Theorem 2 (*[6], Theorem 1.1.6, Corollary 1.2.10*). *Let $\gamma \in \mathbb{R}$. Then $F \in \mathcal{D}(G_\gamma)$ if and only if U is of extended regular variation. Furthermore, for the case $\gamma > 0$, we have that U is of extended regular variation if and only if U is regularly varying.*

Let $\gamma > 0$ and set $tx = y$. Then heuristically, Theorem 2 states that for large t, quantile $U(y)$ is related to a smaller quantile $U(t)$ in the following way

$$U(y) \approx U(t)\left(\frac{y}{t}\right)^\gamma.$$

By setting $t = n/k$, we can further approximate $U(n/k) \approx X_{n-k,n}$. This construction motivates to define an extreme quantile estimator

$$\hat{x}_p = X_{n-k,n}\left(\frac{k}{np}\right)^{\hat{\gamma}_n}, \quad k \in \{1, \dots, n-1\}, \tag{1}$$

where $\hat{\gamma}_n$ is an estimator of the extreme value index γ. For example, one can choose to use the Hill estimator

$$\hat{\gamma}_n = \frac{1}{k}\sum_{i=0}^{k-1} \ln X_{n-i,n} - \ln X_{n-k,n},$$

first introduced in [20]. The Hill estimator is a consistent estimator of γ under i.i.d. observations if the underlying distribution is heavy-tailed and $k = k_n \to \infty$, $k_n/n \to 0$, as $n \to \infty$ [21]. The estimator is also asymptotically normal under certain additional conditions. Asymptotic normality of the Hill estimator has been studied by various authors, see for example [22,23].

For developing asymptotic results in EVT, such as limiting distributions, the domain of attraction condition is often not sufficient. Some second order refinement is necessary. For example, the asymptotic normality of the Hill estimator is often presented under the second order regular variation given in Definition 6, see for example [24]. Also, second order conditions are needed for the asymptotic normality of the univariate extreme quantile estimator and the main result of this article, Theorem 5. Further discussion about second order conditions is given in the Appendix.

Next, let us review a result about the asymptotic normality of the univariate extreme quantile estimator presented in (1).

Theorem 3 (*[6], Theorem 4.3.8*)**.** *Let $X_1, \dots, X_n$ be i.i.d. with distribution F. Let $F \in \mathcal{D}(G_\gamma)$, $\gamma > 0$ and let $\hat{\gamma}_n$ be a corresponding extreme value index estimator based on a threshold sequence k_n. Let x_{p_n} denote the $(1-p_n)$-quantile of F and let*

$$\hat{x}_{p_n} = X_{n-k_n,n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n}.$$

Assume that the following conditions hold:

- *C1.* $k_n \to \infty$, $k_n/n \to 0$, *as* $n \to \infty$;
- *C2.* *U satisfies the second order regular variation condition given in Definition 6;*
- *C3.* $\lim_{n\to\infty} \sqrt{k_n}A(n/k_n) = \lambda \in \mathbb{R}$, *where A is the positive or negative function for U in Definition 6;*
- *C4.* $np_n = o(k_n)$ *and* $\ln(np_n) = o\left(\sqrt{k_n}\right)$, *as* $n \to \infty$;
- *C5.* $\sqrt{k_n}\left(\hat{\gamma}_n - \gamma\right) \xrightarrow{d} \Gamma$, *as $n \to \infty$, where Γ is normally distributed with known expected value possibly depending on γ and ρ and known variance depending on γ (but not on ρ).*

Then, as $n \to \infty$, with $d_n = k_n/np_n$,

$$\frac{\sqrt{k}}{\ln d_n}\left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1\right) \xrightarrow{d} \Gamma.$$

Theorem 3 is applied in proving the convergence of the extreme region estimator proposed in this article, see the Appendix.

Notice that on the first part of Condition C4 of Theorem 3 we require that $p_n \to 0$ fast, as $n \to \infty$. This is a typical requirement in asymptotic results for extreme quantile estimators. If the probability p does not depend on n, then for a large enough sample the quantile x_p would be no longer extreme. However, the probability $p = p_n$ cannot decay to zero arbitrarily fast. Intuitively, the second part of Condition C4 states that p_n has to converge to zero slowly enough compared to the convergence rate of $\hat{\gamma}_n$. For an example, see (13) in the Appendix.

Extreme quantile estimation has been considered also for short-tailed distributions. Note, however, that the short-tailed case requires estimation of a scale function, see [6] for a review. Note also that, by construction, the Hill estimator is always positive, and thus, it is not a suitable estimator for negative extreme value indices. Nevertheless, negative extreme value indices can be estimated using, e.g., the Pickands estimator [25] or the moment estimator [26].

3. Main results

In this section, we propose a new approach for estimating extreme density contours of heavy-tailed elliptical distributions, but first, elliptical distributions and estimation of the location and scatter are reviewed in Section 3.1. Then the estimator is constructed in Section 3.2, and finally, consistency and affine equivariance results are given in Section 3.3.

3.1. Elliptical distributions and estimation of the location–scatter pair

From the perspective of EVT, elliptical distributions form a flexible family of multivariate distributions. For example, the family of elliptical distributions includes multivariate normal distribution that has light tails, but on the other hand, t-distributions are also elliptical distributions but have heavier tails than the normal distribution. Below we define an elliptically distributed random variable.

Definition 4. Random variable $\boldsymbol{X} : \Omega \to \mathbb{R}^m$ is said to be elliptically distributed if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^m$, a positive semidefinite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times m}$ and a function $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}$ such that the characteristic function $\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}$ of $\boldsymbol{X} - \boldsymbol{\mu}$ is of the form $\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}(\boldsymbol{t}) = \phi(\boldsymbol{t}^\intercal \boldsymbol{\Sigma} \boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^m$. We write $\boldsymbol{X} \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ if $\boldsymbol{X}$ is elliptically distributed with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and ϕ.

We work with elliptically distributed random variables with a full rank $\boldsymbol{\Sigma}$. Then we can apply a convenient stochastic representation for elliptically distributed random variables.

Theorem 4 (*[27], Theorem 1*)**.** $\boldsymbol{X} \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *with* $\operatorname{rank}(\boldsymbol{\Sigma}) = k$, $k \leq m$ *if and only if*

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\boldsymbol{S}, \tag{2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^m$, $\mathcal{R} : \Omega \to \mathbb{R}_{\geq 0}$ is a nonnegative random variable, $\boldsymbol{S}$ is uniformly distributed over the unit-sphere $\{\boldsymbol{x} \in \mathbb{R}^k : \boldsymbol{x}^\intercal \boldsymbol{x} = 1\}$, $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times k}$ is a matrix with $\operatorname{rank}(\boldsymbol{\Lambda}) = k$ such that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\intercal$ and random variables $\mathcal{R}$ and $\boldsymbol{S}$ are independent. We call the random variable $\mathcal{R}$ the generating variate, matrix $\boldsymbol{\Sigma}$ the scatter matrix and vector $\boldsymbol{\mu}$ the location vector of the distribution.

The above theorem shows that it is natural to work with elliptically distributed random variables with a full rank scatter matrix $\boldsymbol{\Sigma}$ since otherwise $\boldsymbol{X}$ would live in some linear subspace of $\mathbb{R}^m$.

The generating variate $\mathcal{R}$ plays an important role in the theory of elliptical distributions. In [11] it was proven that elliptically distributed random variable $\boldsymbol{X}$ satisfies multivariate regular variation condition with $\gamma > 0$ if and only if the corresponding generating variate $\mathcal{R}$ is regularly varying with the same index γ. Thus, from the viewpoint of Theorem 2, univariate random variable $\mathcal{R}$ gives all the information about the tail behavior of an elliptically distributed random variable $\boldsymbol{X}$. Notice that the scatter matrix $\boldsymbol{\Sigma}$ and the generating variate $\mathcal{R}$ are only unique up to a positive constant. In order to guarantee uniqueness one could require, for example, that $\det(\boldsymbol{\Sigma}) = 1$, see [28].

Define a norm induced by a symmetric positive definite matrix $\boldsymbol{H} \in \mathbb{R}^{m\times m}$ by

$$\|\boldsymbol{x}\|_{\boldsymbol{H}} = \sqrt{\boldsymbol{x}^\intercal \boldsymbol{H}^{-1}\boldsymbol{x}}, \quad \boldsymbol{x} \in \mathbb{R}^m.$$

Throughout the rest of the article, let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d. copies of an m-variate elliptically distributed random variable $\boldsymbol{X}$ with a full rank scatter matrix Σ and density $f_{\boldsymbol{X}}$. In addition, we assume that the density $f_{\boldsymbol{X}}$ is eventually decreasing, that is, for some $r > 0$ and for all $\boldsymbol{x}, \boldsymbol{y} \in S_r = \{\boldsymbol{z} \in \mathbb{R}^m : \|\boldsymbol{z} - \boldsymbol{\mu}\|_{\Sigma} > r\}$ we have

$$\|\boldsymbol{x} - \boldsymbol{\mu}\|_{\Sigma} < \|\boldsymbol{y} - \boldsymbol{\mu}\|_{\Sigma} \Rightarrow f_{\boldsymbol{X}}(\boldsymbol{x}) > f_{\boldsymbol{X}}(\boldsymbol{y}), \tag{3}$$

and in the set S_r^c we have

$$\inf_{\boldsymbol{x}\in S_r^c} f_{\boldsymbol{X}}(\boldsymbol{x}) > 0. \tag{4}$$

Assumption (3) gives a natural definition for eventually decreasing density for elliptical distributions since $f_{\boldsymbol{X}}$ is constant on ellipsoids $\{\boldsymbol{z} \in \mathbb{R}^m : \|\boldsymbol{z} - \boldsymbol{\mu}\|_{\Sigma} = r\}$ for $r > 0$, see [29, Corollary 4] for details. Together Assumptions (3) and (4) guarantee that extreme quantile regions Q_p are connected for a sufficiently small p, see Section 3.2 for details. Let P be the law of $\boldsymbol{X}$, that is $P(\cdot) = \mathbb{P}(\boldsymbol{X} \in \cdot)$, where $\mathbb{P}$ is the probability measure corresponding to the underlying probability space. Assume that $\hat{\boldsymbol{\mu}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ are estimators of the location $\boldsymbol{\mu}$ and the scatter $\boldsymbol{\Sigma}$, respectively, calculated from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. Then we can denote the Mahalanobis distance between the elliptically distributed random variable $\boldsymbol{X}$, or $\boldsymbol{X}_i$, and its location $\boldsymbol{\mu}$ by

$$R = \|\boldsymbol{X} - \boldsymbol{\mu}\|_{\Sigma}, \;\; R_i = \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_{\Sigma},$$

and the corresponding versions under estimated location and scatter by

$$\hat{R} = \|\boldsymbol{X} - \hat{\boldsymbol{\mu}}_n\|_{\hat{\Sigma}_n}, \;\; \hat{R}_i = \|\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_n\|_{\hat{\Sigma}_n}.$$

Random variables $\hat{R}_i$ will turn out to be crucial in the construction of our extreme quantile region estimator. Consequently, estimators of the location and scatter that still perform in the heavy-tailed setup are required. More precisely, in the consistency result of Theorem 5 we require $\sqrt{n}$-consistent estimators for location and scatter under heavy-tailed generating variate. The literature on the topic of estimation of the location and scatter of elliptical distribution is rich, see [28] and the references therein.

Also note that, for elliptical distributions, all affine equivariant finite population location vectors measure the same population quantity. Moreover, for elliptical distributions, all affine equivariant finite population scatter matrices are proportional to each other. Thus, the corresponding affine equivariant scatter estimators estimate the same population quantity up to multiplication by a positive univariate constant. As our estimation procedure involves Mahalanobis distances with respect to the chosen scatter matrix, the constant vanishes and one can apply any affine equivariant scatter matrix estimator.

We choose to use the minimum covariance determinant (MCD) estimator for the location and scatter [30] in the simulations and the empirical example in Sections 4 and 5, respectively. MCD is based on dividing the whole sample into subsamples, consisting of $0.5 \le \alpha \le 1$ portion of all observations. Then, the subsample for which the computed sample covariance has the smallest determinant is used in estimation. That is, the sample mean and the sample covariance computed from the chosen sample are the estimates of location and scatter, respectively. The parameter α has an effect on the estimation. Namely, as α decreases, MCD estimators become more robust. However, for many models such as multivariate normal distribution and t-distribution with several different degrees of freedom, as α increases, at least to a certain point, MCD estimators become more efficient [31]. In the simulations and the empirical example we choose $\alpha = 0.5$. This choice is motivated by the fact that we work in the heavy-tailed setting. Thus, we know that outlying observations are present, and consequently, we wish to achieve maximal robustness even if the efficiency of the location and scatter estimators suffer. In practice, it is not feasible to compute determinants of sample covariance for each subsample, but there are fast algorithms for approximation of MCD estimates, for example, see [32]. Consistency of the MCD location estimator and the scatter estimator, as well as the asymptotic normality of the former, are considered in [33]. In [34,35] consistency and asymptotic normality of the MCD scatter estimator were studied under more general distributions. For a recent review about the MCD, see [36].

3.2. Construction of the estimator

We next consider the multivariate extreme quantile region estimation under ellipticity. Recall the definition of quantile region given in Section 1. That is, assuming m-variate random variable $\boldsymbol{X}$ has density $f_{\boldsymbol{X}}$, we define quantile region Q_p as

$$Q_p = \{\boldsymbol{x} \in \mathbb{R}^m : f(\boldsymbol{x}) \le \beta\}, \tag{5}$$

where β is chosen such that $P(Q_p) = p$. However, in the case of elliptical distributions we have

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \underbrace{C_{m,\Sigma}\, \|\boldsymbol{x} - \boldsymbol{\mu}\|_{\Sigma}^{1-m} f_{\mathcal{R}}(\|\boldsymbol{x} - \boldsymbol{\mu}\|_{\Sigma})}_{=g(\|\boldsymbol{x}-\boldsymbol{\mu}\|_{\Sigma})}, \quad \boldsymbol{x} \neq \boldsymbol{\mu},$$

where $C_{m,\Sigma} > 0$ is a positive constant that depends on the dimension m and the scatter matrix Σ [29, Corollary 4]. Recall that for some $r > 0$, f_X is decreasing in the set $S_r = \{z \in \mathbb{R}^m : \|z - \mu\|_\Sigma > r\}$ in the sense of Assumption (3). Then it follows that the restriction of g in $\{t \in \mathbb{R} : t > r\}$, denoted by $\tilde{g}$, is decreasing and bijective. Moreover, for a sufficiently small p, that is by choosing β corresponding to p such that $0 < \beta < \inf_{x \in S_r^c} g(\|x - \mu\|_\Sigma)$, we have $Q_p \subset S_r$ by Assumption (4). Then we can express Q_p as

$$Q_p = \left\{ x \in \mathbb{R}^m : g(\|x - \mu\|_\Sigma) \le \beta \right\} = \left\{ x \in \mathbb{R}^m : \|x - \mu\|_\Sigma \ge \tilde{g}^{-1}(\beta) \right\}.$$

Thus, for a sufficiently small p, extreme quantile regions can be characterized as

$$Q_p = \left\{ x \in \mathbb{R}^m : \|x - \mu\|_\Sigma \ge r_p \right\}, \tag{6}$$

where r_p is chosen such that $P(Q_p) = p$. Since random variable X is absolutely continuous we can choose r_p uniquely. Actually, from the definition of elliptical distribution it follows that $R \overset{d}{=} \mathcal{R}$. Indeed, r_p is the $(1-p)$-quantile of the generating variate $\mathcal{R}$ as

$$P(Q_p) = \mathbb{P}(R \ge r_p) = 1 - F_{\mathcal{R}}(r_p) = p,$$

leading to

$$r_p = F_{\mathcal{R}}^{-1}(1-p) = U_{\mathcal{R}}(1/p).$$

Under ellipticity, the characterization given by Eq. (6) is more natural than the definition given by (5). Firstly, r_p can be easily computed explicitly if p corresponding to Q_p is given. Secondly, and more importantly, r_p corresponding to a very small p can be estimated with the tools of univariate extreme value theory. On the contrary, [3,4] relied on multivariate regular variation in the estimation of β, which leads to a more general but also more complicated estimator than ours.

Now, our task is to estimate Q_p for a very small p. We assume that p is so small that for Q_p we can use the representation given by Eq. (6). Ideally, we would like to estimate r_p with extreme quantile estimator (1) given the sample $R_1, \ldots, R_2$. However, estimation is complicated by the fact that, in practice, we do not know the true location or scatter but only have the multivariate sample $X_1, \ldots, X_n$. Thus, we estimate the true location μ and scatter Σ with estimators $\hat{\mu}_n$ and $\hat{\Sigma}_n$, respectively, and use approximations $\hat{R}_1, \ldots, \hat{R}_n$ in the estimation of the $(1-p)$-quantile of the generating variate,

$$\hat{r}_p = \hat{R}_{n-k_n,n} \left(\frac{k_n}{np} \right)^{\hat{\gamma}_n},$$

where

$$\hat{\gamma}_n = \frac{1}{k_n} \sum_{i=0}^{k-1} \ln \left(\frac{\hat{R}_{n-i,n}}{\hat{R}_{n-k,n}} \right).$$

The estimator $\hat{\gamma}_n$ is called the separating Hill estimator. Consistency and asymptotic normality of the estimator $\hat{\gamma}_n$ are proved in [13]. Utilizing the separating Hill estimator, we define an estimator for the extreme quantile region Q_p as

$$\hat{Q}_p = \left\{ x \in \mathbb{R}^m : \|x - \hat{\mu}_n\|_{\hat{\Sigma}_n} \ge \hat{r}_p \right\}. \tag{7}$$

The estimator $\hat{Q}_p$ consists of three components: the estimator of the location $\hat{\mu}_n$, the estimator of the scatter matrix $\hat{\Sigma}_n$ and the extreme quantile estimator of the generating variate $\hat{r}_p$. Estimated quantile region $\hat{Q}_p$ is a complement of an open ellipsoid at location $\hat{\mu}_n$. The matrix $\hat{\Sigma}_n$ gives the shape of the ellipsoid, $\hat{\mu}_n$ gives the center of the ellipsoid, and $\hat{r}_p$ gives the scale of the ellipsoid. If $\hat{\Sigma}_n$ is proportional to the identity matrix, the ellipsoid is spherical. For heavy-tailed distributions $\hat{r}_p$ is large. Consequently, the heavier the tail, the larger the ellipsoid (assuming that the scatter is fixed).

3.3. Consistency and affine equivariance

Denote symmetric difference between two sets $S \subset \mathbb{R}^m$ and $T \subset \mathbb{R}^m$ by $S \triangle T = (S \setminus T) \cup (T \setminus S)$. Since the extreme quantile region Q_p and the corresponding estimate are no longer scalars but subsets of $\mathbb{R}^m$, it is not straightforward to formulate the consistency result for the estimator $\hat{Q}_p$. However, in many applications where some multivariate region is estimated, such as in multivariate density estimation [2] and estimation of depth contours [37, Lemma 2], it is useful to formulate consistency in terms of probability mass on the symmetric difference. One motivation for this is that the map $P(S \triangle T) : \mathcal{B} \times \mathcal{B} \mapsto [0,1]$ is a pseudometric in probability space $(\mathbb{R}^m, \mathcal{B}, P)$ where $\mathcal{B}$ is the Borel sigma-algebra on $\mathbb{R}^m$.

In the context of multivariate extreme quantile estimation we require that $p_n \to 0$ fast, as $n \to \infty$. Then both the extreme quantile region $Q_p = Q_{p_n}$ and the estimated region $\hat{Q}_p = \hat{Q}_{p_n}$ shrink as $n \to \infty$. Hence

$$P(Q_{p_n} \triangle \hat{Q}_{p_n}) \overset{\mathbb{P}}{\to} 0, \quad n \to \infty,$$

is not a sufficient requirement for consistency. Otherwise, we could choose, for example, $\hat{Q}_{p_n} = \emptyset$. Instead, we must require that $P(Q_{p_n} \triangle \hat{Q}_{p_n})$ converges to zero in probability fast as $n \to \infty$. More precisely, we require that

$$\frac{P(Q_{p_n} \triangle \hat{Q}_{p_n})}{p_n} \overset{\mathbb{P}}{\to} 0, \quad n \to \infty.$$

Above notion of refined consistency is typical in the literature of extreme quantile region estimation, see [3,4,8]. Also, the fact $p_n \to 0$ as $n \to \infty$ guarantees that eventually Q_{p_n} are connected and have the representation given by (6).

We are now ready to consider the consistency of the elliptical extreme quantile region estimator $\hat{Q}_p$.

Theorem 5. *Let $\boldsymbol{X}$ be an m-variate absolutely continuous elliptically distributed random variable with generating variate $\mathcal{R}$, location vector $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$. Additionally, assume that the density $f_{\boldsymbol{X}}$ of $\boldsymbol{X}$ satisfies Assumptions (3) and (4). Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d. copies of $\boldsymbol{X}$. Let $F_{\mathcal{R}} \in \mathcal{D}(G_\gamma)$, $\gamma > 0$, and let $\hat{\gamma}_n$ denote the corresponding separating Hill estimator. Let Q_{p_n} denote the elliptical extreme $(1-p_n)$-quantile region and let $\hat{Q}_{p_n}$ be the corresponding estimator given in Eq. (7). Estimators $\hat{\gamma}_n$ and $\hat{Q}_{p_n}$ are based on a threshold sequence k_n, location estimator $\hat{\boldsymbol{\mu}}_n$ and scatter estimator $\hat{\boldsymbol{\Sigma}}_n$. Assume that the following conditions hold:*

- *C1.* $k_n \to \infty$, $k_n/n \to 0$, *as* $n \to \infty$;
- *C2.* $U_{\mathcal{R}}$ *satisfies the second order extended regular variation condition given in Definition 5;*
- *C3.* $\lim_{n\to\infty} \sqrt{k_n} A(n/k_n) = \lambda \in \mathbb{R}$, *where A is the positive or negative function for $U_{\mathcal{R}}$ in Definition 6;*
- *C4.* $np_n = o(k_n)$, $\ln\left(np_n\right) = o\left(\sqrt{k_n}\right)$ *and* $1/p_n = O\left(n^{1/(2\gamma)}\right)$, *as* $n \to \infty$;
- *C5.* $\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}) = O_{\mathbb{P}}(1)$ *and* $\sqrt{n}(\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}) = O_{\mathbb{P}}(1)$.

Then as $n \to \infty$,

$$\frac{P(\hat{Q}_{p_n} \triangle Q_{p_n})}{p_n} \xrightarrow{\mathbb{P}} 0.$$

By comparing Theorem 5 to Theorem 3, there is an additional part in Condition C4. In Section 2 we already discussed that the second part of Condition C4 is related to the rate of convergence of the extreme quantile estimator. Similarly, the third part is related to the rate of convergence of the scatter estimator $\hat{\boldsymbol{\Sigma}}_n$, see the proof of Theorem 5 in the Appendix for details. However, the third part of Condition C4 is not too stringent. That is, we can still find sequences p_n that satisfy Condition C4. For example, we can set

$$p_n = n^{-\frac{1}{2\gamma}} \quad \text{and}$$

$$k_n = \left\lceil n^\beta \right\rceil, \text{ where } \max\left\{0, 1 - \frac{1}{2\gamma}\right\} < \beta < 1.$$

Also, note that $\sqrt{n}$-consistent estimators for the location and scatter exist even in the heavy-tailed framework, see Section 3.1 for a detailed discussion.

We next consider affine equivariance of our proposed extreme quantile region estimator. Let $\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{X} + \boldsymbol{b}$ where $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{B} \in \mathbb{R}^{m\times m}$ is invertible. Now from the representation (2) it follows that

$$\boldsymbol{Y} \overset{d}{=} (\boldsymbol{B}\boldsymbol{\mu} + \boldsymbol{b}) + \mathcal{R}\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{S}.$$

Thus the random variable $\boldsymbol{Y}$ is also elliptically distributed with the same generating variate as $\boldsymbol{X}$, location $\boldsymbol{\mu}' = \boldsymbol{B}\boldsymbol{\mu} + \boldsymbol{b}$ and scatter matrix $\boldsymbol{\Sigma}' = \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\intercal$. We now have

$$\|\boldsymbol{Y} - \boldsymbol{\mu}'\|_{\boldsymbol{\Sigma}'} = \|\boldsymbol{X} - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}},$$

by affine invariance of the Mahalanobis distance, see Lemma 7. Hence the assumption of ellipticity guarantees that the constructed estimator $\hat{Q}_n$ is affine equivariant as long as the location estimator $\hat{\boldsymbol{\mu}}_n$ and the scatter estimator $\hat{\boldsymbol{\Sigma}}_n$ are affine equivariant.

Theorem 6. *Let $\boldsymbol{B} \in \mathbb{R}^{m\times m}$ be invertible and let $\boldsymbol{b} \in \mathbb{R}^m$. Let $\mathcal{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ denote a sample of an elliptically distributed random variable $\boldsymbol{X}$. Let $\boldsymbol{Y}_i = \boldsymbol{B}\boldsymbol{X}_i + \boldsymbol{b}$ and $\mathcal{Y} = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$. Let $\hat{\boldsymbol{\mu}}_n(\mathcal{X})$, $\hat{\boldsymbol{\Sigma}}_n(\mathcal{X})$ and $\hat{\boldsymbol{\mu}}_n(\mathcal{Y})$, $\hat{\boldsymbol{\Sigma}}_n(\mathcal{Y})$ be estimators of the location and scatter calculated from $\mathcal{X}$ and $\mathcal{Y}$, respectively. Assume that $\hat{\boldsymbol{\mu}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ are affine equivariant in the following sense:*

$$\hat{\boldsymbol{\mu}}_n(\mathcal{Y}) = \boldsymbol{B}\hat{\boldsymbol{\mu}}_n(\mathcal{X}) + \boldsymbol{b}, \quad \hat{\boldsymbol{\Sigma}}_n(\mathcal{Y}) = \boldsymbol{B}\hat{\boldsymbol{\Sigma}}_n(\mathcal{X})\boldsymbol{B}^\intercal.$$

Denote

$$\hat{R}_i^x = \|\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_n(\mathcal{X})\|_{\hat{\boldsymbol{\Sigma}}_n(\mathcal{X})}, \quad \hat{R}_i^y = \|\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_n(\mathcal{Y})\|_{\hat{\boldsymbol{\Sigma}}_n(\mathcal{Y})},$$

$$\hat{\gamma}_n(\mathcal{X}) = \frac{1}{k_n}\sum_{i=0}^{k-1} \ln\left(\frac{\hat{R}^x_{n-i,n}}{\hat{R}^x_{n-k,n}}\right), \quad \hat{\gamma}_n(\mathcal{Y}) = \frac{1}{k_n}\sum_{i=0}^{k-1} \ln\left(\frac{\hat{R}^y_{n-i,n}}{\hat{R}^y_{n-k,n}}\right),$$

$$\hat{r}^x_{p_n} = \hat{R}^x_{n-k_n,n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n(\mathcal{X})}, \quad \hat{r}^y_{p_n} = \hat{R}^y_{n-k_n,n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n(\mathcal{Y})}.$$

Let

$$\hat{Q}^x_{p_n} = \left\{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_n(\mathcal{X})\|_{\hat{\boldsymbol{\Sigma}}_n(\mathcal{X})} \geq \hat{r}^x_{p_n}\right\}, \quad \hat{Q}^y_{p_n} = \left\{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_n(\mathcal{Y})\|_{\hat{\boldsymbol{\Sigma}}_n(\mathcal{Y})} \geq \hat{r}^y_{p_n}\right\}.$$

Now

$$\hat{Q}^y_{p_n} = \boldsymbol{B}\hat{Q}^x_{p_n} + \boldsymbol{b},$$

where for all $S \subset \mathbb{R}^m$, $\boldsymbol{B}S + \boldsymbol{b}$ denotes the set $\{\boldsymbol{B}\boldsymbol{x} + \boldsymbol{b} : \boldsymbol{x} \in S\}$.

Table 1
Simulation results for the elliptical extreme quantile region estimator $\hat{Q}_{p_n}$ for the two-dimensional settings. The table contains the medians calculated from $s = 100$ approximated relative errors $P(\hat{Q}_{p_n} \triangle Q_{p_n})/p_n$ for all combinations of distributions (centered spherical Cauchy distribution, centered elliptical Cauchy distribution and centered spherical t-distribution with 4 degrees of freedom), $n \in \{1000, 5000\}$, $p_n \in \{2/n, 1/n, 1/(2n)\}$ and $k_n \in \{0.05n, 0.1n, 0.2n\}$. In all scenarios location and scatter are estimated with MCD for $\alpha = 0.5$.

Distribution	Value of p_n	$n = 1000$			$n = 5000$		
		$k_n = 0.05n$	$k_n = 0.1n$	$k_n = 0.2n$	$k_n = 0.05n$	$k_n = 0.1n$	$k_n = 0.2n$
Centered spherical Cauchy distribution	$p_n = 2/n$	0.28	0.24	0.22	0.17	0.17	0.14
	$p_n = 1/n$	0.34	0.29	0.25	0.19	0.19	0.16
	$p_n = 1/(2n)$	0.36	0.33	0.29	0.21	0.21	0.17
Centered elliptical Cauchy distribution	$p_n = 2/n$	0.28	0.24	0.21	0.16	0.17	0.14
	$p_n = 1/n$	0.34	0.28	0.25	0.18	0.18	0.16
	$p_n = 1/(2n)$	0.35	0.32	0.28	0.20	0.20	0.17
Centered spherical t-distribution with 4 degrees of freedom	$p_n = 2/n$	0.31	0.39	0.66	0.29	0.54	0.81
	$p_n = 1/n$	0.38	0.47	0.75	0.36	0.61	0.87
	$p_n = 1/(2n)$	0.45	0.54	0.82	0.42	0.68	0.90

4. Simulation study

In this section, we provide a simulation study to illustrate the finite sample performance of our estimator in two- and three-dimensional settings. In three-dimensional cases, we consider two centered spherical distributions, Cauchy distribution and t-distribution with four degrees of freedom. That is, for both distributions we set $\boldsymbol{\mu} = 0$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. In two-dimensional settings simulations are performed for three different distributions. Namely, a centered elliptical Cauchy distribution with nontrivial scatter is considered in addition to the aforementioned centered and spherical Cauchy and t-distributions. In two dimensions, for the nontrivial scatter we set

$$\boldsymbol{\Sigma} = \begin{pmatrix} 11 & 10.5 \\ 10.5 & 11 \end{pmatrix}.$$

Cauchy distribution is quite heavy-tailed since the extreme value index corresponding to the distribution is $\gamma = 1$. On the other hand, t-distribution with four degrees of freedom is less heavy-tailed than the Cauchy distribution and has extreme value index equal to $\gamma = 1/4$.

For each distribution we estimate extreme quantile region corresponding to $p_n \in \{2/n, 1/n, 1/(2n)\}$ for sample sizes $n \in \{1000, 5000\}$ with two different estimators. We use the elliptical extreme quantile region estimator $\hat{Q}_{p_n}$ with the scatter estimated with MCD. For MCD we set $\alpha = 0.5$. We compare our estimator to another extreme quantile region estimator developed in [4]. We denote this alternative estimator by $\bar{Q}_{p_n}$. The estimator $\bar{Q}_{p_n}$ is based on halfspace depth, and as such, it is computationally heavy, but does not require ellipticity. Under ellipticity, density contours and halfspace depth contours coincide. Consequently, estimators $\hat{Q}_{p_n}$ and $\bar{Q}_{p_n}$ estimate the same population quantity for elliptical distributions. Estimates $\hat{Q}_{p_n}$ and $\bar{Q}_{p_n}$ are computed with different values of $k_n \in \{0.05n, 0.1n, 0.2n\}$. The estimator $\bar{Q}_{p_n}$ does not take into account the estimation of the location $\boldsymbol{\mu}$. Thus, we assume the location of the distribution to be known for the fair comparison between $\hat{Q}_{p_n}$ and $\bar{Q}_{p_n}$. Approximation of the relative errors $P(\hat{Q}_{p_n} \triangle Q_{p_n})/p_n$ and $P(\bar{Q}_{p_n} \triangle Q_{p_n})/p_n$ is based on approximating integrals with Riemann sums in polar coordinates, see [38] for details of the computation. Each simulation scenario with different values for n, p_n, k_n and different distributions is repeated $s = 100$ times.

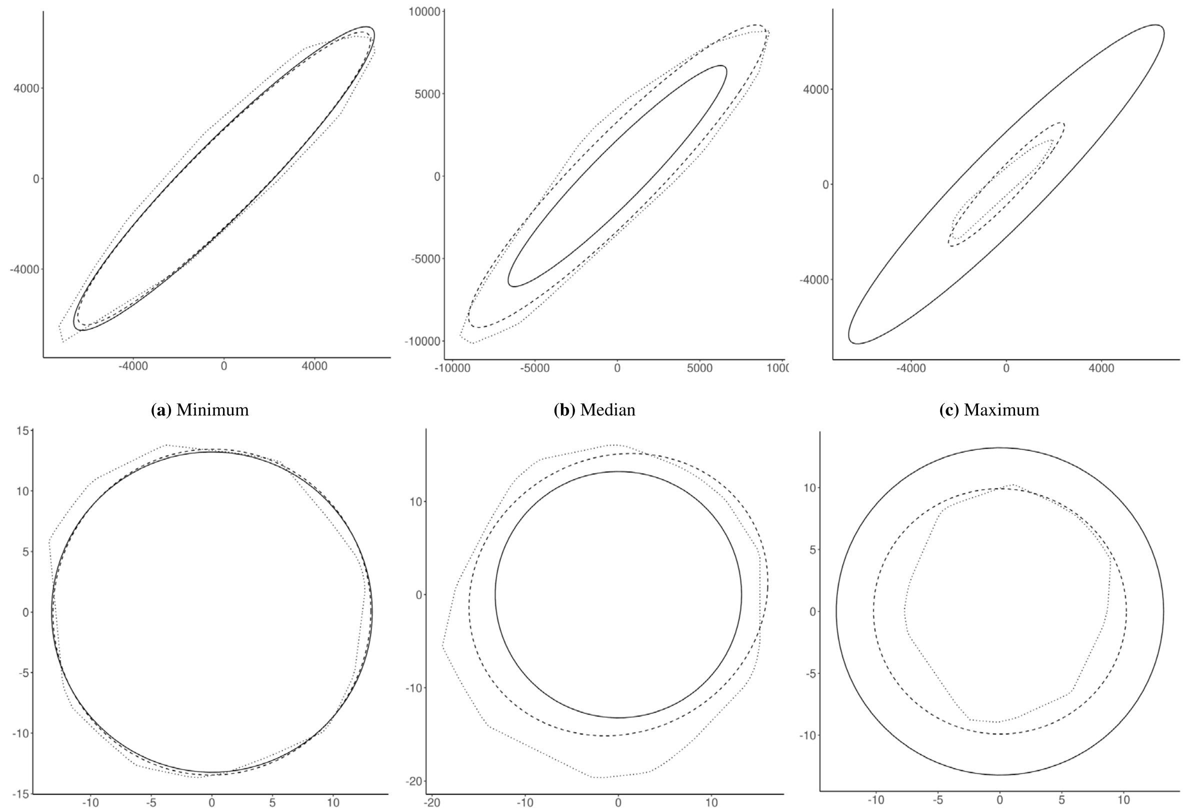
Tables 1 and 2 show simulation results for the two-dimensional settings for the estimators $\hat{Q}_{p_n}$ and $\bar{Q}_{p_n}$, respectively. More precisely, the tables include medians of the approximated relative errors $P(\hat{Q}_{p_n} \triangle Q_{p_n})/p_n$ and $P(\bar{Q}_{p_n} \triangle Q_{p_n})/p_n$. Results for the spherical and elliptical Cauchy distribution are identical for the estimator $\hat{Q}_{p_n}$ up to numerical errors in the integration. This is expected, since our estimator $\hat{Q}_{p_n}$ is affine equivariant, see Theorem 6. Results for the spherical and elliptical Cauchy distribution are also quite similar for the estimator $\bar{Q}_{p_n}$, even though the estimator is not fully affine equivariant, see [4, Proposition 3]. Our estimator $\hat{Q}_{p_n}$ performs marginally better than the estimator $\bar{Q}_{p_n}$ for the Cauchy distribution. For the Cauchy distribution, relative errors decrease for both estimators as the sample size n increases. For the t-distribution, that is the less heavy-tailed case, the performance is slightly weaker for both estimators. It also seems that, especially for the t-distribution, the choice of k has a significant effect on the performance of the estimators. In other words, simulations show that the estimators $\hat{Q}_{p_n}$ and $\bar{Q}_{p_n}$ are not robust for the choice of k. In practice, it is not easy to choose k optimally. The choice of k for the Hill estimator is discussed, e.g., in [39,40].

Fig. 1 shows the best, the median and the worst estimates with respect to the relative errors $P(\hat{Q}_{p_n} \triangle Q_{p_n})/p_n$ and $P(\bar{Q}_{p_n} \triangle Q_{p_n})/p_n$ for particular scenarios when the sample size is set to $n = 1000$. Figures illustrate that there is a considerable amount of variation in the performance of both estimators. However, this is expected in the context of extreme value theory.

Fig. 2 shows boxplots of $s = 100$ relative errors for both estimators $\hat{Q}_{p_n}$ and $\bar{Q}_{p_n}$ for certain three-dimensional scenarios with $n = 1000$, $p_n \in \{2/n, 1/n, 1/(2n)\}$ and both selected distributions (the centered spherical Cauchy distribution and the centered spherical t-distribution with four degrees of freedom). The elliptical estimator $\hat{Q}_{p_n}$ performs marginally better than the competitor $\bar{Q}_{p_n}$ based on medians of the relative errors. Interestingly, it seems that the elliptical estimator has a smaller variation than the depth-based estimator.

Table 2
Simulation results for the extreme quantile region estimator $\bar{Q}_{p_n}$ based on halfspace depth for the two-dimensional settings. The table contains the medians calculated from $s = 100$ approximated relative errors $P(\bar{Q}_{p_n} \triangle Q_{p_n})/p_n$ for all combinations of distributions (centered spherical Cauchy distribution, centered elliptical Cauchy distribution and centered spherical t-distribution with 4 degrees of freedom), $n \in \{1000, 5000\}$, $p_n \in \{2/n, 1/n, 1/(2n)\}$ and $k_n \in \{0.05n, 0.1n, 0.2n\}$.

Distribution	Value of p_n	$n = 1000$			$n = 5000$		
		$k_n = 0.05n$	$k_n = 0.1n$	$k_n = 0.2n$	$k_n = 0.05n$	$k_n = 0.1n$	$k_n = 0.2n$
Centered spherical Cauchy distribution	$p_n = 2/n$	0.38	0.32	0.28	0.21	0.20	0.17
	$p_n = 1/n$	0.42	0.36	0.31	0.23	0.22	0.18
	$p_n = 1/(2n)$	0.49	0.39	0.33	0.26	0.24	0.18
Centered elliptical Cauchy distribution	$p_n = 2/n$	0.54	0.39	0.29	0.24	0.17	0.13
	$p_n = 1/n$	0.58	0.42	0.31	0.28	0.19	0.14
	$p_n = 1/(2n)$	0.61	0.45	0.33	0.31	0.20	0.15
Centered spherical t-distribution with 4 degrees of freedom	$p_n = 2/n$	0.59	0.44	0.34	0.32	0.24	0.56
	$p_n = 1/n$	0.59	0.44	0.44	0.34	0.30	0.68
	$p_n = 1/(2n)$	0.59	0.57	0.59	0.34	0.41	0.77



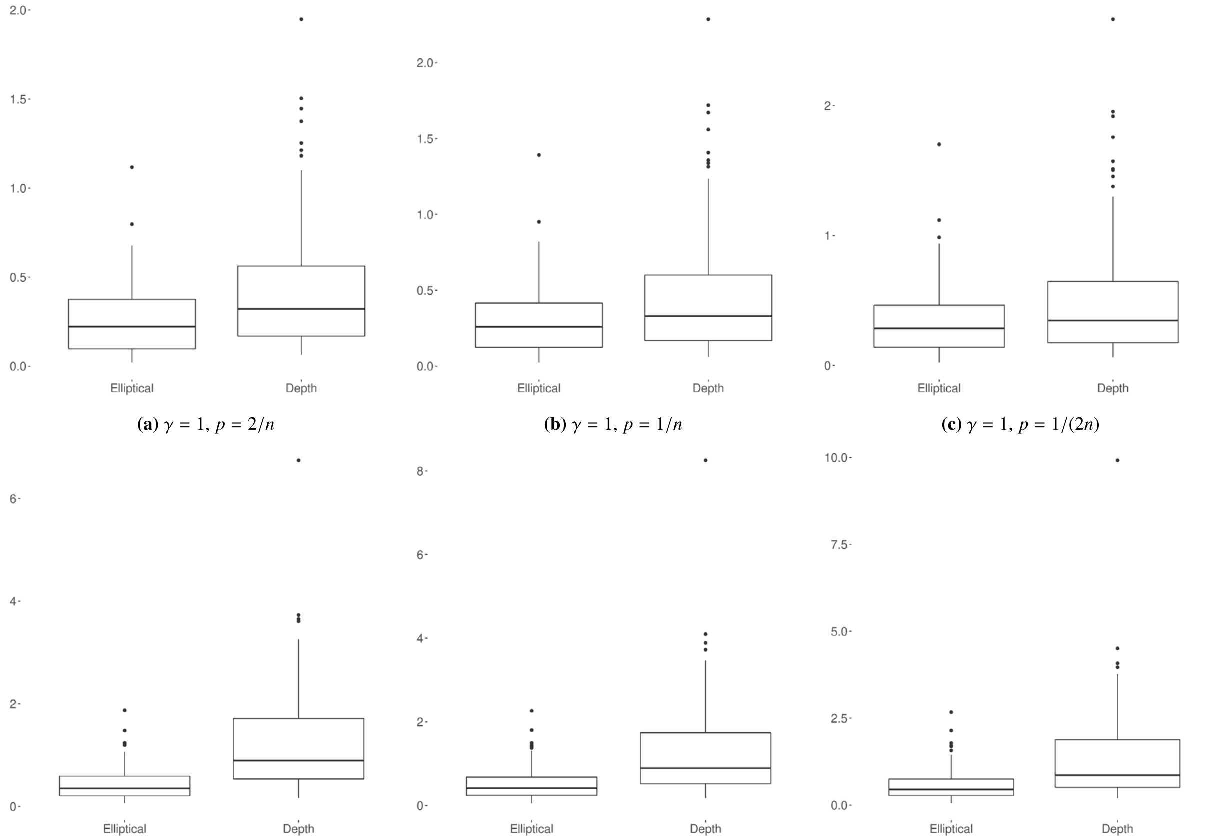
(a) Minimum **(b)** Median **(c)** Maximum

(d) Minimum **(e)** Median **(f)** Maximum

Fig. 1. True and estimated quantile regions for the two-dimensional scenarios where $n = 1000$ and $p_n = 1/(2n)$: ———, true; – – – – $\hat{Q}_{p_n}$; .... $\bar{Q}_{p_n}$. Upper and lower rows correspond to the centered elliptical Cauchy distribution and the centered elliptical t-distribution with 4 degrees of freedom, respectively. Estimates in the upper row were calculated with $k_n = 0.2n$ and estimates in the lower row were calculated with $k = 0.05n$. The leftmost figures correspond to the smallest relative error from $s = 100$ repetitions. Middle figures correspond to the cases with relative error closest to the median and the rightmost figures correspond to the maximum relative error.

We also performed simulations for the elliptical extreme quantile region estimator $\hat{Q}_{p_n}$ with samples generated from the m-dimensional centered spherical Cauchy distribution when $m \in \{2, \ldots, 30\}$. Based on the simulations it seems that if the sample size n is fixed but dimension m increases, then the performance of the estimator $\hat{Q}_{p_n}$ deteriorates slightly. This is visible from the fact that the medians of the errors computed from $s = 100$ repetitions increase as m increases (as dimension m grows from 2 to 30, the median error becomes approximately 2-3 times larger). For details, see Github repository `elliptical-sim` [38].

Lastly, Fig. 3 shows an example where the true quantile regions are not elliptically shaped. For the example we choose a skew-elliptical distribution. Skew-elliptical distributions are a generalization of elliptical distributions where skewness of the tails is



(a) $\gamma = 1,\ p = 2/n$ **(b)** $\gamma = 1,\ p = 1/n$ **(c)** $\gamma = 1,\ p = 1/(2n)$

(d) $\gamma = 1/4,\ p = 2/n$ **(e)** $\gamma = 1/4,\ p = 1/n$ **(f)** $\gamma = 1/4,\ p = 1/(2n)$

Fig. 2. Boxplots of relative errors $P(\hat{Q}_{p_n} \triangle Q_{p_n})/p_n$ (Elliptical) and $P(\bar{Q}_{p_n} \triangle Q_{p_n})/p_n$ (Depth) for $m = 100$ repetitions of selected three-dimensional scenarios with $n = 1000$. Upper row corresponds to the centered spherical Cauchy distribution ($\gamma = 1$) and the lower row corresponds to the centered spherical t-distribution with four degrees of freedom ($\gamma = 1/4$). Left, center and right columns correspond to $p = 2/n$, $p = 1/n$ and $p = 1/(2n)$ respectively. Scenarios correspond the value of k for which the median relative error is the smallest.

allowed. See [41] for a review of skew-elliptical distributions. More specifically, we choose skewed t-distribution with four degrees of freedom, location $\boldsymbol{\mu} = 0$, scale equal to the identity matrix $\boldsymbol{I}$ and skewness parameter $\boldsymbol{\delta} = \left(1/\sqrt{11} \;\; -3/\sqrt{11}\right)$. That is, we consider the distribution of $\boldsymbol{Y} = [\boldsymbol{X} | X_0 > 0]$, where the random vector $(X_0\ \boldsymbol{X})$ has t-distribution with the location $\boldsymbol{\mu} = 0$ and the scatter

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \boldsymbol{\delta}^\intercal \\ \boldsymbol{\delta} & \boldsymbol{I} \end{pmatrix}.$$

The value of the extreme value index for this distribution is $\gamma = 1/4$. In this case we set $n = 5000$, $k = 0.05n$ and estimate quantile regions corresponding to $p_n \in \{2/n, 1/(2n)\}$ with elliptical extreme quantile region estimator $\hat{Q}_{p_n}$. We also estimated the location and the scatter of the distribution with MCD and set $\alpha = 0.5$, even though the distribution is not elliptical. Clearly, the estimator fails in the sense that the shape of the estimated quantile regions is elliptical while the real quantile regions are not. When ellipticity assumption is not satisfied but multivariate regular variation condition still is, one should use, for example, the estimator developed in [3].

5. Empirical illustration

Here we present a real data example in a financial context. Similar examples with the same data for the depth extreme quantile regions and extreme directional multivariate quantiles were provided in [4,9]. Data consists of daily prices for three different stock market indices from the 2nd of July 2001 to the 29th of June 2007: S&P 500 from the USA, FTSE 100 from England and Nikkei 225 from Japan. First, for each series $\boldsymbol{X}^{(i)} = (X_1, \dots, X_{1565})$, $i \in \{1, 2, 3\}$, we compute daily log returns $Y_t^{(i)} = \log\left(X_t^{(i)}/X_{t-1}^{(i)}\right)$, where $t \in \{2, \dots, 1565\}$. However, for the 5% level of significance, the null hypothesis of serial independence of the Ljung–Box test is rejected for the daily log returns corresponding to Nikkei 225. Thus, it is not appropriate to apply the extreme quantile region estimator for the log returns $\boldsymbol{Y}_t = (Y_t^{(1)}, Y_t^{(2)}, Y_t^{(3)})$ directly. Consequently, we follow the procedure in [4] and filter each of the three time series by first fitting EGARCH(1, 1) models, see [42], and then by collecting innovations $\boldsymbol{Z}_t = (Z_t^{(1)}, Z_t^{(2)}, Z_t^{(3)})$.

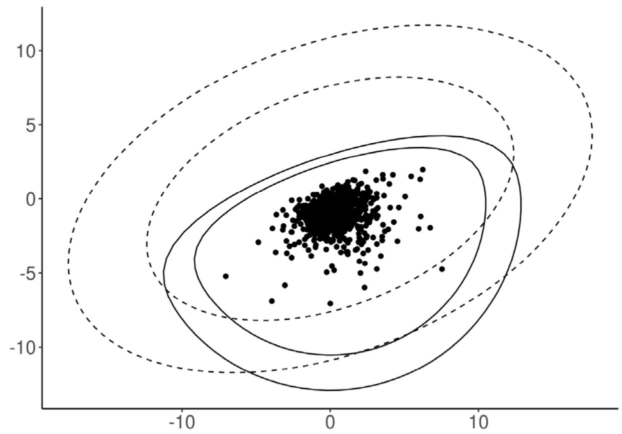


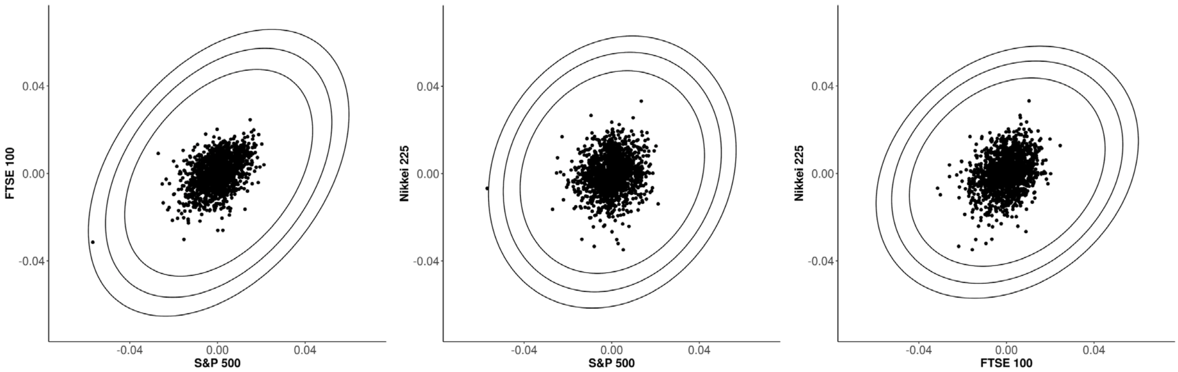
Fig. 3. True and estimated quantile regions when the elliptical extreme quantile region estimator $\hat{Q}_{p_n}$ was used for the skewed t-distribution. In this case $n = 5000$, $k = 0.05n$ and $p_n \in \{2/n, 1/(2n)\}$: •, observations:———, true; – – – – $\hat{Q}_{p_n}$.

We estimate parameters of EGARCH(1, 1) models by maximizing quasi-likelihood and assuming t-distributed innovations $Z_t^{(i)}$ with unknown degrees of freedom. Now, serial independence is not rejected for innovations with 5% level of significance. For details of the computations, see Github repository `elliptical-empirical` [43].

In addition to the independence of observations, we test the assumptions of ellipticity and regular variation of the estimated generating variate. Multiple tests for (multivariate) ellipticity are available in the literature, for example, see [44–48]. We use a Pearson's chi-squared type test [45]. With this test, multivariate ellipticity is not rejected with 5% level of significance. To protect against the type II error in statistical testing, we visualize the univariate marginals by histograms. Based on the approximate symmetry of the histograms, the marginals seem elliptical. We complement the visualizations by bivariate scatter plots. All the scatter plots of the bivariate marginals have an elliptical shape. This further supports the ellipticity assumption as the marginals of an elliptically distributed random variable are elliptically distributed as well. Furthermore, we utilized Q–Q plots to assess whether all the univariate marginals come from the same location-scale family. Points in the Q–Q plots fall approximately on a straight line (excluding one outlying point), which suggests that the univariate marginals come from the same location–scale family. Figures of histograms of $Z^{(i)}$, and scatter plots and Q–Q plots for each pair of $(Z^{(i)}, Z^{(j)})$, $i \neq j$, can be found in Github repository `elliptical-empirical` [43].

Next, let us consider the multivariate regular variation assumption. Multivariate regular variation implies that both tails of each univariate marginal should have the same extreme value indices. When we set $k = 80$, the estimated extreme value indices for both tails of each $Z^{(i)}$ are 0.178, 0.178, 0.223, 0.225, 0.255 and 0.261 from the smallest to the largest. Then based on the asymptotic normality of the Hill estimator one can test if the maximal difference between the extreme value indices is significant. The aforementioned procedure is applied for the data set of this empirical example in [4], and the null hypothesis that the extreme value indices are equal was not rejected. If the extreme value indices of the tails are different from each other one can, for example, transform each univariate marginal (to, e.g, Pareto distribution) to make the multivariate regular variation assumption more plausible, but this is not necessary here. However, in the elliptical setting multivariate regular variation is equivalent to regular variation of the generating variate. Thus, we can just test regular variation of the generating variate. For testing regular variation we use the PE-test [49]. The test involves a free parameter $\eta > 0$, for which we choose the value $\eta = 1/2$ as suggested in [49]. For the number of the tail observations we choose $k = 160$. Note that we do not have observations of the generating variate R_i but only approximated Mahalanobis distances $\hat{R}_i$ are available. Thus, we use estimated Mahalanobis distances $\hat{R}_i$ for performing the PE-test when the location and the scatter are estimated with MCD for $\alpha = 0.5$. The assumption of regular variation is not rejected for the 5% significance level. Notice also that multivariate regular variation of log returns was tested in [50].

Now that we have tested the assumptions of independence, ellipticity and regular variation, we compute bivariate elliptical extreme quantile regions for each pair of innovations $(Z_t^{(i)}, Z_t^{(j)})$, $i \neq j$, $t \in \{2, \ldots, 1565\}$. Predicted quantile regions for the corresponding log returns $(Y_{1566}^{(i)}, Y_{1566}^{(j)})$ are then obtained by affine retransformation by means of the predicted standard deviations and the estimated offsets of the EGARCH(1, 1) models for each country. This prediction corresponds to the 2nd of July 2007 (the next trading day). Fig. 4 shows predicted quantile regions for each pair of the log returns $(Y_{1566}^{(i)}, Y_{1566}^{(j)})$. Regions are computed for probabilities $p \in \{1/2000, 1/5000, 1/10000\}$ and $k = 160$. For each pair of innovations $(Z_t^{(i)}, Z_t^{(j)})$, the location and the scatter are estimated with MCD for $\alpha = 0.5$. Fig. 4 can be used as a tool for inferring the atypicality of log returns $Y_t^{(i)}$. Moreover, points on the boundary of the predicted extreme quantile regions can be used in stress testing. The shape of the regions uncovers the dependency between different stock indices. For example, the shape of the regions in Fig. 4(a) is ellipsoidal with axes of symmetry not equal to the main axes which tells that the risk could be grossly over- or underestimated by using only univariate extreme quantile estimators of the marginals, since univariate estimates do not take into account the dependency structure. On the other hand, the shape of the predicted quantile regions in Fig. 4(b) is closer to spherical. If an investor would like to invest in two index funds, she would possibly



(a) USA versus UK **(b)** USA versus Japan **(c)** UK versus Japan

Fig. 4. Predicted bivariate extreme quantile regions corresponding to the log returns of the 2nd of July 2007 for $p \in \{1/2000, 1/5000, 1/10000\}$. In the computation we chose $k = 160$. Location and scatter were estimated with MCD ($\alpha = 0.5$). Plotted points are obtained by affine transformation of the innovations by means of the predicted standard deviations and the estimated offsets of the EGARCH(1, 1) models for each country.

prefer the pair S&P 500 and Nikkei 225 over S&P 500 and FTSE 100 as the risks related to the former pair seem less dependent. Note also that the estimates of the extreme value indices of the generating variate in all three cases are close to each other (0.20 for USA versus UK, 0.18 for USA versus Japan and 0.18 for UK versus Japan), which is indicated by the fact that the extreme quantile regions for all three pairs are of a similar scale.

Next, we demonstrate the use of elliptical extreme quantile regions in outlier detection. We can say that an observation is an outlier if the corresponding innovation $\boldsymbol{Z}_t$ lies in the estimated extreme quantile region $\hat{Q}_p$ for a fixed small value of p, i.e.,

$$\|\boldsymbol{Z}_t - \hat{\boldsymbol{\mu}}_n\|^2_{\hat{\boldsymbol{\Sigma}}_n} \geq \hat{r}_p,$$

where $\hat{\boldsymbol{\mu}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ are MCD estimates for the location and scatter of the trivariate elliptical distribution of the innovations. Again, in the computation of the MCD estimates we choose $\alpha = 0.5$. For $p = 1/5000$ and $k = 160$ we obtain the following estimates,

$$\hat{\boldsymbol{\mu}}_n = \begin{pmatrix} 0.00 & 0.02 & 0.03 \end{pmatrix}^\intercal, \quad \hat{\boldsymbol{\Sigma}}_n = \begin{pmatrix} 0.99 & 0.44 & 0.14 \\ 0.44 & 0.99 & 0.25 \\ 0.14 & 0.25 & 0.91 \end{pmatrix}, \quad \hat{r}_p = 7.69.$$

With the above estimates, the day 27th of February 2007 was deemed as an outlier. According to [4] the Chinese market index dropped by 9% on the same day, which broke the 10-year record.

Here we showed an empirical example where the original data had to be transformed in order to apply the extreme quantile region estimator $\hat{Q}_p$. Our estimator may be suitable also in other cases where the original data can be bijectively mapped such that the transformed data satisfies the ellipticity assumption.

CRediT authorship contribution statement

Jaakko Pere: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Pauliina Ilmonen:** Conceptualization, Formal Analysis, Writing – review & editing, Supervision. **Lauri Viitasaari:** Conceptualization, Formal analysis, Writing – review & editing, Supervision.

Acknowledgments

The authors wish to thank the two anonymous referees for their insightful comments and careful reading that helped to improve the article.

Jaakko Pere gratefully acknowledges support from the Vilho, Yrjö and Kalle Väisälä Foundation. Pauliina Ilmonen gratefully acknowledges support from the Academy of Finland, Finland, decision number 346308 (The Centre of Excellence in Randomness and Structures). We acknowledge the computational resources provided by the Aalto Science-IT project.

Appendix. Proofs

The proofs of the main results of the article are provided in this appendix. We start by proving several technical lemmas that are then applied in proving the main theorems.

A vector norm induced by a symmetric positive definite matrix $\boldsymbol{H} \in \mathbb{R}^{m\times m}$ is denoted by

$$\|\boldsymbol{x}\|_{\boldsymbol{H}} = \sqrt{\boldsymbol{x}^\intercal \boldsymbol{H}^{-1}\boldsymbol{x}}, \quad \boldsymbol{x} \in \mathbb{R}^m.$$

For $\boldsymbol{H} = \boldsymbol{I}$, we use the notation

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_{\boldsymbol{I}}.$$

A matrix norm induced by a vector norm $\|\cdot\|$ is denoted by

$$\|\boldsymbol{B}\| = \sup\left\{\frac{\|\boldsymbol{B}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} : \boldsymbol{x} \in \mathbb{R}^m \setminus \{\boldsymbol{0}\}\right\},$$

for $\boldsymbol{B} \in \mathbb{R}^{m\times m}$. For example, for a symmetric positive definite matrix $\boldsymbol{H}$, we have that

$$\|\boldsymbol{H}\| = \lambda_{\max},$$

where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{H}$. Complement of an open ellipsoid with location $\boldsymbol{h} \in \mathbb{R}^m$, shape determined by a symmetric positive definite matrix $\boldsymbol{H}$ and scaling factor $r > 0$ is denoted by

$$E(\boldsymbol{h}, \boldsymbol{H}, r) = \{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x} - \boldsymbol{h}\|_{\boldsymbol{H}} \geq r\}.$$

Let $\boldsymbol{X}$ be an elliptically distributed random variable with generating variate $\mathcal{R}$, location $\boldsymbol{\mu}$ and scatter $\boldsymbol{\Sigma}$. Also assume that $\boldsymbol{X}$ satisfies assumptions of Theorem 5. Then for a sufficiently small p_n we can represent $Q_{p_n} = \{\boldsymbol{x} \in \mathbb{R}^m : f(\boldsymbol{x}) \leq \beta\}$ as

$$Q_{p_n} = E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r_{p_n}),$$

where r_{p_n} is the $(1-p_n)$-quantile of the generating variate $\mathcal{R}$. For details of the derivation, see Section 3.2. Without a loss of generality we can assume the above representation for Q_{p_n} throughout the proofs since by the assumptions of Theorem 5 we have that $p_n \to 0$, as $n \to \infty$.

The next lemma gives an upper bound for the length of the longest semiaxis of an open ellipsoid. The result is applied in the proof of Lemma 2.

Lemma 1. *Let $\boldsymbol{h} \in \mathbb{R}^m$, $\boldsymbol{H} \in \mathbb{R}^{m\times m}$ be a positive definite matrix and let $r > 0$. Then*

$$\sup\left\{\|\boldsymbol{x} - \boldsymbol{h}\|^2 : \boldsymbol{x} \in (E(\boldsymbol{h}, \boldsymbol{H}, r))^c\right\} \leq r^2\|\boldsymbol{H}\|.$$

Proof of Lemma 1. Notice that

$$(E(\boldsymbol{h}, \boldsymbol{H}, r))^c = \left\{\boldsymbol{x} : \|\boldsymbol{x} - \boldsymbol{h}\|^2_{r^2\boldsymbol{H}} < 1\right\}.$$

Matrix $\boldsymbol{H}$ is positive definite. Thus, by spectral theorem, we have that $\boldsymbol{H} = \boldsymbol{B}\boldsymbol{B}$, where

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^\intercal,$$

$\boldsymbol{U}$ is orthogonal, $\boldsymbol{\Lambda}^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_m^{1/2})$ and where $\lambda_1 \geq \cdots \geq \lambda_m$ are the eigenvalues of the matrix $\boldsymbol{H}$. Now

$$\|\boldsymbol{x} - \boldsymbol{h}\|^2 = \|(r\boldsymbol{B})(r^{-1}\boldsymbol{B}^{-1})(\boldsymbol{x} - \boldsymbol{h})\|^2 \leq \|r\boldsymbol{B}\|^2\|\boldsymbol{x} - \boldsymbol{h}\|^2_{r^2\boldsymbol{H}} = r^2\|\boldsymbol{H}\|\|\boldsymbol{x} - \boldsymbol{h}\|^2_{r^2\boldsymbol{H}}$$

and it follows that

$$\sup_{\boldsymbol{x}\in(E(\boldsymbol{H},\boldsymbol{h},r))^c}\|\boldsymbol{x} - \boldsymbol{h}\|^2 \leq r^2\|\boldsymbol{H}\| \sup_{\boldsymbol{x}\in(E(\boldsymbol{H},\boldsymbol{h},r))^c}\|\boldsymbol{x} - \boldsymbol{h}\|^2_{r^2\boldsymbol{H}} \leq r^2\|\boldsymbol{H}\|. \quad \square$$

Lemma 2. *Let $\hat{\boldsymbol{\mu}}_n, \boldsymbol{\mu} \in \mathbb{R}^m$, let $\hat{\boldsymbol{\Sigma}}_n, \boldsymbol{\Sigma} \in \mathbb{R}^{m\times m}$ be positive definite, and let $r_{p_n}, \hat{r}_{p_n} > 0$. Let*

$$\tilde{r}_{p_n} = \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}} + \hat{r}_{p_n}\left(\|\hat{\boldsymbol{\Sigma}}_n\|\|\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}_n^{-1}\| + 1\right), \quad \tilde{Q}_{p_n} = E\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \max(\tilde{r}_{p_n}, r_{p_n})\right).$$

Then

$$\tilde{Q}_{p_n} \subset Q_{p_n} \cap \hat{Q}_{p_n}.$$

Proof of Lemma 2. Since $\max(\tilde{r}_{p_n}, r_{p_n}) \geq r_{p_n}$, we have that $\tilde{Q}_{p_n} \subset Q_{p_n}$. Thus it is sufficient to prove that $E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tilde{r}_{p_n}) \subset \hat{Q}_{p_n}$, or equivalently,

$$\hat{Q}^c_{p_n} \subset (E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tilde{r}_{p_n}))^c. \tag{8}$$

We start by proving that

$$\hat{Q}^c_{p_n} \subset \left(E\left(\hat{\boldsymbol{\mu}}_n, \boldsymbol{\Sigma}, \dot{r}_{p_n}\right)\right)^c, \tag{9}$$

where

$$\dot{r}_{p_n} = \hat{r}_{p_n}\left(\left\|\hat{\boldsymbol{\Sigma}}_n\right\|\left\|\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}_n^{-1}\right\| + 1\right).$$

Let $\boldsymbol{x} \in \hat{Q}^c_{p_n}$. Then

$$\begin{aligned}\|\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n\|^2_{\boldsymbol{\Sigma}} &= \left(\|\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n\|^2_{\boldsymbol{\Sigma}} - \|\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n\|^2_{\hat{\boldsymbol{\Sigma}}_n}\right) + \|\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n\|^2_{\hat{\boldsymbol{\Sigma}}_n} < \left|(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n)^\intercal \left(\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}_n^{-1}\right)(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n)\right| + \hat{r}^2_{p_n} \\ &\le \|\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n\|^2 \left\|\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}_n^{-1}\right\| + \hat{r}^2_{p_n} \le \sup_{\boldsymbol{x}\in\hat{Q}^c_{p_n}} \|\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n\|^2 \left\|\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}_n^{-1}\right\| + \hat{r}^2_{p_n} \\ &\le \hat{r}^2_{p_n}\left(\left\|\hat{\boldsymbol{\Sigma}}_n\right\| \left\|\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}_n^{-1}\right\| + 1\right) \le \hat{r}^2_{p_n}\left(\left\|\hat{\boldsymbol{\Sigma}}_n\right\| \left\|\boldsymbol{\Sigma}^{-1} - \hat{\boldsymbol{\Sigma}}_n^{-1}\right\| + 1\right)^2 = \dot{r}^2_{p_n}.\end{aligned}$$

Thus $\boldsymbol{x} \in (E(\hat{\boldsymbol{\mu}}_n, \boldsymbol{\Sigma}, \dot{r}_{p_n}))^c$ and Inclusion (9) holds.

Next let us prove that

$$(E(\hat{\boldsymbol{\mu}}_n, \boldsymbol{\Sigma}, \dot{r}_{p_n}))^c \subset (E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tilde{r}_{p_n}))^c. \tag{10}$$

Let $\boldsymbol{x} \in (E(\hat{\boldsymbol{\mu}}_n, \boldsymbol{\Sigma}, \dot{r}_n))^c$. Now

$$\begin{aligned}\|\boldsymbol{x}-\boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}} &= \|(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n) + (\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu})\|^2_{\boldsymbol{\Sigma}} \le \left(\|(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_n)\|_{\boldsymbol{\Sigma}} + \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}\right)^2 \\ &< \left(\dot{r}_{p_n} + \|\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}\right)^2 = \tilde{r}^2_{p_n}\end{aligned}$$

Thus $\boldsymbol{x} \in (E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tilde{r}_{p_n}))^c$ and Inclusion (10) holds. By combining Inclusions (9) and (10) we have that also Inclusion (8) holds, which completes the proof. □

Next, let us give two lemmas about the relations between R_i and $\hat{R}_i$. The first lemma states that $\hat{R}$ is a consistent estimator of R.

Lemma 3. *Suppose $\boldsymbol{X}$ is absolutely continuous elliptically distributed random variable with location $\boldsymbol{\mu}$ and scatter $\boldsymbol{\Sigma}$. Assume that $\hat{\boldsymbol{\mu}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\mu}$ and $\hat{\boldsymbol{\Sigma}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}$, as $n \to \infty$. Let $R = \|\boldsymbol{X}-\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}$ and $\hat{R} = \|\boldsymbol{X}-\hat{\boldsymbol{\mu}}_n\|_{\hat{\boldsymbol{\Sigma}}_n}$. Then*

$$\hat{R} \xrightarrow{\mathbb{P}} R, \quad n \to \infty.$$

Proof of Lemma 3. We have that

$$|\hat{R}^2 - R^2| \le \underbrace{\left|\|\boldsymbol{X}-\hat{\boldsymbol{\mu}}_n\|^2_{\hat{\boldsymbol{\Sigma}}_n} - \|\boldsymbol{X}-\hat{\boldsymbol{\mu}}_n\|^2_{\boldsymbol{\Sigma}}\right|}_{\text{Term I}} + \underbrace{\left|\|\boldsymbol{X}-\hat{\boldsymbol{\mu}}_n\|^2_{\boldsymbol{\Sigma}} - \|\boldsymbol{X}-\boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}}\right|}_{\text{Term II}}.$$

Now let us consider terms I and II separately. For term I we have

$$\mathrm{I} = \left|(\boldsymbol{X}-\hat{\boldsymbol{\mu}}_n)^\intercal \left(\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}^{-1}\right)(\boldsymbol{X}-\hat{\boldsymbol{\mu}}_n)\right| \le \|\boldsymbol{X}-\hat{\boldsymbol{\mu}}_n\|^2 \left\|\hat{\boldsymbol{\Sigma}}_n^{-1} - \boldsymbol{\Sigma}^{-1}\right\| = O_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1), \quad n \to \infty.$$

Similarly, for term II we apply reverse triangle inequality to get

$$\left|\|(\boldsymbol{X}-\boldsymbol{\mu}) + (\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}_n)\|_{\boldsymbol{\Sigma}} - \|\boldsymbol{X}-\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}\right| \le \|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}_n\|_{\boldsymbol{\Sigma}} = o_{\mathbb{P}}(1)$$

from which we get by continuous mapping theorem that $\mathrm{II} = o_{\mathbb{P}}(1)$. Thus, by continuous mapping theorem

$$\hat{R} \xrightarrow{\mathbb{P}} R, \quad n \to \infty. \quad \square$$

The second lemma about relationship between R_i and $\hat{R}_i$ shows the effect of replacing order statistics $R_{n-k,n}$ with $\hat{R}_{n-k,n}$ when $k_n \to \infty$, $k_n/n \to \infty$, as $n \to \infty$.

Lemma 4 (*[13], Lemma 2.2*). *Suppose $\boldsymbol{X}$ has absolutely continuous elliptical distribution and let $\boldsymbol{X}_1, \ldots \boldsymbol{X}_n$ be i.i.d. copies of $\boldsymbol{X}$. Let $R_i = \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}$ and $\hat{R}_i = \|\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_n\|_{\hat{\boldsymbol{\Sigma}}_n}$, $i \in \{1, \ldots, n\}$. If $k_n \to \infty$, $k_n/n \to 0$, as $n \to \infty$, then we have*

$$\left|R^2_{n-k_n,n} - \hat{R}^2_{n-k_n,n}\right| \le K_n R^2_{n-k_n,n}, \quad n \to \infty,$$

where K_n is a sequence of nonnegative random variables. If $\hat{\boldsymbol{\mu}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\mu}$ and $\hat{\boldsymbol{\Sigma}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}$, as $n \to \infty$, then we can find a sequence K_n such that $K_n \xrightarrow{\mathbb{P}} 0$, as $n \to \infty$.

Now let us review an equivalent formulation of the domain of attraction condition given in Definition 1.

Theorem 7 (*[6], Theorem 1.1.6*). *Let $\gamma \in \mathbb{R}$, let F be a cumulative distribution function of a univariate random variable and let U be the tail quantile function corresponding to F. Then $F \in \mathcal{D}(G_\gamma)$ if and only if there exists a positive function a such that*

$$\lim_{t\to\infty} t\bar{F}(a(t)x + U(t)) = (1+\gamma x)^{-1/\gamma}, \tag{11}$$

for all $x > 0$ with $1 + \gamma x > 0$. Here, for $\gamma = 0$, the right-hand side of the equation is interpreted as e^{-x}.

Below we present two different second order conditions. Both are needed in the proof of Theorem 5. In the definitions of the second order conditions we restrict to the case $\gamma > 0$ as our primary interest is to study heavy-tailed distributions.

Definition 5 (*Second Order Extended Regular Variation*)**.** The tail quantile function U satisfies the second order extended regular variation condition if there exists a positive or a negative function $\tilde{A}$ with $\lim_{t\to\infty} \tilde{A}(t) = 0$ such that for all $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma-1}{\gamma}}{\tilde{A}(t)} = \frac{1}{\tilde{\rho}} \left(\frac{x^{\gamma+\tilde{\rho}}-1}{\gamma+\tilde{\rho}} - \frac{x^\gamma-1}{\gamma} \right),$$

where $\gamma > 0$ and $\tilde{\rho} < 0$.

Definition 6 (*Second Order Regular Variation*)**.** The tail quantile function U satisfies the second order regular variation condition if there exists a positive or a negative function A with $\lim_{t\to\infty} A(t) = 0$ such that for all $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho},$$

where $\gamma > 0$ and $\rho < 0$.

Regular variation and extended regular variation are equivalent conditions in the case $\gamma > 0$, see Theorem 2. However, above second order conditions are not equivalent [6, Example 2.3.11]. Only the other implication holds, that is, if U is of extended regular variation with a function $\tilde{A}$ and a second order parameter $\tilde{\rho} < 0$, then it is of second order regular variation with a different function A and possibly a different second order parameter $\rho < 0$. [51, Corollary 3.1.]. A refined version of Relation (11) holds under the second order extended regular variation condition.

Lemma 5 (*[52], Lemma 2.4.1*)**.** *Let F be a cumulative distribution function of a univariate random variable and let U be the tail quantile function corresponding to F. Suppose F satisfies the second order extended regular variation condition given in Definition 5 with $\gamma > 0$ and $\tilde{\rho} < 0$. Also, assume that $x_0 > -1/\gamma$. Then*

$$\lim_{t\to\infty} \sup_{x\geq x_0} \left| \frac{(1+\gamma x)^{-1/\gamma}}{t\bar{F}(a(t)x + U(t))} - 1 \right| = 0.$$

The proof of Theorem 5 relies on the following lemma.

Lemma 6. *Let $\boldsymbol{X}$ be an m-variate elliptically distributed random variable with absolutely continuous generating variate $\mathcal{R}$, location vector $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$ and let $R = \|\boldsymbol{X} - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}$. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be i.i.d. copies of $\boldsymbol{X}$. Let $F_{\mathcal{R}} \in \mathcal{D}(G_\gamma)$, $\gamma > 0$ and let $\hat{\gamma}_n$ denote the corresponding separating Hill estimator based on a threshold sequence k_n. Let $\hat{\boldsymbol{\mu}}_n$ and $\hat{\boldsymbol{\Sigma}}_n$ denote estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. Let*

$$\hat{r}_{p_n} = \hat{R}_{n-k_n,n} \left(\frac{k_n}{np_n} \right)^{\hat{\gamma}_n},$$

where $\hat{R}_i = \|\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}_n\|_{\hat{\boldsymbol{\Sigma}}_n}$. Suppose the following conditions:

C1. $k_n \to \infty$, $k_n/n \to 0$ *as* $n \to \infty$;
C2. $U_{\mathcal{R}}$ *satisfies the second order extended regular variation condition given in Definition 5;*
C3. $\lim_{n\to\infty} \sqrt{k_n} A(n/k_n) = \lambda \in \mathbb{R}$, *where A is the positive or negative function for $U_{\mathcal{R}}$ in Definition 6;*
C4. $np_n = o(k_n)$ *and* $\ln(np_n) = o\left(\sqrt{k_n}\right)$, *as* $n \to \infty$;
C5. $\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}) = O_{\mathbb{P}}(1)$ *and* $\sqrt{n}(\hat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}) = O_{\mathbb{P}}(1)$.

Then as $n \to \infty$,

$$\frac{\bar{F}_R(\hat{r}_{p_n} + Z_n)}{p_n} \xrightarrow{\mathbb{P}} 1,$$

for any sequence $Z_n = O_{\mathbb{P}}(1)$.

Proof of Lemma 6. The proof of this lemma is similar to the proof of Theorem A.1. in [8]. For reader's convenience we provide all the details here as Theorem A.1. in [8] does neither involve estimated Mahalanobis distances $\hat{R}_i$ nor a sequence Z_n.

This proof relies on second order extended regular variation of R. Note that, as $R \overset{d}{=} \mathcal{R}$, second order extended regular variation of the generating variate $\mathcal{R}$ implies second order extended regular variation of R.

To simplify notations, denote

$$b = b(n/k_n) = U_R\left(\frac{n}{k_n}\right), \ \hat{b} = \hat{b}(n/k_n) = \hat{R}_{n-k_n,n}, a = a(n/k_n), \ \hat{a} = \hat{a}_{n/k_n} = \hat{R}_{n-k_n,n}\hat{\gamma}_n \quad \text{and}$$

$$Y_n = \frac{\hat{b} - b}{a} + \frac{\hat{a}}{a}\frac{d_n^{\hat{\gamma}_n} - 1}{\hat{\gamma}_n} + \frac{Z_n}{a},$$

where $d_n = k_n/np_n$. Now $\frac{Z_n}{a} = O_{\mathbb{P}}\left(\frac{1}{a}\right)$ and

$$\frac{1-F_R(\hat{r}_{p_n}+O_{\mathbb{P}}(1))}{p_n} = d_n \frac{n}{k_n}\bar{F}(aY_n+b) = d_n \frac{\frac{n}{k_n}\bar{F}(aY_n+b)}{(1+\gamma Y_n)^{-1/\gamma}}(1+\gamma Y_n)^{-1/\gamma} - d_n(1+\gamma Y_n)^{-1/\gamma} + d_n(1+\gamma Y_n)^{-1/\gamma}$$
$$= \underbrace{d_n(1+\gamma Y_n)^{-1/\gamma}}_{=I_1}\left(\underbrace{\left(\frac{\frac{n}{k_n}\bar{F}(aY_n+b)}{(1+\gamma Y_n)^{-1/\gamma}}-1\right)}_{=I_2}+1\right).$$

Thus it is sufficient to prove that $I_1 \xrightarrow{\mathbb{P}} 1$ and $I_2 \xrightarrow{\mathbb{P}} 0$ as $n\to\infty$. We continue by proving that

$$d_n^{\hat{\gamma}_n-\gamma} \xrightarrow{\mathbb{P}} 1,\ \frac{\hat{b}-b}{a} \xrightarrow{\mathbb{P}} 0,\ \frac{\hat{a}}{a} \xrightarrow{\mathbb{P}} 1, \quad n\to\infty.$$

After we have proven the above relations we proceed by examining terms I_1 and I_2 separately.

Since

$$\sqrt{k_n}(\hat{\gamma}_n-\gamma) = O_{\mathbb{P}}(1) \tag{12}$$

by [13, Corollary 2.2], we have that

$$d_n^{\hat{\gamma}-\gamma} = \exp\left(\underbrace{\sqrt{k_n}(\hat{\gamma}_n-\gamma)}_{=O_{\mathbb{P}}(1)}\underbrace{\left(\frac{\ln k_n}{\sqrt{k_n}}-\frac{\ln(np_n)}{\sqrt{k_n}}\right)}_{\to 0,\ n\to\infty}\right) \xrightarrow{\mathbb{P}} 1. \tag{13}$$

By [6, Theorem 2.4.1] we have

$$\frac{R_{n-k_n,n}-b}{a} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right), \tag{14}$$

and by [6, Lemma 1.2.9],

$$\lim_{n\to\infty}\frac{b}{a} = \frac{1}{\gamma}. \tag{15}$$

Notice that since both $\hat{R}_{n-k_n,n}$ and $R_{n-k_n,n}$ are nonnegative, it follows from Lemma 4 that

$$\left|\hat{b}-R_{n-k_n,n}\right|^2 \le \left|\hat{b}^2-R^2_{n-k_n,n}\right| \le K_n R^2_{n-k_n,n},$$

where K_n is a nonnegative sequence of random variables such that $K_n \xrightarrow{\mathbb{P}} 0$, as $n\to\infty$. Thus,

$$\left|\hat{b}-R_{n-k_n,n}\right| \le L_n R_{n-k_n,n}, \tag{16}$$

where $L_n = \sqrt{K_n}$. Now, as K_n is nonnegative, continuous mapping theorem yields $L_n \xrightarrow{\mathbb{P}} 0$, as $n\to\infty$. By combining Eqs. (14), (15) and (16) we have that

$$\left|\frac{\hat{b}-b}{a}\right| \le \left|\frac{\hat{b}-R_{n-k_n,n}}{a}\right| + \left|\frac{R_{n-k_n,n}-b}{a}\right| \le L_n\frac{R_{n-k_n,n}}{a}+o_{\mathbb{P}}(1)$$
$$= L_n\underbrace{\left(\frac{R_{n-k_n,n}-b}{a}+\frac{b}{a}\right)}_{\xrightarrow{\mathbb{P}} 1/\gamma}+o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0, \quad n\to\infty.$$

Similarly,

$$\left|\frac{\hat{a}}{a}-1\right| = |\hat{\gamma}_n|\left|\frac{\hat{b}-R_{n-k_n,n}}{a}+\frac{R_{n-k_n,n}-b}{a}+\frac{b}{a}-\frac{1}{\hat{\gamma}_n}\right| \le |\hat{\gamma}_n|\left(\frac{L_nR_{n-k_n,n}}{a}+\left|\frac{R_{n-k_n,n}-b}{a}\right|+\left|\frac{b}{a}-\frac{1}{\hat{\gamma}_n}\right|\right) \xrightarrow{\mathbb{P}} 0, \quad n\to\infty. \tag{17}$$

Let us now consider term I_1. Notice that

$$1+\gamma Y_n = \gamma\underbrace{\frac{\hat{b}-b}{a}}_{\xrightarrow{\mathbb{P}} 0}+\underbrace{\left(1-\frac{\hat{a}}{a}\frac{\gamma}{\hat{\gamma}_n}\right)}_{\xrightarrow{\mathbb{P}} 0}+\frac{\hat{a}}{a}\frac{\gamma}{\hat{\gamma}_n}d_n^{\hat{\gamma}_n}+\underbrace{O_{\mathbb{P}}\left(\frac{1}{a}\right)}_{\xrightarrow{\mathbb{P}} 0} = d_n^{\gamma}\underbrace{\frac{\hat{a}}{a}\frac{\gamma}{\hat{\gamma}_n}d_n^{\hat{\gamma}_n-\gamma}}_{\xrightarrow{\mathbb{P}} 1}+o_{\mathbb{P}}(1)$$
$$= d_n^{\gamma}\left(1+o_{\mathbb{P}}(1)\right),$$

and consequently, by using continuous mapping theorem,

$$I_1 = d_n(d_n^\gamma)^{-1/\gamma}\left(1+o_{\mathbb{P}}(1)\right)^{-1/\gamma} = (1+o_{\mathbb{P}}(1))^{-1/\gamma} \xrightarrow{\mathbb{P}} 1, \quad n\to\infty.$$

Next, consider term I_2. Let $M\in\mathbb{R}$. Notice that

$$\mathbb{P}\left(Y_n \le M\right) = \mathbb{P}(d_n^\gamma \le O_{\mathbb{P}}(1)) + o(1) \to 0, \quad n\to\infty.$$

Choose $x_0 > -\frac{1}{\gamma}$ and let $\varepsilon>0$. Then

$$\mathbb{P}\left(|I_2|>\varepsilon\right) \le \mathbb{P}\left(|I_2|>\varepsilon, Y_n \ge x_0\right) + \underbrace{\mathbb{P}\left(Y_n < x_0\right)}_{\to 0}.$$

Now by Lemma 5 we have that

$$\mathbb{P}\left(|I_2|>\varepsilon, Y_n\ge x_0\right) \le \mathbb{P}\left(\sup_{x\ge x_0}\left|\frac{t\bar{F}(ax+b)}{(1+\gamma x)^{-1/\gamma}}-1\right|>\varepsilon\right)\to 0, \quad n\to\infty.$$

We have proven that $I_1 \xrightarrow{\mathbb{P}} 1$ and that $I_2 \xrightarrow{\mathbb{P}} 0$ as $n\to\infty$, which completes the proof. $\square$

Now we are ready to prove Theorem 5.

Proof of Theorem 5. Let $\tilde{Q}_{p_n}$ be as in Lemma 2, then $\tilde{Q}_{p_n}\subset Q_{p_n}\cap\hat{Q}_{p_n}$. Now it follows from triangle inequality that

$$\begin{aligned}\frac{P(Q_{p_n}\,\triangle\,\hat{Q}_{p_n})}{p_n} &\le \frac{P(Q_{p_n}\,\triangle\,\tilde{Q}_{p_n})}{p_n} + \frac{P(\tilde{Q}_{p_n}\,\triangle\,\hat{Q}_{p_n})}{p_n} = \left(\frac{P(Q_{p_n})}{p_n}-\frac{P(\tilde{Q}_{p_n})}{p_n}\right)+\left(\frac{P(\hat{Q}_{p_n})}{p_n}-\frac{P(\tilde{Q}_{p_n})}{p_n}\right)\\ &= \frac{P(Q_{p_n})}{p_n} - 2\frac{P(\tilde{Q}_{p_n})}{p_n} + \frac{P(\hat{Q}_{p_n})}{p_n}.\end{aligned}$$

By definition of Q_{p_n} we have that $P(Q_{p_n}) = p_n$. Thus, it is sufficient to prove that, as $n\to\infty$

$$\frac{P(\hat{Q}_{p_n})}{p_n} \xrightarrow{\mathbb{P}} 1 \tag{18}$$

and

$$\frac{P(\tilde{Q}_{p_n})}{p_n} \xrightarrow{\mathbb{P}} 1. \tag{19}$$

First, let us show that Relation (18) holds. By Lemma 3, we have that

$$R-\hat{R} \xrightarrow{\mathbb{P}} 0, \quad n\to\infty.$$

Now, by using Lemma 6, we obtain

$$\frac{P(\hat{Q}_{p_n})}{p_n} = \frac{\mathbb{P}(\hat{R}\ge\hat{r}_{p_n})}{p_n} = \frac{\mathbb{P}(R\ge\hat{r}_{p_n}+(R-\hat{R}))}{p_n} = \frac{\bar{F}_R(\hat{r}_{p_n}+(R-\hat{R}))}{p_n} = \frac{\bar{F}_R(\hat{r}_{p_n}+o_{\mathbb{P}}(1))}{p_n} \xrightarrow{\mathbb{P}} 1, \quad n\to\infty.$$

Next, we show that Relation (19) holds. Let us start by proving that

$$\tilde{r}_{p_n} = \hat{r}_{p_n} + O_{\mathbb{P}}(1).$$

Remember that $\sqrt{n}(\hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu})$ and $\sqrt{n}(\hat{\boldsymbol{\Sigma}}_n-\boldsymbol{\Sigma})$ are bounded in probability. Thus,

$$\tilde{r}_{p_n} = \hat{r}_{p_n} + \left\|\hat{\boldsymbol{\mu}}_n-\boldsymbol{\mu}\right\|_{\Sigma} + \hat{r}_{p_n}\|\hat{\boldsymbol{\Sigma}}_n\|\left\|\hat{\boldsymbol{\Sigma}}_n^{-1}-\boldsymbol{\Sigma}^{-1}\right\| = \hat{r}_{p_n} + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) + O_{\mathbb{P}}\left(\frac{\hat{r}_{p_n}}{\sqrt{n}}\right).$$

Hence it is sufficient to show that

$$\frac{\hat{r}_{p_n}}{\sqrt{n}} = O_{\mathbb{P}}(1).$$

Second order extended regular variation implies that

$$\lim_{t\to\infty}\frac{U_R(t)}{t^\gamma} = c \text{ for some } c\in(0,\infty),$$

see [6, page 49]. Since $r_{p_n} = U_R(1/p_n)$, we now have that

$$\lim_{n\to\infty}\frac{r_{p_n}}{(1/p_n)^\gamma} = c. \tag{20}$$

Additionally, by Theorem 3, we have that

$$\frac{R_{n-k_n,n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n}}{r_{p_n}} \xrightarrow{\mathbb{P}} 1, \quad n\to\infty. \tag{21}$$

Since $R_{n-k_n,n} \xrightarrow{a.s.} \infty$, as $n\to\infty$ [13, Lemma 2.1], then the quantity $\hat{R}_{n-k_n}/R_{n-k_n}$ is almost surely well-defined for large n and by Eq. (16),

$$\left|\frac{\hat{R}_{n-k_n}}{R_{n-k_n}}-1\right| = \left|\frac{\hat{R}_{n-k_n}-R_{n-k_n}}{R_{n-k_n}}\right| \le L_n \xrightarrow{\mathbb{P}} 0, \quad n\to\infty. \tag{22}$$

Now, by combining Eqs. (21) and (22), we have that

$$\frac{\hat{r}_{p_n}}{r_{p_n}} = \frac{\hat{R}_{n-k_n,n}}{R_{n-k_n,n}}\frac{R_{n-k_n,n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}_n}}{r_{p_n}} \xrightarrow{\mathbb{P}} 1, \quad n\to\infty. \tag{23}$$

By Condition 4 we have that $p_n^{-\gamma}n^{-1/2}=O(1)$, and combining Eqs. (20) and (23) gives

$$\frac{\hat{r}_{p_n}}{\sqrt{n}} = \underbrace{\frac{r_{p_n}}{(1/p_n)^{\gamma}}}_{\to c}\underbrace{\frac{\hat{r}_{p_n}}{r_{p_n}}}_{\xrightarrow{\mathbb{P}}1}\underbrace{\frac{(1/p_n)^{\gamma}}{\sqrt{n}}}_{=O(1)} = O_{\mathbb{P}}(1).$$

Now, by Lemma 6 we have

$$\frac{P(E(\boldsymbol{\mu},\boldsymbol{\Sigma},\tilde{r}_{p_n}))}{p_n} = \frac{\bar{F}_R(\hat{r}_{p_n}+O_{\mathbb{P}}(1))}{p_n} \xrightarrow{\mathbb{P}} 1, \quad n\to\infty.$$

Recall that

$$\min\{x,y\} = \frac{x+y-|x-y|}{2}, \quad x,y\in\mathbb{R}.$$

Thus

$$\begin{aligned}\frac{P(\tilde{Q}_{p_n})}{p_n} &= \frac{\min\left\{P(Q_{p_n}),P(E(\boldsymbol{\mu},\boldsymbol{\Sigma},\tilde{r}_{p_n}))\right\}}{p_n} = \frac{1}{2}\left(\frac{P(Q_{p_n})}{p_n}+\frac{P(E(\boldsymbol{\mu},\boldsymbol{\Sigma},\tilde{r}_{p_n}))}{p_n}\right)-\frac{1}{2}\left|\frac{P(Q_{p_n})}{p_n}-\frac{P(E(\boldsymbol{\mu},\boldsymbol{\Sigma},\tilde{r}_{p_n}))}{p_n}\right| \\ &\xrightarrow{\mathbb{P}} 1, \quad n\to\infty.\end{aligned}$$

We have now proven Relations (18) and (19), which completes the proof. □

Before we prove affine equivariance of the estimator $\hat{Q}_{p_n}$, we give a simple lemma about affine invariance of Mahalanobis distance.

Lemma 7. *Let $\boldsymbol{H}\in\mathbb{R}^{m\times m}$ be a symmetric positive definite matrix and let $\boldsymbol{B}\in\mathbb{R}^{m\times m}$ be an invertible matrix. Let $\boldsymbol{x}\in\mathbb{R}^m$ and $\boldsymbol{y}=\boldsymbol{B}\boldsymbol{x}+\boldsymbol{b}$. Then*

$$\|\boldsymbol{y}-\boldsymbol{b}\|_{\boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^\intercal} = \|\boldsymbol{x}\|_{\boldsymbol{H}}$$

We call this property the affine invariance of Mahalanobis distance.

Proof of Lemma 7. Using the invertibility of matrix $\boldsymbol{B}$, it is straightforward to obtain

$$\|\boldsymbol{y}-\boldsymbol{b}\|_{\boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^\intercal} = \|\boldsymbol{B}\boldsymbol{x}\|_{\boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^\intercal} = \sqrt{(\boldsymbol{B}\boldsymbol{x})^\intercal(\boldsymbol{B}\boldsymbol{H}\boldsymbol{B}^\intercal)^{-1}(\boldsymbol{B}\boldsymbol{x})} = \sqrt{\boldsymbol{x}^\intercal\boldsymbol{B}^\intercal(\boldsymbol{B}^\intercal)^{-1}\boldsymbol{H}^{-1}\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{x}} = \sqrt{\boldsymbol{x}^\intercal\boldsymbol{H}^{-1}\boldsymbol{x}} = \|\boldsymbol{x}\|_{\boldsymbol{H}}. \quad \square$$

We are now ready to prove Theorem 6.

Proof of Theorem 6. Using the affine invariance of Mahalanobis distance and affine equivariance of the given location and scatter estimators, we obtain

$$\hat{R}_i^y = \hat{R}_i^x.$$

Consequently,

$$\hat{r}_{p_n}^x = \hat{r}_{p_n}^y,$$

and thus,

$$Q_{p_n}^x = E(\hat{\boldsymbol{\mu}}_n(\mathcal{X}),\hat{\boldsymbol{\Sigma}}_n(\mathcal{X}),\hat{r}_{p_n}^y).$$

Now it is sufficient to prove that

$$\boldsymbol{B}E(\hat{\boldsymbol{\mu}}_n(\mathcal{X}), \hat{\boldsymbol{\Sigma}}_n(\mathcal{X}), \hat{r}^y_{p_n}) + \boldsymbol{b} = Q^y_{p_n}.$$

Let $\boldsymbol{y} \in \boldsymbol{B}E(\hat{\boldsymbol{\mu}}_n(\mathcal{X}), \hat{\boldsymbol{\Sigma}}_n(\mathcal{X}), \hat{r}^y_{p_n}) + \boldsymbol{b}$. Then $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{b}$, where $\boldsymbol{x} \in E(\hat{\boldsymbol{\mu}}_n(\mathcal{X}), \hat{\boldsymbol{\Sigma}}_n(\mathcal{X}), \hat{r}^y_{p_n})$. By affine invariance of the Mahalanobis distance we have that

$$\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}_n(\mathcal{Y})\|_{\hat{\Sigma}_n(\mathcal{Y})} = \|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_n(\mathcal{X})\|_{\hat{\Sigma}_n(\mathcal{X})} \geq \hat{r}^y_{p_n}.$$

Thus $\boldsymbol{y} \in Q^y_{p_n}$ and consequently,

$$\boldsymbol{B}E(\hat{\boldsymbol{\mu}}_n(\mathcal{X}), \hat{\boldsymbol{\Sigma}}_n(\mathcal{X}), \hat{r}^y_{p_n}) + \boldsymbol{b} \subset Q^y_{p_n}.$$

Reverse inclusion is proven similarly. Let $\boldsymbol{y} \in Q^y_{p_n}$ and $\boldsymbol{x} = \boldsymbol{B}^{-1}(\boldsymbol{y} - \boldsymbol{b})$. Now, by affine invariance of the Mahalanobis distance,

$$\|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_n(\mathcal{X})\|_{\hat{\Sigma}_n(\mathcal{X})} = \|\boldsymbol{y} - \hat{\boldsymbol{\mu}}_n(\mathcal{Y})\|_{\hat{\Sigma}_n(\mathcal{Y})} \geq \hat{r}^y_{p_n}.$$

Thus $\boldsymbol{x} \in E(\hat{\boldsymbol{\mu}}_n(\mathcal{X}), \hat{\boldsymbol{\Sigma}}_n(\mathcal{X}), \hat{r}^y_{p_n})$ and consequently, $\boldsymbol{y} \in \boldsymbol{B}E(\hat{\boldsymbol{\mu}}_n(\mathcal{X}), \hat{\boldsymbol{\Sigma}}_n(\mathcal{X}), \hat{r}^y_{p_n}) + \boldsymbol{b}$. That is, we have

$$Q^y_{p_n} \subset \boldsymbol{B}Q(\hat{\boldsymbol{\mu}}_n(\mathcal{X}), \hat{\boldsymbol{\Sigma}}_n(\mathcal{X}), \hat{r}^y_{p_n}) + \boldsymbol{b},$$

which completes the proof. □

References

[1] R. Serfling, Quantile functions for multivariate analysis: approaches and applications, Stat. Neerl. 56 (2) (2002) 214–232.
[2] A.B. Tsybakov, On nonparametric estimation of density level sets, Ann. Statist. 25 (3) (1997) 948–969.
[3] J.-J. Cai, J.H. Einmahl, L. De Haan, Estimation of extreme risk regions under multivariate regular variation, Ann. Statist. 39 (3) (2011) 1803–1826.
[4] Y. He, J.H. Einmahl, Estimation of extreme depth-based quantile regions, J. R. Stat. Soc. Ser. B Stat. Methodol. 79 (2) (2017) 449–461.
[5] J.H. Einmahl, L. de Haan, A. Krajina, Estimating extreme bivariate quantile regions, Extremes 16 (2) (2013) 121–145.
[6] L. De Haan, A. Ferreira, Extreme value theory: An introduction, in: Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.
[7] S.I. Resnick, Heavy-Tail Phenomena: Probabilistic and Statistical Modeling, Springer Science & Business Media, 2007.
[8] J.H. Einmahl, J. Li, R.Y. Liu, Thresholding events of extreme in simultaneous monitoring of multiple risks, J. Amer. Statist. Assoc. 104 (487) (2009) 982–992.
[9] R. Torres, E. Di Bernardino, H. Laniado, R.E. Lillo, On the estimation of extreme directional multivariate quantiles, Comm. Statist. Theory Methods 49 (22) (2020) 5504–5534.
[10] S. Girard, G. Stupfler, Intriguing properties of extreme geometric quantiles, REVSTAT-Stat. J. 15 (1) (2017) 107–139.
[11] H. Hult, F. Lindskog, Multivariate extremes, aggregation and dependence in elliptical distributions, Adv. Appl. Probab. 34 (3) (2002) 587–608.
[12] Y. Dominicy, P. Ilmonen, D. Veredas, Multivariate Hill estimators, Internat. Statist. Rev. 85 (1) (2017) 108–142.
[13] M. Heikkilä, Y. Dominicy, P. Ilmonen, On multivariate separating Hill estimator under estimated location and scatter, Statistics 53 (2) (2019) 301–320.
[14] S. Girard, G. Stupfler, Extreme geometric quantiles in a multivariate regular variation framework, Extremes 18 (2015) 629–663.
[15] P. Chaudhuri, On a geometric notion of quantiles for multivariate data, J. Am. Stat. Assoc. 91 (434) (1996) 862–872.
[16] R. Torres, R.E. Lillo, H. Laniado, A directional multivariate value at risk, Insurance Math. Econom. 65 (2015) 111–123.
[17] R. Torres, C. De Michele, H. Laniado, R.E. Lillo, Directional multivariate extremes in environmental phenomena, Environmetrics 28 (2) (2017) e2428.
[18] R.A. Fisher, L.H.C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, Math. Proc. Camb. Phil. Soc. 24 (2) (1928) 180–190.
[19] B. Gnedenko, Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire, Ann. of Math. 44 (3) (1943) 423–453, URL http://www.jstor.org/stable/1968974.
[20] B.M. Hill, A simple general approach to inference about the tail of a distribution, Ann. Statist. (1975) 1163–1174.
[21] D.M. Mason, Laws of large numbers for sums of extreme values, Ann. Probab. (1982) 754–764.
[22] P. Hall, On some simple estimates of an exponent of regular variation, J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (1) (1982) 37–42.
[23] E. Haeusler, J.L. Teugels, On asymptotic normality of Hill's estimator for the exponent of regular variation, Ann. Statist. (1985) 743–756.
[24] L.d. De Haan, L. Peng, Comparison of tail index estimators, Stat. Neerl. 52 (1) (1998) 60–70.
[25] J. Pickands III, et al., Statistical inference using extreme order statistics, Ann. Statist. 3 (1) (1975) 119–131.
[26] A.L. Dekkers, J.H. Einmahl, L. De Haan, A moment estimator for the index of an extreme-value distribution, Ann. Statist. (1989) 1833–1855.
[27] S. Cambanis, S. Huang, G. Simons, On the theory of elliptically contoured distributions, J. Multivariate Anal. 11 (3) (1981) 368–385.
[28] D. Paindaveine, A canonical definition of shape, Statist. Probab. Lett. 78 (14) (2008) 2240–2247.
[29] G. Frahm, Generalized Elliptical Distributions: Theory and Applications (Ph.D. thesis), Universität zu Köln, 2004.
[30] P.J. Rousseeuw, Multivariate estimation with high breakdown point, Math. Stat. Appl. 8 (283–297) (1985) 37.
[31] C. Croux, G. Haesbroeck, Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, J. Multivariate Anal. 71 (2) (1999) 161–190.
[32] P.J. Rousseeuw, K.V. Driessen, A fast algorithm for the minimum covariance determinant estimator, Technometrics 41 (3) (1999) 212–223.
[33] R. Butler, P. Davies, M. Jhun, Asymptotics for the minimum covariance determinant estimator, Ann. Statist. (1993) 1385–1400.
[34] E.A. Cator, H.P. Lopuhaä, Asymptotic expansion of the minimum covariance determinant estimators, J. Multivariate Anal. 101 (10) (2010) 2372–2388.
[35] E.A. Cator, H.P. Lopuhaä, Central limit theorem and influence function for the MCD estimators at general multivariate distributions, Bernoulli 18 (2) (2012) 520–551.
[36] M. Hubert, M. Debruyne, P.J. Rousseeuw, Minimum covariance determinant and extensions, Wiley Interdiscip. Rev. Comput. Stat. 10 (3) (2018) e1421.
[37] X. He, G. Wang, Convergence of depth contours for multivariate datasets, Ann. Statist. (1997) 495–504.
[38] J. Pere, elliptical-sim, , 2024, Github repository, Github, https://github.com/perej1/elliptical-sim/releases/tag/v0.4.0.
[39] J. Danielsson, L. de Haan, L. Peng, C.G. de Vries, Using a bootstrap method to choose the sample fraction in tail index estimation, J. Multivariate Anal. 76 (2) (2001) 226–248.
[40] M.I. Gomes, O. Oliveira, The bootstrap methodology in statistics of extremes—choice of the optimal sample fraction, Extremes 4 (2001) 331–358.
[41] M.D. Branco, D.K. Dey, A general class of multivariate skew-elliptical distributions, J. Multivariate Anal. 79 (1) (2001) 99–113.

[42] D.B. Nelson, Conditional heteroskedasticity in asset returns: A new approach, Econometrica (1991) 347–370.
[43] J. Pere, elliptical-empirical, 2024, Github repository, Github, https://github.com/perej1/elliptical-empirical/releases/tag/v0.2.0.
[44] V. Koltchinskii, L. Sakhanenko, Testing for ellipsoidal symmetry of a multivariate distribution, in: High Dimensional Probability II, Springer, 2000, pp. 493–510.
[45] F.W. Huffer, C. Park, A test for elliptical symmetry, J. Multivariate Anal. 98 (2) (2007) 256–281.
[46] A. Manzotti, F.J. Pérez, A.J. Quiroz, A statistic for testing the null hypothesis of elliptical symmetry, J. Multivariate Anal. 81 (2) (2002) 274–285.
[47] J.R. Schott, Testing for elliptical symmetry in covariance-matrix-based analyses, Statist. Probab. Lett. 60 (4) (2002) 395–404.
[48] S. Babić, L. Gelbgras, M. Hallin, C. Ley, Optimal tests for elliptical symmetry: specified and unspecified location, Bernoulli 27 (4) (2021) 2189–2216.
[49] J. Hüsler, D. Li, On testing extreme value conditions, Extremes 9 (2006) 69–86.
[50] J.H. Einmahl, F. Yang, C. Zhou, Testing the multivariate regular variation model, J. Bus. Econom. Statist. 39 (4) (2021) 907–919.
[51] M.I.F. Alves, M.I. Gomes, L. de Haan, C. Neves, A note on second order conditions in extreme value theory: linking general and heavy tail conditions, REVSTAT-Stat. J. 5 (3) (2007) 285–304.
[52] D. Li, On Extreme Value Approximation to Tails of Distribution Functions (Ph.D. thesis), Erasmus University Rotterdam, 2004.