CentER

Discussion Paper

No. 2006–104

EXTREME VALUE THEORY APPROACH TO SIMULTANEOUS MONITORING AND TRESHOLDING OF MULTIPLE RISK INDICATORS

By John H.J. Einmahl, Jun Li, Regina Y. Liu

October 2006

ISSN 0924-7815

Tilburg University

Extreme Value Theory Approach to Simultaneous Monitoring and Thresholding of Multiple Risk Indicators [1]

John H.J. Einmahl, Jun Li, Regina Y. Liu

Abstract

Risk assessments often encounter extreme settings with very few or no occurrences in reality. Inferences about risk indicators in such settings face the problem of insufficient data. Extreme value theory is particularly well suited for handling this type of problems. This paper uses a multivariate extreme value theory approach to establish thresholds for signaling levels of risk in the context of simultaneous monitoring of multiple risk indicators. The proposed threshold system is well justified in terms of extreme multivariate quantiles, and its sample estimator is shown to be consistent. As an illustration, the proposed approach is applied to developing a threshold system for monitoring airline performance measures. This threshold system assigns different risk levels to observed airline performance measures. In particular, it divides the sample space into regions with increasing levels of risk. Moreover, in the univariate case, such a thresholding technique can be used to determine a suitable cut-off point on a runway for holding short of landing aircrafts. This cut-off point is chosen to ensure a certain required level of safety when allowing simultaneous operations on two intersecting runways in order to ease air traffic congestion.

Key words: Extreme value theory, extreme quantile, multiple risk indicators, multivariate quantile, rare event, statistics of extremes, threshold system.

---

[1] John H.J. Einmahl is Professor, Department of Econometrics & OR and CentER, Tilburg University, Tilburg, The Netherlands (Email: *j.h.j.einmahl@uvt.nl*). Jun Li is Assistant Professor, Department of Statistics, University of California, Riverside, CA 92521-0138 (Email: *jun.li@ucr.edu*). Regina Y. Liu is Professor, Department of Statistics, Rutgers University, Hill Center, Piscataway, NJ 08854-8019 (E-mail: *rliu@stat.rutgers.edu*). This research is supported in part by grants from the *National Science Foundation*, the *National Security Agency*, and the *Federal Aviation Administration*. The discussion on aviation safety in this paper reflects the views of the authors, who are solely responsible for the accuracy of the analysis results presented herein, and does not necessarily reflect the official view or policy of the FAA. The data set used in this paper has been partially masked in order to protect confidentiality.

1 Introduction

Many real life problems require identifying some threshold point such that the probability of exceeding the threshold is no greater than a prescribed level, say p. The task here is simply to locate the $(1-p)$-th quantile of the underlying distribution. If p is not too small, then the usual quantile estimator based on the empirical distribution generally provides a good solution. However, if p is very small, then there may not be sufficient observations in the sample to provide such a useful quantile estimate. For example, given a sample of 100 observations, the $(1-p)$-th quantile for $p = 0.001$ would not be well estimated by the usual quantile estimator due to the no occurrence of such observations. When making inferences for such extreme events with very few or no occurrences, extreme value theory is particularly useful and it has been applied to many real life applications with much success. A classical example is the calculation of the height of sea dikes in the Netherlands in the context of flood prevention (e.g. for $p = 10^{-4}$ per year). More applications can be found in finance and insurance, such as estimating the so-called Value-at-Risk and the related stress testing for equity portfolios or determining premiums for insurance contracts. Other fields of application include sports statistics, meteorology, and seismology. There are several excellent treatises on the subject and its applications, see, for example, Embrechts, Klüppelberg and Mikosch (1997), Coles (2001), Beirlant et al. (2004), and de Haan and Ferreira (2006).

The goal of this paper is to discuss extreme value theory, apply it to deriving extreme quantiles, and develop inference for these quantiles. This project is motivated by the need for a threshold system for flagging extreme risks in an aviation monitoring scheme which can be useful to monitoring agencies such as the FAA (Federal Aviation Administration). The notion of univariate quantiles is well defined, and the extreme quantile is well treated in univariate extreme value theory. However, for different purposes, there may be different notions of *multivariate* quantiles. In this paper, we propose to define a $(1-p)$-th quantile as a lower orthant (quadrant in the bivariate case) of the sample space for which the exceedance probability of any component variate is no more than p. This proposed multivariate extreme quantile is suitable for thresholding in risk assessment in the multivariate setting, since any observation falls beyond the proposed quantile would imply that at least one of its component variates exceeds a certain allowable threshold. We provide an estimator for the proposed quantile, and show its consistency. Note that the definition of consistency in this case also requires some modification from the usual definition of consistency due to the extremely small value of p. Furthermore, to broaden the applicability of our threshold system, we allow the multivariate extreme quantile to take into account different weights assigned to different component variables. Different weights may arise in different applications, and they can be used to reflect the perceived difference in importance of the exceedance in individual component variables. A such example is elaborated in Section 4 in the application of aviation risk assessment.

Another aviation application, which we will not pursue in depth in this paper, is the task of choosing for an airport runway a threshold point beyond which the runway crossing could be allowed. Due to the recent explosive growth in air traffic, the shortage of runway capacity remains the bottle neck for most airport operations and causes much delay and congestion in air traffic. While the construction of additional runways is being sought in due process, the FAA may consider implementing the so-called LAHSO (land and hold short operations) on aircraft landings to help ease air traffic. LAHSO would require all aircraft landings to be accomplished before a predetermined hold short point on the runway. The advantage of implementing LAHSO is to free up the capacity of a certain portion of the runway to allow for other usage, and, in turn, to reduce air traffic congestion. To establish an acceptable land-and-hold-short point on the runway, public safety concerns generally require that the portion of the runway from its touchdown to the hold short point constitute an available safe landing distance. Specifically, this requires that the probability that the full-stop of a landing aircraft occurs beyond the hold short point is no more than one-out-of-ten-million. This amounts to determining the (1-0.0000001)-th quantile of the distribution of landing distance for all aircrafts. Typically a data set consists of the landing distances of about 1000 aircrafts on a given airport runway. Since $1000 \times 0.0000001 = 0.0001 << 1$, this is a setting with no occurrence. The univariate extreme quantile estimator discussed in Section 2 would be ideally suited for this application.

The paper is organized as follows. In Section 2, we review briefly extreme value theory in the univariate setting, and discuss estimators for extreme quantiles. In Section 3, we review extreme value theory in the bivariate setting, propose a definition for bivariate extreme quantiles and provide corresponding estimators. We also show that these estimators are consistent, in a well justified sense of consistency. Finally, in Section 4, we apply the proposed notions of extreme quantiles to establish a threshold system for a monitoring scheme for assigning different levels of risk to observed measurements. Specifically, two aviation risk indicators for monitoring the performance of air carriers are used to demonstrate our approach to constructing a threshold system. This threshold system divides the sample space into regions with increasing levels of risk. These regions, for example, are referred to as "informational", "expected", "advisory" and "concern" in aviation safety analysis. We discuss in detail the construction of these thresholds, as well as the application of the thresholds to a real data set.

2 Monitoring one risk indicator – Univariate extreme quantile

Assume that $X_1, \ldots, X_n$ is a random sample from an unknown, univariate, continuous distribution function F. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of $X_1, \ldots, X_n$. Our task is to obtain the $(1-p)$-th quantile of F, or more specifically, to obtain x_p such that

$$x_p = \inf\{x \in I\!R : P(X > x) \leq p\}.$$

The straightforward nonparametric estimator of x_p is the usual quantile estimator based on the empirical distribution, namely

$$\tilde{x}_p = \inf\{x \in I\!R : \sum_{i=1}^{n} I_{\{X_i > x\}}/n \leq p\}.$$

However, if p is small, there may not be sufficient observations in the sample to render this estimate useful in practice. For example, with a sample of 1000 observations, the 0.0001-th quantile would not be well estimated by the above formula. For the inference related to such extreme quantiles of a probability distribution, extreme value theory is very useful, as shown below.

Statistical inference generally involves the central limit theorem, which characterizes the limiting distribution of the sum $S_n := X_1 + X_2 + \cdots + X_n$. In extreme value theory, our focus is mainly on the sample maximum rather than the sum. Specifically, we would search for a sequence of positive numbers $\{a_n; n \geq 1\}$ and another sequence of numbers $\{b_n; n \geq 1\}$, such that

$$\lim_{n\to\infty} P\left(\frac{X_{n:n} - b_n}{a_n} \leq x\right) = G(x) \tag{2.1}$$

for all $x \in I\!R$ at which the limiting distribution function G is continuous. Here G is a non-degenerate distribution function. If such sequences a_n and b_n exist, F is said to be in the *domain of attraction* of G, denoted by $F \in D(G)$. If $F \in D(G)$, then much of the tail behavior of F can be characterized by G. Fisher and Tippett (1928) and Gnedenko (1943) have shown that G (apart from a location and scale constant) is of the form

$$G(x) = G_\gamma(x) = \exp\big(-(1+\gamma x)^{-1/\gamma}\big), \quad 1+\gamma x > 0, \;\; \gamma \in I\!R \tag{2.2}$$

(by convention, $(1+\gamma x)^{-1/\gamma} = e^{-x}$ for $\gamma = 0$).

These distributions are referred to as extreme value distributions.

The parameter γ is called the extreme value index. It characterizes the tail behavior of F in terms of its degree of heaviness. More specifically:

i) $\gamma > 0$ (G is referred to as a Fréchet distribution) $\Longrightarrow$ F has a heavy tail,

ii) $\gamma < 0$ (G is referred to as a reverse Weibull distribution) $\Longrightarrow$ F has a finite endpoint,

iii) $\gamma = 0$ (G is referred to as a Gumbel distribution) $\Longrightarrow$ F has a light tail.

For example, a Cauchy distribution is a heavy tailed distribution and its corresponding γ is 1; a uniform distribution on the interval [0,1] has a finite endpoint and its corresponding γ is -1; and a normal distribution is attracted by the Gumbel distribution with the corresponding $\gamma = 0$.

Clearly, the parameter γ determines G. To estimate γ, we define

$$M_n^{(j)} = \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n-i:n} - \log X_{n-k:n})^j, \quad 1 < k < n, \quad j \in I\!N, \tag{2.3}$$

$$\hat{\gamma}_n^+ = M_n^{(1)},$$

$$\hat{\gamma}_n^- = 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1}. \tag{2.4}$$

The estimator $\hat{\gamma}_n^+$ was proposed in Hill (1975), and is generally referred to as the Hill estimator. It has been shown that $\hat{\gamma}_n^+$ is consistent and asymptotically normal, when

$\gamma > 0$. Dekkers, Einmahl and de Haan (1989) have constructed the moment estimator

$$\hat{\gamma}_n = \hat{\gamma}_n^+ + \hat{\gamma}_n^-, \tag{2.5}$$

and shown that this estimator is consistent and asymptotically normal for a general $\gamma \in I\!R$.

We now return to the task of using extreme value theory to estimate an extreme quantile. We first observe that (2.1) implies (by taking logarithms)

$$\lim_{t\to\infty} t(1 - F(a_t x + b_t)) = -\log G_\gamma(x) = (1 + \gamma x)^{-1/\gamma}, \quad G_\gamma(x) > 0,$$

where t now runs through $I\!R^+$, and a_t and b_t are defined by interpolation. Setting $y = a_t x + b_t$, we obtain heuristically

$$1 - F(y) \approx \frac{1}{t}\left(1 + \gamma \frac{y - b_t}{a_t}\right)^{-1/\gamma}.$$

Since the p-th quantile of F, x_p, satisfies $1 - F(x_p) = p$, the above approximation yields, with $t = \frac{n}{k}$,

$$x_p \approx \frac{(\frac{k}{np})^\gamma - 1}{\gamma} a_{n/k} + b_{n/k}\,. \tag{2.6}$$

The normalizing sequences $a_{n/k}$ and $b_{n/k}$ can be estimated by

$$\begin{aligned} \hat{b}_{n/k} &= X_{n-k:n}, \\ \hat{a}_{n/k} &= X_{n-k:n} M_n^{(1)} (1 - \hat{\gamma}_n^-). \end{aligned} \tag{2.7}$$

Plugging in (2.6) the above estimators as well as the estimator from (2.5), we obtain the following estimator for x_p

$$\hat{x}_p = \frac{(\frac{k}{np})^{\hat{\gamma}_n} - 1}{\hat{\gamma}_n} \hat{a}_{n/k} + \hat{b}_{n/k}, \tag{2.8}$$

see Dekkers, Einmahl and de Haan (1989).

2.1 The choice of k

Since the expressions (2.3) to (2.8) above all involve k, the properties of the estimators $\hat{x}_p$ obviously depend on the choice of k. The value k can be viewed as the effective sample size for tail extrapolations. If k is too small, then the estimator tends to have a large variance, whereas if k is too large, then the bias tends to dominate. This point can be easily illustrated using the LAHSO project as an example. Since larger aircrafts generally require longer landing distances on the runway, the higher landing distance values observed in the data set should be more relevant for the inference for the extreme landing pattern. If too many landing distances observed from the small aircrafts are included in determining the hold short point, which amounts to choosing too large a k, the outcome is likely to be quite biased.

One commonly used heuristic approach for choosing k in practice is to plot the estimated quantile $\hat{x}_p$ versus k, and choose a k which corresponds to the first stable part of the plot. This visual approach is simple but lacks precise statistical justification. Moreover, in many situations, it can be difficult to identify the first stable part of the plot. To overcome this problem, we may look for the theoretically optimal k by minimizing the mean squared error of $\hat{x}_p$, which is defined as

$$MSE(n,k) = E(\hat{x}_p - x_p)^2. \tag{2.9}$$

Unfortunately, the optimal choice of k clearly depends on the unknown x_p. This problem can be circumvented by considering an analogue of this mean squared error, that contains no unknown parameters and hence can be computed from the data only. This analogue is obtained by replacing x_p by an estimator different from the one in (2.8). Following this idea, Ferreira, de Haan and Peng (2003) defined

$$\begin{aligned}
\hat{\gamma}_{n,1}(k) &= M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1} \quad (= \hat{\gamma}_n), \\
\hat{\gamma}_{n,2}(k) &= \sqrt{\frac{M_n^{(2)}}{2}} + 1 - \frac{2}{3}\left(1 - \frac{M_n^{(1)} M_n^{(2)}}{M_n^{(3)}}\right)^{-1}, \\
\hat{a}_{n/k,1} &= \frac{1}{2} X_{n-k:n} M_n^{(1)} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1} \quad (= \hat{a}_{n/k}), \\
\hat{a}_{n/k,2} &= \frac{2}{3} X_{n-k:n} M_n^{(1)} \left(1 - \frac{M_n^{(1)} M_n^{(2)}}{M_n^{(3)}}\right)^{-1}.
\end{aligned} \tag{2.10}$$

Recall that

$$M_n^{(j)} = \frac{1}{k}\sum_{i=0}^{k-1} (\log X_{n-i:n} - \log X_{n-k:n})^j, \quad 1 < k < n,$$

as defined in (2.3). The following two estimators for x_p can then be obtained:

$$\hat{x}_{n,1}(k) = X_{n-k:n} + \hat{a}_{n/k,1} \frac{(\frac{k}{np})^{\hat{\gamma}_{n,1}(k)} - 1}{\hat{\gamma}_{n,1}(k)},$$

$$\hat{x}_{n,2}(k) = X_{n-k:n} + \hat{a}_{n/k,2} \frac{\left(\frac{k}{np}\right)^{\hat{\gamma}_{n,2}(k)} - 1}{\hat{\gamma}_{n,2}(k)} . \tag{2.11}$$

Note that $\hat{x}_{n,1}(k)$ above is the same as $\hat{x}_p$ in (2.8), and $\hat{x}_{n,2}(k)$ is an alternative estimator for x_p. Ferreira, de Haan and Peng (2003) then considered replacing $MSE(n,k)$ in (2.9) with

$$E\left(\hat{x}_{n,1}(k) - \hat{x}_{n,2}(k)\right)^2 , \tag{2.12}$$

and proceeded to develop a double bootstrap procedure for (2.12) as a way to determine the optimal k in an asymptotic version of (2.9). The detailed algorithm can be outlined in the following steps:

1) Randomly draw a bootstrap sample $\{X_i^*, 1 \le i \le n_1\}$ from $\{X_i, 1 \le i \le n\}$ with $n_1 < n$;

2) Select $\{X_i^*, 1 \le i \le n_2\}$, a subset of size n_2 from the bootstrap sample in step 1, where $n_2 = n_1^2/n < n_1$;

3) Compute $\hat{x}_{n_1,1}(k)$, $\hat{x}_{n_1,2}(k)$, $\hat{x}_{n_2,1}(k)$ and $\hat{x}_{n_2,2}(k)$ in (2.11) based on the two bootstrap samples obtained respectively in steps 1 and 2;

4) Repeat steps 1-3 independently, sufficiently many, say B, times.
   Calculate, for $i = 1, 2$,

$$\widehat{MSE}^*(n_i, k) = \frac{1}{B} \sum_{j=1}^{B} \left(\hat{x}^{*(j)}_{n_i,1}(k) - \hat{x}^{*(j)}_{n_i,2}(k) \right)^2 , \tag{2.13}$$

   where $\hat{x}^{*(j)}_{n_i,1}(k)$ and $\hat{x}^{*(j)}_{n_i,2}(k)$ are the $\hat{x}_{n_i,1}(k)$ and $\hat{x}_{n_i,2}(k)$ based on the j-th bootstrap sample.

5) Find a $\hat{k}_i$ which minimizes $\widehat{MSE}^*(n_i, k)$, $i = 1, 2$ ($\hat{k}_i$ not too close to 1 or n_i).

6) The optimal choice of k in the estimator $\hat{x}_{n,1}(k)$ is then given by

$$\hat{k}_0 = \frac{\hat{k}_1^2}{\hat{k}_2} g(\hat{\gamma}_n, \hat{\rho}), \tag{2.14}$$

   where, if $\hat{\gamma}_n > 0$,

$$g(\hat{\gamma}_n, \hat{\rho}) = \left(\frac{\hat{\rho}^2}{(1-\hat{\rho})^2} \right)^{1/(1-2\hat{\rho})} .$$

For $\hat{\gamma}_n < 0$ the expression for g can be obtained similarly. The details can be found in Ferreira, de Haan and Peng (2003) for details. To proceed with the case of $\hat{\gamma}_n > 0$, we consider

$$\hat{\rho} = 3 + \frac{6}{T_n - 3}, \tag{2.15}$$

with

$$T_n = \frac{\hat{g}_1 - \hat{g}_2}{\hat{g}_2 - \hat{g}_3}, \quad \hat{g}_j = \frac{j+1}{j}\left(1 - \frac{M_n^{(j)} M_n^{(1)}}{M_n^{(j+1)}}\right), \; j = 1, 2, 3 \;\; \text{(see (2.3) for } M_n^{(\cdot)}\text{)}.$$

The estimator $\hat{\rho}$ was constructed in Fraga Alves, de Haan and Lin (2003). Clearly, $\hat{\rho}$ also depends on k. Plot $\hat{\rho}$ against k and choose the $\hat{\rho}$-value of the first stable part of the plot. Generally we require that the corresponding k-values are not too small.

Once k is chosen following the above procedure, the estimate for the extreme quantile $\hat{x}_p$ in (2.8) can be obtained immediately.

3 Monitoring two risk indicators – Multivariate extreme quantiles

We now consider an application of extreme value theory in the bivariate case to establish a threshold system for the simultaneous monitoring of two measurements, which are possibly correlated. Although we present here only bivariate extreme quantiles, the extension to the higher dimensional case is straightforward. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from an unknown, continuous distribution function F. Denote with $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$ the marginal distributions of F. Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ and $Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}$ denote the order statistics of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. Similar to the univariate extreme value theory, F is assumed to belong to the domain of attraction of an extreme value distribution. In other words, there exist sequences $\{a_{1n} > 0; n \geq 1\}$, $\{b_{1n}; n \geq 1\}$, $\{a_{2n} > 0; n \geq 1\}$ and $\{b_{2n}; n \geq 1\}$ such that

$$\left(\frac{X_{n:n} - b_{1n}}{a_{1n}}, \frac{Y_{n:n} - b_{2n}}{a_{2n}}\right) \xrightarrow{d} G(x, y) \tag{3.1}$$

where G has non-degenerate marginal distributions. Clearly, this implies that $G(x, \infty)$ and $G(\infty, y)$ are univariate extreme value distributions. Therefore, with properly chosen sequences, we can obtain that

$$G_1(x) := G(x, \infty) = \exp(-(1 + \gamma_1 x)^{-1/\gamma_1}),$$

and

$$G_2(x) := G(\infty, y) = \exp(-(1 + \gamma_2 x)^{-1/\gamma_2}),$$

for some γ_1, $\gamma_2 \in I\!R$, where $1 + \gamma_1 x > 0$, and $1 + \gamma_2 x > 0$.

For deriving extreme quantiles in the bivariate case, in addition to the quantiles from the two marginal distributions, the tail dependence structure between the two component variables is also an important feature. We briefly describe this tail dependence structure. Denote with C the distribution function of the pair $(1 - F_1(X_1), 1 - F_2(Y_1))$. Note that (3.1) implies that

$$\lim_{t \downarrow 0} \frac{1}{t} C(tx, ty) = x + y - l(x, y), \tag{3.2}$$

where

$$l(x, y) = -\log G\left(\frac{x^{-\gamma_1} - 1}{\gamma_1}, \frac{y^{-\gamma_2} - 1}{\gamma_2}\right).$$

A bivariate probability distribution function F is said to have a *tail dependence function* l if (3.2) holds for $x, y \geq 0$. We list below two key properties of l:

(i) $l(tx, ty) = tl(x, y)$, for all $t, x, y \geq 0$ (often referred to as the *homogeneity* property),

(ii) $\max(x, y) \leq l(x, y) \leq x + y$, where the equality on the left hand side is attained when X_1 and Y_1 are completely positive dependent in the tail, and the equality on the right hand side is attained when X_1 and Y_1 are independent in the tail (often referred to as *asymptotic independence*).

3.1 Defining multivariate extreme quantiles for simultaneous thresholds

One of the main tasks in simultaneous monitoring of multiple measurements is to identify proper threshold points for which the exceedance probabilities are within certain predetermined values. For example, if we are to monitor a pair of measurements (X, Y), from F, our task would be to find threshold points x and y such that, for a predetermined value p,

$$P(X > x \text{ or } Y > y) = p. \tag{3.3}$$

Obviously, there exist infinitely many choices of (x, y) which satisfy the above condition. Different applications may also force additional constraints on the condition (3.3). One common constraint, which is also required in our applications in Section 4, is that

$$cP(X > x) = P(Y > y), \tag{3.4}$$

where the positive constant c indicates the different weights assigned to the two marginal tail probabilities. The value c can be chosen to reflect the different degrees of importance attached to the marginal variables, and is generally chosen in advance to address some particular practical concerns. For example, $c = 1$ implies that events of exceedance of either variable are viewed with equal importance. If c is chosen to be greater than 1 (which is the case in our application in Section 4), then the exceedance in Y is viewed as more important or more critical.

For a very small p, (3.2) implies that

$$\begin{aligned} p &= P(X > x \text{ or } Y > y) \\ &= 1 - F(x, y) \\ &= 1 - F_1(x) + 1 - F_2(y) - C(1 - F_1(x), 1 - F_2(y)) \\ &\approx l\,(1 - F_1(x), 1 - F_2(y)) \\ &= l(p_1, p_2) \end{aligned} \tag{3.5}$$

where $p_1 = 1 - F_1(x) = P(X > x)$, $p_2 = 1 - F_2(y) = P(Y > y)$. Since $cp_1 = p_2$, see (3.4),

$$p \approx l(p_1, cp_1) = p_1 l(1, c)\,,$$

so

$$p_1 \approx \frac{p}{l(1,c)}\,, \quad p_2 \approx \frac{cp}{l(1,c)}\,. \tag{3.6}$$

The discussion above shows that the estimation of a bivariate extreme quantile can be essentially decomposed into two parts, namely i) the estimation of the marginal quantiles, and ii) the estimation of $l(1, c)$. Part i) can be addressed in a similar fashion as in the univariate case discussed in Section 2, although we have to estimate p_1 and p_2 now. Part ii) is the remaining task and is addressed in the next subsection.

3.2 Estimating the Tail Dependence Function

Following the definition of $l(x,y)$ in (3.2), the empirical tail dependence function of F based on (X_i, Y_i), $i = 1, \ldots, n$, is proposed in Huang (1992), see also Einmahl, de Haan and Li (2006), and is defined as

$$\hat{l}_{n,k}(x,y) = k^{-1} \sum_{j=1}^{n} I_{\left[X_j \ge X_{n-[kx]+1:n} \text{ or } Y_j \ge Y_{n-[ky]+1:n}\right]}. \tag{3.7}$$

Note that this estimator of $l(x,y)$ again depends on the choice of k. It is shown to be consistent and asymptotically normal. As seen in the univariate extreme quantile estimation in Section 2, we shall find the optimal k by minimizing the mean squared error of $\hat{l}_{n,k}(x,y)$, i.e.,

$$MSE(n,k) = E\left(\hat{l}_{n,k}(x,y) - l(x,y)\right)^2. \tag{3.8}$$

Clearly, the optimal choice of k here depends on the unknown $l(x,y)$. Mimicking the idea in the univariate case, we may circumvent this difficulty by replacing (3.8) by an auxiliary statistic.

In Section 5.4 of Peng (1998) the following alternative estimator of $l(x,y)$ was introduced

$$\tilde{l}_{n,k}(x,y) = \hat{l}_{n,k}(2x, 2y) - \hat{l}_{n,k}(x,y). \tag{3.9}$$

Note that using the homogeneity property of l, it follows that $\tilde{l}_{n,k}(x,y)$ is a consistent estimator of $l(x,y)$. Now, replacing $MSE(n,k)$ in (3.8) by

$$E\left(\tilde{l}_{n,k}(x,y) - \hat{l}_{n,k}(x,y)\right)^2,$$

Peng (1998) then derived a double bootstrap procedure to find the optimal k for estimating $l(x,y)$. This bootstrap procedure is similar to the one we have presented in Section 2 for the estimation of extreme quantiles and is thus omitted here. We only mention that in step 6) we can obtain similarly

$$\hat{k}_0 = \frac{\hat{k}_1^2}{\hat{k}_2} g(\hat{\rho}),$$

where

$$g(\hat{\rho}) = \left(\frac{2(2^{1+\hat{\rho}} - 1)^2}{(2^{1+\hat{\rho}} - 2)^2}\right)^{-1/(2\hat{\rho}+1)}.$$

We now take

$$\hat{\rho} = -\frac{\log L_{n,k}}{\log 2},$$

where

$$L_{n,k} = \frac{2\hat{l}_{n,k}(\frac{1}{2}, \frac{1}{2}) - \hat{l}_{n,k}(1,1)}{\hat{l}_{n,k}(1,1) - \frac{1}{2}\hat{l}_{n,k}(2,2)}.$$

This estimator is derived following Fraga Alves, de Haan and Lin (2003), p. 156.

Finally, we are ready to describe the procedure for estimating the extreme quantile (x, y), such that $P(X > x, \text{or } Y > y) = p$ and $cp_1 = p_2$. The procedure is outlined as follows:

Step a) Obtain the estimate $\tilde{l}(1, c)$ (as given in (3.9)) for $l(1, c)$ by using the optimal k obtained from the above bootstrap procedure.

Step b) Following (3.6), estimate the marginal tail probabilities p_1 and p_2 by

$$\hat{p}_1 = \frac{p}{\tilde{l}(1,c)}, \qquad \hat{p}_2 = \frac{cp}{\tilde{l}(1,c)}.$$

Step c) Apply $\hat{p}_1$ and $\hat{p}_2$ to (2.8) to obtain the corresponding estimators for the marginal quantiles $\hat{x}_{\hat{p}_1}$ and $\hat{y}_{\hat{p}_2}$. Here the optimal k should be obtained from the bootstrap procedure given in Section 2. Finally, we propose the resulting $(\hat{x}_{\hat{p}_1}, \hat{y}_{\hat{p}_2})$ as an estimator for (x, y).

3.3 Consistency of the quantile estimators

Next, we show that the extreme quantile estimator $(\hat{x}_{\hat{p}_1}, \hat{y}_{\hat{p}_2})$ achieves the desired probability level and satisfies the constraint (3.4), asymptotically. Before we proceed with the bivariate case, we first prove some asymptotic results in the univariate case. Let $\bar{F} = 1 - F$, and $q_\gamma(x) = \int_1^x s^{\gamma-1} \log s \, ds$. Also, recall the definition of $\hat{x}_p$ in (2.8). Observe that our p should depend on n and tend to zero (if p would be fixed, many observations would exceed x_p when n is sufficiently large). Therefore the usual notion of consistency in terms of estimation difference is not appropriate here. We consider using the ratio instead.

Theorem 3.1. Assume that

(a) $np = O(1)$,

(b) $\frac{k}{n} \to 0$, $k \to \infty$,

(c) $q_\gamma(d_n)/(d_n^\gamma \sqrt{k}) \to 0$, with $d_n = \frac{k}{np}$ (hence $\gamma > -\frac{1}{2}$),

(d) F satisfies the following second order refinement of the domain of attraction condition: there exists a function A with $\lim_{t\to\infty} A(t) = 0$ and constant sign near infinity, such that

$$\lim_{t\to\infty} \frac{t\bar{F}(a_t x + b_t) - (1+\gamma x)^{-1/\gamma}}{A(t)} = (1+\gamma x)^{-1-1/\gamma} H_{\gamma,\rho}\left((1+\gamma x)^{-1/\gamma}\right),$$

for all x with $1 + \gamma x > 0$ and some $\rho < 0$, where

$$H_{\gamma,\rho}(x) = \frac{1}{\rho}\left(\frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma}\right),$$

(e) $A_n = \sqrt{k}(\frac{\hat{a}_{n/k}}{a_{n/k}} - 1) = O_p(1)$, $B_n = \sqrt{k}(\frac{X_{n-k:n} - b_{n/k}}{a_{n/k}}) = O_p(1)$, and $\Gamma_n = \sqrt{k}(\hat{\gamma_n} - \gamma) = O_p(1)$.

Then we have

$$\frac{\bar{F}(\hat{x}_p)}{p} \xrightarrow{p} 1. \tag{3.10}$$

The proof of Theorem 3.1 is given in the appendix.

Remark 3.1. In fact, Theorem 3.1 holds for any estimators of $a_{n/k}, b_{n/k}$ and γ as long as the $O_p(1)$ requirements in (e) are fulfilled.

Remark 3.2. If $\hat{x}_{\hat{p}}$ is calculated from (2.8) based on a random $\hat{p}$, such that $\hat{p}/p \xrightarrow{p} c_0$ holds for some $c_0 \in (0, \infty)$, then under our assumptions on p it can also be shown easily that

$$\frac{\bar{F}(\hat{x}_{\hat{p}})}{\hat{p}} \xrightarrow{p} 1.$$

To proceed with the bivariate case, we define $Q_p = (-\infty, x] \times (-\infty, y]$ such that $F(x, y) = 1 - p$ and $c\,(1 - F_1(x)) = 1 - F_2(y)$ for some predetermined value $c \in (0, \infty)$. Let $\hat{Q}_p$ denote the estimator given by the aforementioned procedure, i.e.,

$$\hat{Q}_p = (-\infty, \hat{x}_{\hat{p}_1}] \times (-\infty, \hat{y}_{\hat{p}_2}]$$

where $\hat{x}_{\hat{p}_1}$ and $\hat{y}_{\hat{p}_2}$ are the $1-\hat{p}_1$-th and $1-\hat{p}_2$-th quantile estimators of F_1 and F_2, respectively, with $\hat{p}_1 = \frac{p}{\hat{l}(1,c)}$ and $\hat{p}_2 = c\hat{p}_1 = \frac{cp}{\hat{l}(1,c)}$. Observe that for estimating $l(1,c)$ we have to choose a k, and that for estimating $x_{\hat{p}_1}$ and $y_{\hat{p}_2}$ we have to choose k_1 and k_2, say. Our main theoretical result is stated in the theorem below. A related result can be found in de Haan and Huang (1995).

Theorem 3.2. Assume $np = O(1)$, $\frac{k}{n}, \frac{k_1}{n}, \frac{k_2}{n} \to 0$, $k, k_1, k_2 \to \infty$. Also assume that F is in the domain of attraction of a bivariate extreme distribution, and that both of the marginal distributions, F_1 and F_2, satisfy the conditions (c)-(e) listed in Theorem 3.1. Then we have

$$\frac{P\left(\hat{Q}_p \triangle Q_p\right)}{p} \xrightarrow{p} 0,$$

where $\triangle$ denotes the symmetric difference.

The proof is also given in the appendix.

Remark 3.3. Theorem 3.2 immediately implies

$$\frac{1 - F\left(\hat{x}_{\hat{p}_1}, \hat{y}_{\hat{p}_2}\right)}{p} \xrightarrow{p} 1.$$

4 Application: Simultaneous thresholding two risk indicators

As an illustrative example for the utility of extreme quantiles, we now apply the threshold system derived in Section 3 to assist the FAA project on simultaneous monitoring of multiple aviation risk indicators. One of the main responsibilities of the FAA is to monitor and regulate all air carriers in terms of aviation safety. The FAA regularly conducts surveillance inspections on all air carriers and inspections findings are carefully analyzed and monitored. To increase the efficiency of the monitoring scheme, the FAA hopes to embed a threshold system in the monitoring scheme, which can assign inspections findings with proper indications of their levels of risk. Specifically, the regions (or ranges) corresponding to the different risk levels are termed

- *informational* (colored green)
- *expected* (colored blue)

- *advisory* (colored yellow)
- *concern* (colored red)

and they indicate increasing levels of risk. Following the procedure described in Section 3, we can determine thresholds that correspond to given exceedance probabilities for any risk indicators of interest. This threshold system can provide a concrete measure of the inspection results in terms of the severity of potential flaws and serve as a guideline for the general rating of the safety performance of each carrier.

Our application concerns the monitoring of two risk indicators. Specifically, they are air carrier performance measures: *incident rate* (IR) and *operational unfavorable ratio* (OU). In the aviation industry, the OU is perceived as "twice as important" as the IR. The data set is collected by the FAA from 10 air carriers of similar service type and fleet size over a period of 57 months, from July 1993 to March 1998, see the scatter plot in Figure 1. Each data point represents a monthly observation of (IR, OU) from a given carrier. Since both IR and OU are measures of non-conformance, the higher the values the more severe the potential flaws.

The purpose of the FAA project is to identify the region which contains the worst 0.15% of all possible performances and label it as the region of *concern*. This implies that the *concern* region corresponds to the sample space which is beyond the joint upper 0.0015-th quantile. The region would be labeled as *advisory* if contains the worst 1% of all possible performances which are not yet the worst 0.15%. Thus, the *advisory* region corresponds to the sample space which is beyond the joint upper 0.01-th quantile, but below the joint upper 0.0015-th quantile. The region would be labeled as *informational* if it contains the best 5% of all possible performances. The remaining region on the sample space would then be labeled as *expected* and it contains observations which are viewed as having met the FAA expectation under normal circumstances. There are 570 data points in total. Note that $570 \times 0.0015 = 0.855$, which implies that on average there is less than one observations in the *concern* region. Also $570 \times 0.01 = 5.7$ which is quite small. This setting of very few or no occurrence is ideal for the application of multivariate statistics of extremes. We discuss first the construction of the *concern* and *advisory* regions. The *informational* region (and hence also the *expected* region) will be constructed at the end of this section using empirical

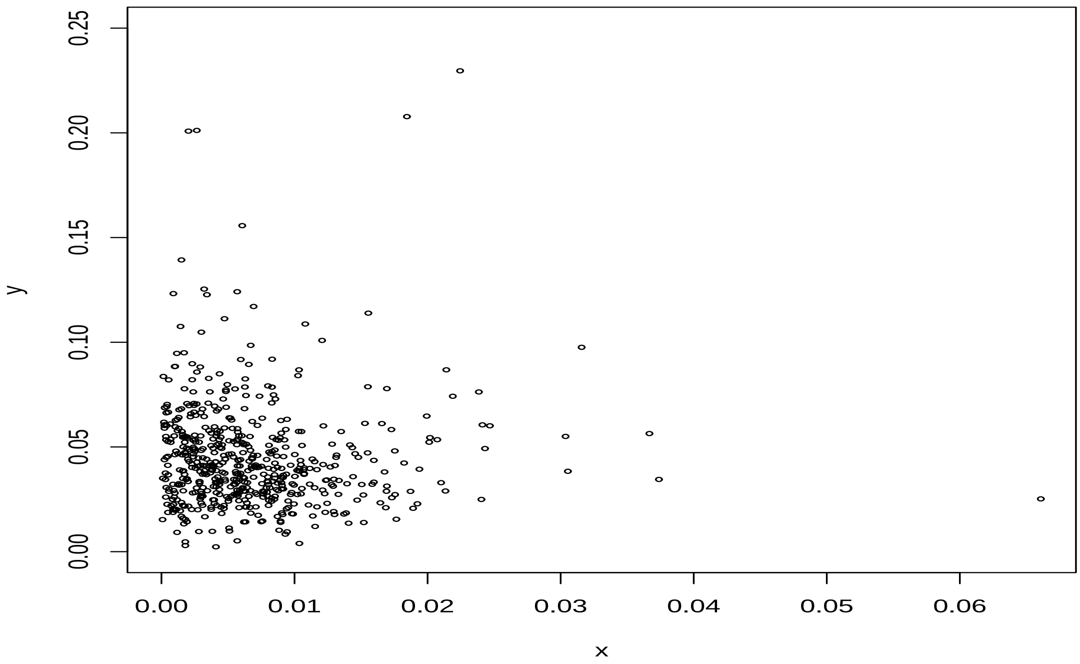


Figure 1: Scatter plot for air carrier performance measures.

process theory, since on average relatively many ($570 \times 0.05 = 28.5$) observations fall in that region.

For different goals and interpretations for different applications, there can be different definitions of multivariate quantiles. In the context of the FAA application, since higher observed value implies worse performance, we interpret a multivariate observation as flawed or at risk if any of its component measurements exceeds a certain threshold. This leads us to consider in Section 3 the quantiles as regions of the form $(-\infty, x] \times (-\infty, y]$ (quadrants) such that $F(x, y) = 1 - p$. The constraint that OU is "twice as important" as IR can be translated into the expression

$$2\,(1 - F_1(x)) = 1 - F_2(y) \quad \text{i.e.} \quad 2p_1 = p_2$$

if we denote IR as X and OU as Y. Altogether, our task now amounts to finding x and y which can satisfy the conditions $1 - F(x, y) = p$ and $2\,(1 - F_1(x)) = 1 - F_2(y)$, for $p = 0.0015$ and 0.01 (respectively for *concern* region and *advisory* region). This setting fits exactly the framework discussed in Section 3 with $c = 2$.

Before we proceed with the procedure given in Section 3 to solve the problem above, we first need to verify from the data set that the assumption for bivariate extreme value theory hold. In other words, we need to check if F is in the domain of attraction of an extreme value distribution. To this end, we have applied the test proposed in Einmahl, de Haan and Li (2006) to our data set and failed to reject the null hypothesis that F is in the domain of attraction of an extreme value distribution. Therefore, we can safely move to the next step to apply the procedure in Section 3 to our data set. For illustration purpose, we show the step-by-step results only for $p = 0.0015$. The same procedure applies to $p = 0.01$ and is omitted. The final estimates of the quantile for $p = 0.01$ will be mentioned later.

We first begin with estimating the tail dependence function $l(1, 2)$. To obtain the optimal k for the estimation of $l(1, 2)$, we carry out the bootstrap procedure listed in Section 3 with $n_1 = n^{0.95}$ and $B = 10000$. In order to avoid the few non-convergence situations, we choose to use a multi-stage bootstrap procedure for which we bootstrap $m = 200$ times for each of $r = 50$ replications. With this multi-stage bootstrap, we obtain 50 sets of k_1 and k_2. It can be shown that the bootstrap works well only if k_1 and k_2 can satisfy $k_2 \leq k_1 \leq \frac{n_1}{n_2} k_2$. As it turns out, there are only 37 pairs (out of 50 pairs) of such (k_1, k_2). To obtain the optimal k_0, we need to estimate ρ, and then examine the plot of $\hat{\rho}$ vs. k. This plot is given in Figure 2. In the plot, the horizontal line corresponds to $\hat{\rho} = 1.635$

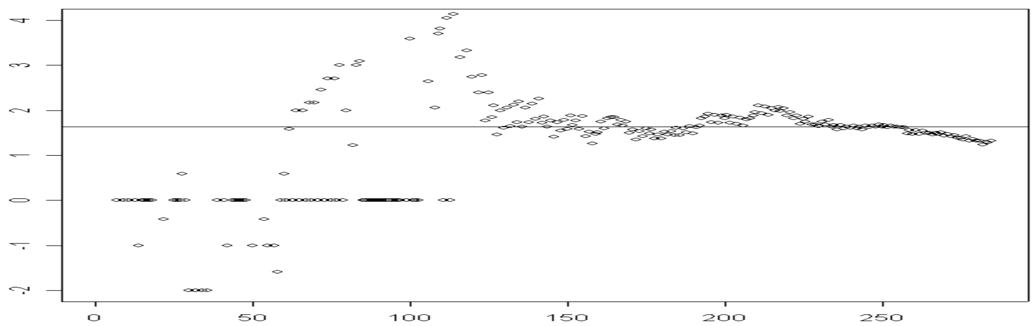


Figure 2: $\hat{\rho}$ vs. k.

which is our choice of the estimate of ρ. Using this $\hat{\rho}$ and those 37 pairs of k_1 and k_2, we can obtain 37 estimates of the optimal k_0. Plugging these k_0's in (3.9) and (3.7) leads to

37 estimates for $\tilde{l}(1,2)$, which yield a mean 2.702. Figure 3 shows the plot of $\tilde{l}(1,2)$ vs. k with the horizontal line at 2.702. The estimation based on these optimal k_0's (which are derived from the bootstrap procedure) appears to be reasonably satisfactory.

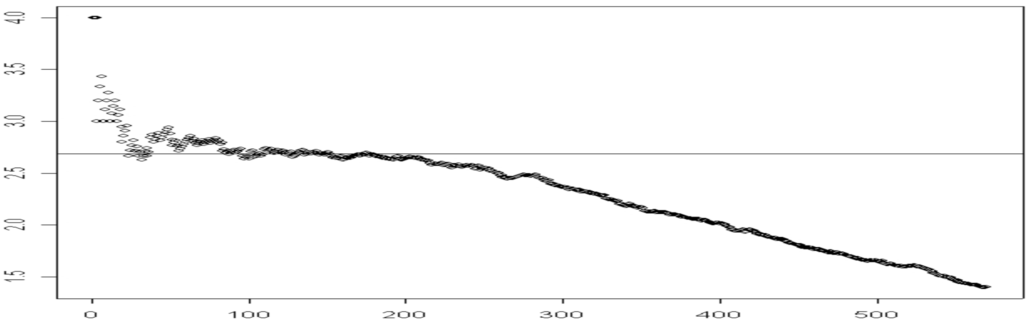


Figure 3: $\tilde{l}(1,2)$ vs. k.

We can now plug the estimate $\tilde{l}(1,2)$ into (3.6) to obtain the estimates for the two marginal tail probabilities, which turn out to be $\hat{p}_1 = 0.00056(= 0.0015/2.702)$ and $\hat{p}_2 = 0.00111(= 2\hat{p}_1)$. Once these two tail probabilities are determined, we can simply follow the procedure described in Section 2 for estimating a univariate extreme quantile to obtain the joint upper 0.0015-th quantile. To begin with, we need to check if the two marginal distributions are in the univariate domain of attraction of an extreme value distribution. Here we have used the tests devised in Dietrich, de Haan and Hüsler (2002) and Drees, de Haan and Li (2006). The test results turn out to be affirmative. Figure 4 shows the estimated γ plots for both X and Y. The plots in Figure 4 clearly show that both of the marginal distributions have positive γ and thus have heavy tails.

For each marginal, we proceed using the (multi-stage) bootstrap procedure in Section 2 to determine the optimal k for the quantile estimate, with again $n_1 = n^{0.95}$ and $B = 10000$ ($m = 200$ and $r = 50$). In each replication, we essentially need to find k_i which minimizes $\widehat{MSE}^*(n_i,k) = \frac{1}{m}\sum_{j=1}^m \left(\hat{x}^{*(j)}_{n_i,1}(k) - \hat{x}^{*(j)}_{n_i,2}(k)\right)^2$, $i = 1,2$. Figure 5 is the plot of $\widehat{MSE}^*(n_i,k)$ vs. k from one replication based on OU for $i = 1,2$. Both plots above show that $\widehat{MSE}^*$ achieves its global minimum at either end of the range of k. This means k_i is either very small or very large (close to n_i). Since neither could be a practical estimate, we add some constraints to the range of the possible k_i, $i = 1,2$ by focusing only k_i's that yield a local

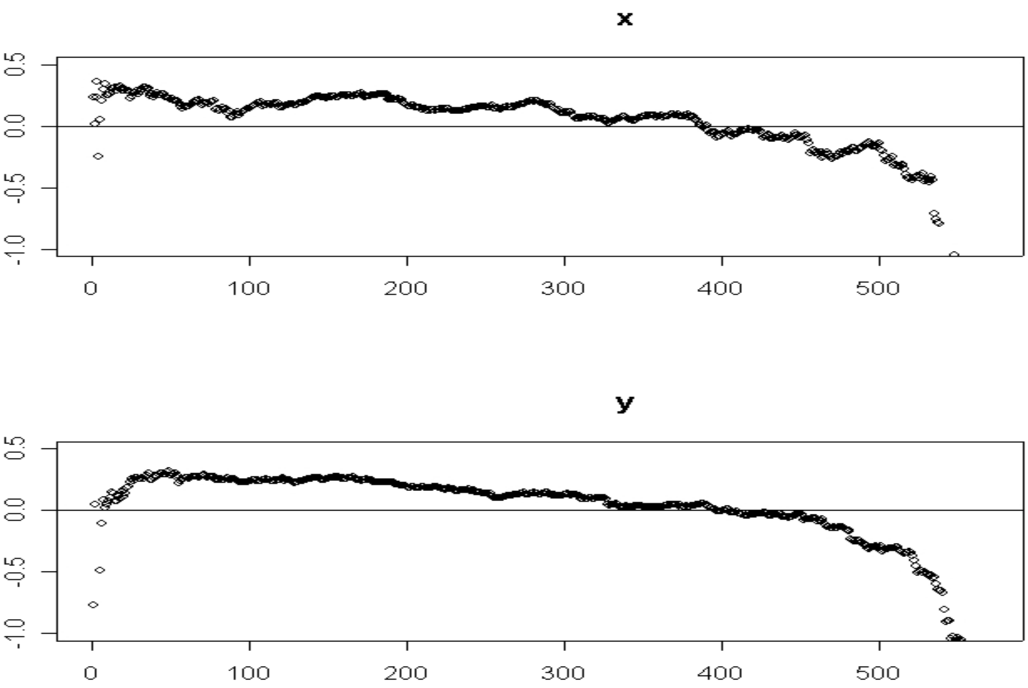
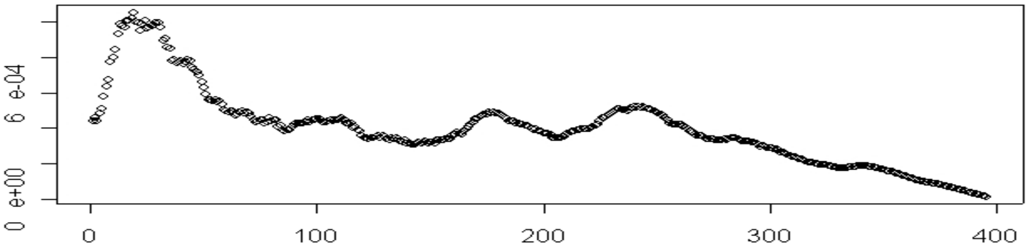


Figure 4: $\hat{\gamma}$ vs. k.

minimum instead. With this consideration, we observe that there exist local minimums in the range (20 to 240) and (10 to 175) respectively in the upper and lower panels of Figure 5. Thus, we may try to find the optimal k_i's, in those intervals to achieve the local minimum. The same procedure will be applied to obtain the optimal k_i, $i = 1, 2$, for IR. Among the resulting 50 pairs of k_1's and k_2's, only 22 pairs for IR and 32 pairs for OU remain after imposing the constraint $k_2 \leq k_1 \leq \frac{n_1}{n_2} k_2$. Those remaining pairs of k_1 and k_2 can then be used in (2.14) to obtain the estimate for k_0. To obtain the estimates for k_0 from the expression (2.14), we only need to find $\hat{\rho}$. Based on (2.15), the plots of $\hat{\rho}$ vs. k for both IR and OU are given in Figure 6, The horizontal lines in those plots represent our $\hat{\rho}$, which are -0.356 for IR and -0.456 for OU respectively. Plugging these values along with the remaining k_i, $i = 1, 2$, into (2.14) and then (2.8), we can obtain the mean of the estimates of the marginal quantiles, which are 0.064 for IR, 0.252 for OU. Therefore, our estimated upper joint 0.0015-th quantile is (0.064, 0.252). Figure 7 shows the plot of the estimated quantiles vs. k with the horizontal lines for the final estimation. From these plots, we can



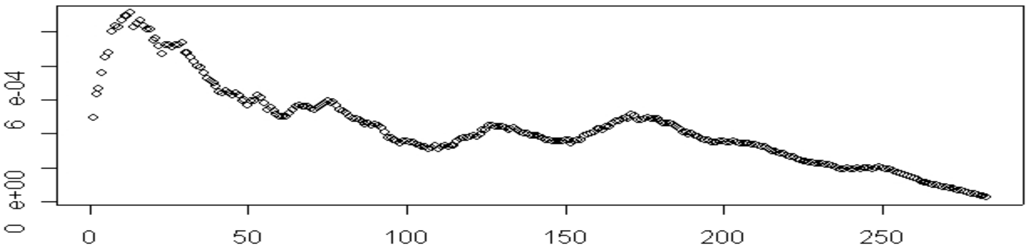


Figure 5: $\widehat{MSE}^*(n_i, k)$ vs. k. (The upper and lower plots respectively are with bootstrap sample sizes n_1 and n_2.)

see that the optimal k_0 obtained from the bootstrap procedure works well for both IR and OU.

Following the same procedure, we also obtain the estimated upper joint 0.01-th quantile (0.036, 0.150).

Finally, we discuss the *informational* region. Obviously, both components in this region should assume low values, namely a region of the form $(-\infty, x] \times (-\infty, y]$, with $F(x, y) = \tilde{p}$. Let Q_1 and Q_2 denote the left-continuous quantile functions corresponding to F_1 and F_2 respectively. The constraint that OU is "twice as important" as IR is then translated into $x = Q_1(2t), y = Q_2(t)$ for some $t \in (0, 1)$. For a given $\tilde{p}$, let t_0 be a t which satisfies the above conditions. We can now define the estimated *informational* region by

$$\left(-\infty, \frac{1}{2}(X_{2n\hat{t}:n} + X_{2n\hat{t}+1:n})\right] \times \left(-\infty, \frac{1}{2}(Y_{n\hat{t}:n} + Y_{n\hat{t}+1:n})\right],$$

where $\hat{t}$ is the smallest t such that nt is an integer and $\sum_{i=1}^{n} I_{[X_i \le X_{2nt:n}; Y_i \le Y_{nt:n}]} \ge n\tilde{p}$. Applying this with $\tilde{p} = 0.05$ to our data, the values correspond to IR and OU are (0.0032,

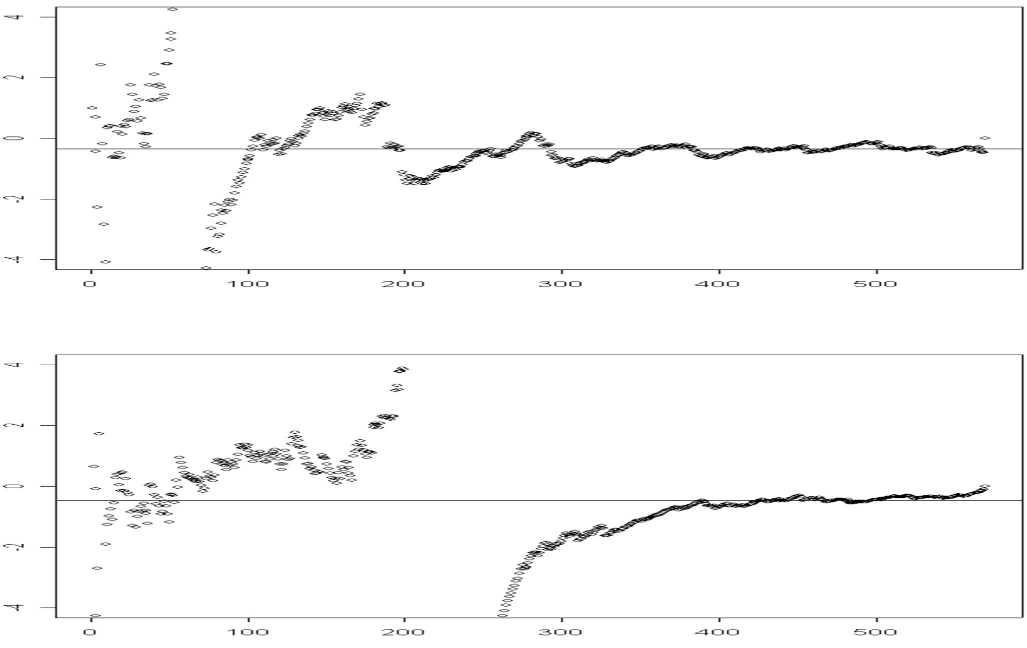


Figure 6: $\hat{\rho}$ vs. k for IR and OU.

0.0238).

We briefly show the consistency of this procedure, i.e.

$$F\left(\frac{1}{2}(X_{2n\hat{t}:n} + X_{2n\hat{t}+1:n}), \frac{1}{2}(Y_{n\hat{t}:n} + Y_{n\hat{t}+1:n})\right) \xrightarrow{p} \tilde{p}.$$

Note that, contrary to the extreme value approach, we now assume that the given probability $\tilde{p}$ is fixed. For the consistency, it suffices to show that $F(X_{2n\hat{t}:n}, Y_{n\hat{t}:n}) \xrightarrow{p} \tilde{p}$. Write $U_i = F_1(X_i), V_i = F_2(Y_i), i = 1, \ldots, n$, and denote the order statistics of the U_i and V_i in the usual way. Moreover, let $\tilde{C}$ denote the distribution function of the pairs (U_i, V_i), and let Q_{1n} and Q_{2n} be the empirical quantile functions of the U_i and V_i, respectively. We have then

$$\begin{aligned} & |F(X_{2n\hat{t}:n}, Y_{n\hat{t}:n}) - \tilde{p}| = |\tilde{C}(U_{2n\hat{t}:n}, V_{n\hat{t}:n}) - \tilde{p}| \\ = \;& |\tilde{C}(U_{2n\hat{t}:n}, V_{n\hat{t}:n}) - \tilde{C}(2t_0, t_0)| = |\tilde{C}(Q_{1n}(2\hat{t}), Q_{2n}(\hat{t})) - \tilde{C}(2t_0, t_0)| \\ \leq \;& |\tilde{C}(Q_{1n}(2\hat{t}), Q_{2n}(\hat{t})) - \tilde{C}(2\hat{t}, \hat{t})| + |\tilde{C}(2\hat{t}, \hat{t}) - \tilde{C}(2t_0, t_0)| \\ \leq \;& |Q_{1n}(2\hat{t}) - 2\hat{t}| + |Q_{2n}(\hat{t}) - \hat{t}| + |\tilde{C}(2\hat{t}, \hat{t}) - \tilde{C}(2t_0, t_0)|, \qquad (4.1) \end{aligned}$$

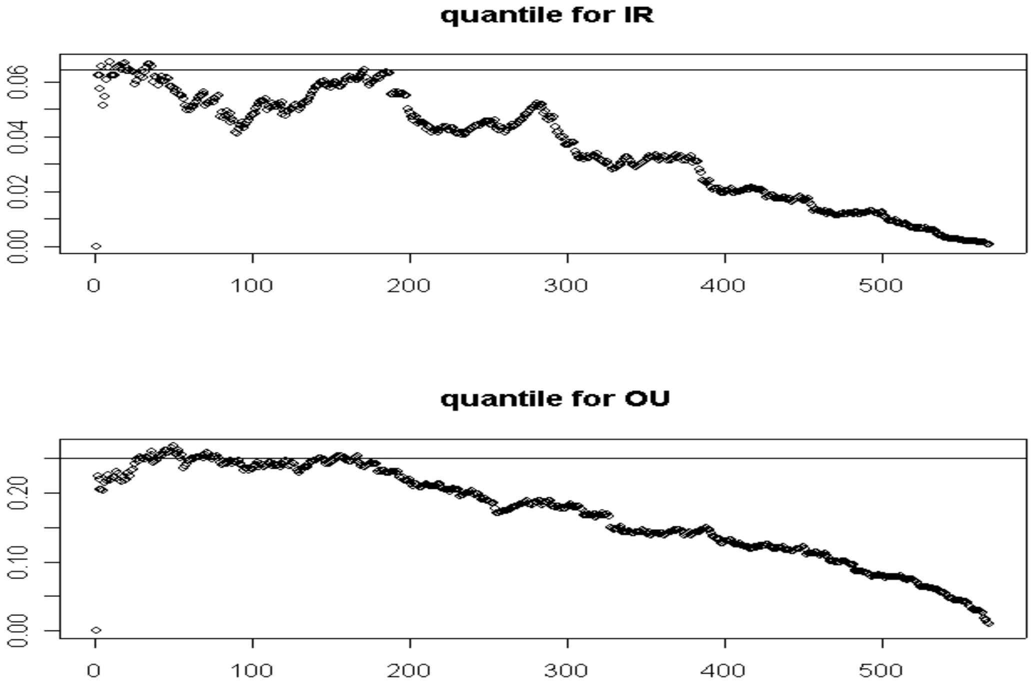


Figure 7: The extreme quantile estimates for IR and OU vs. k.

where the last inequality follows because $\tilde{C}$ has uniform-(0,1) marginals. The first two terms tend to zero in probability because of the Glivenko-Cantelli theorem for the uniform quantile process. Set $\tilde{C}_n(u, v) = H_n(\bar{Q}_{1n}(u), \bar{Q}_{2n}(v))$, where H_n is the bivariate empirical distribution function of the (U_i, V_i). Now from the uniform consistency of $\tilde{C}_n$ as an estimator of $\tilde{C}$ and the definition of $\hat{t}$, it can be shown that the third term in the right hand side of (4.1) also tends to zero in probability, which renders the proof complete.

Finally, the estimated threshold regions are shown in Figure 8, where for better viewing both coordinates are presented in the log scale. The upper right region corresponds to *concern* and should be colored red, the next upper region corresponds to *advisory* and should be yellow. The lower rectangle is the green *informational* region and the blue *expected* region is between green and yellow.

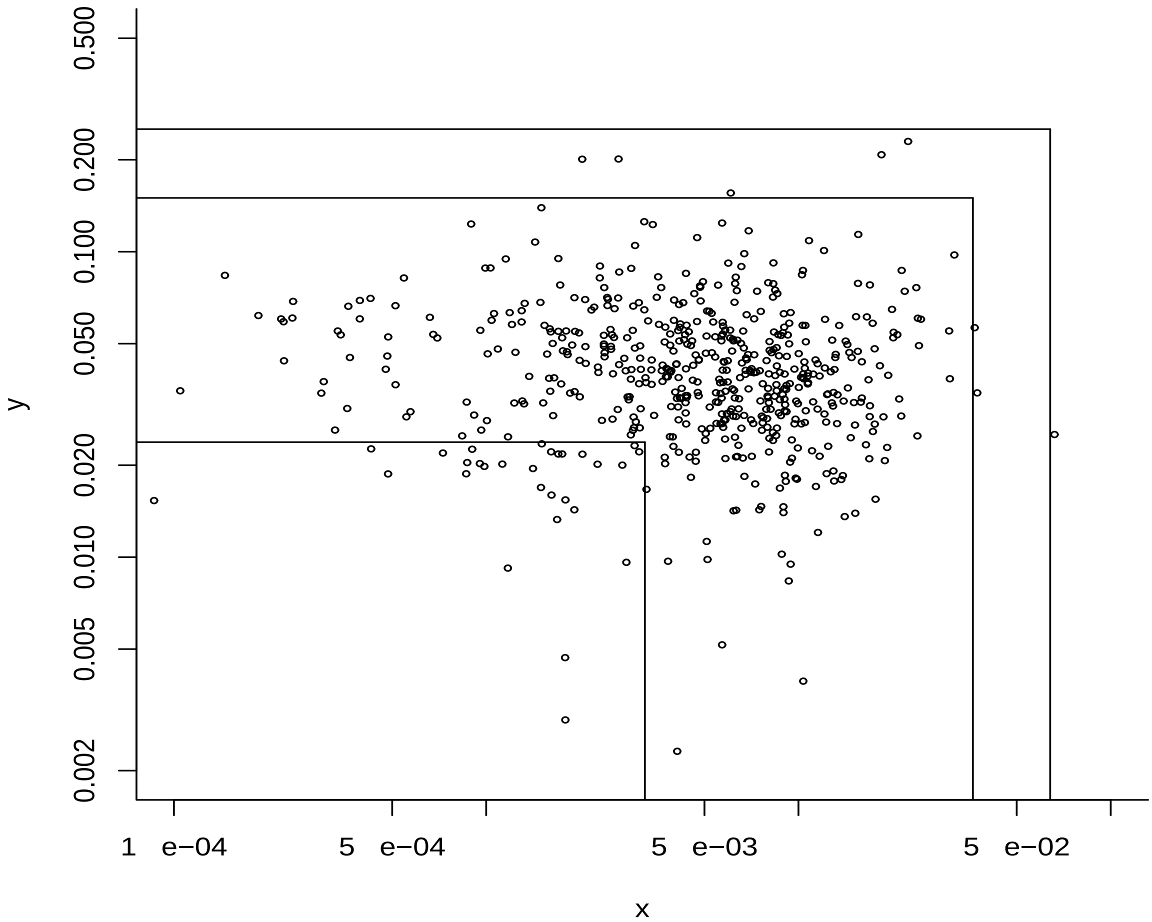


Figure 8: Threshold system: the four designated regions *w.r.t.* the scatter plot.

5 Appendix

Proof of Theorem 3.1. We briefly write $\hat{a} = \hat{a}_{n/k}, \hat{b} = \hat{b}_{n/k}, a = a_{n/k}, b = b_{n/k}$ and $\hat{\gamma} = \hat{\gamma}_n$. Observe that $d_n \to \infty$. We first show that for $-\frac{1}{2} < \gamma < 0$,

$$P\left(\frac{\hat{b} - b}{a} + \frac{\hat{a}}{a}\frac{d_n^{\hat{\gamma}} - 1}{\hat{\gamma}} < -\frac{1}{\gamma}\right) \to 0 \quad (n \to \infty). \tag{5.2}$$

Note that, the condition (e) implies

$$1 + \frac{\hat{\gamma}\left(-\frac{1}{\gamma} - \frac{B_n}{\sqrt{k}}\right)}{1 + \frac{A_n}{\sqrt{k}}} = O_p\left(\frac{1}{\sqrt{k}}\right). \tag{5.3}$$

Applying condition (c) twice and also (e), we obtain

$$d_n^{\hat{\gamma}}\sqrt{k} = d_n^{\gamma}\sqrt{k}e^{\frac{\log d_n}{\sqrt{k}}\cdot\Gamma_n}, \tag{5.4}$$

which tends to infinity in probability. Combination of (5.3) and (5.4) easily yields (5.2).

From Lemma 2.4.1 in Li (2004) in conjunction with (5.2), we obtain for all $\gamma > -\frac{1}{2}$

$$\frac{\bar{F}(\hat{x}_p)}{p} = d_n\left(1+\gamma\left(\frac{\hat{b}-b}{a}+\frac{\hat{a}}{a}\frac{d_n^{\hat{\gamma}}-1}{\hat{\gamma}}\right)\right)^{-\frac{1}{\gamma}}(1+o_p(1)).$$

Therefore it remains to show that

$$d_n\left(1+\gamma\left(\frac{\hat{b}-b}{a}+\frac{\hat{a}}{a}\frac{d_n^{\hat{\gamma}}-1}{\hat{\gamma}}\right)\right)^{-\frac{1}{\gamma}} \xrightarrow{p} 1. \tag{5.5}$$

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

$$q_\gamma(x) \sim \begin{cases} \frac{1}{\gamma}x^\gamma \log x & \gamma > 0 \\ \frac{1}{2}(\log x)^2 & \gamma = 0 \\ \frac{1}{\gamma^2} & \gamma < 0 \end{cases}$$

Observe that condition (c) is equivalent to

$$\begin{cases} \frac{\log d_n}{\sqrt{k}} \to 0, & \text{for } \gamma > 0, \\ \frac{(\log d_n)^2}{\sqrt{k}} \to 0, & \text{for } \gamma = 0, \\ d_n^\gamma\sqrt{k} \to \infty, & \text{for } \gamma < 0. \end{cases} \tag{5.6}$$

Next we prove (5.5); part of the proof is similar to that of Proposition 8.2.9 in de Haan and Ferreira (2006). First, we take $\gamma \neq 0$:

$$\begin{aligned} & d_n\left(1+\gamma\left(\frac{\hat{b}-b}{a}+\frac{\hat{a}}{a}\frac{d_n^{\hat{\gamma}}-1}{\hat{\gamma}}\right)\right)^{-\frac{1}{\gamma}} \\ = \; & d_n\left(1+\gamma\frac{B_n}{\sqrt{k}}+\left(1+\frac{A_n}{\sqrt{k}}\right)\left(1-\frac{\Gamma_n}{\sqrt{k}\hat{\gamma}}\right)\left(d_n^{\hat{\gamma}}-1\right)\right)^{-\frac{1}{\gamma}} \\ =: \; & d_n\left(\left(1+\frac{D_n}{\sqrt{k}}\right)d_n^{\hat{\gamma}}+\frac{E_n}{\sqrt{k}}\right)^{-\frac{1}{\gamma}} \\ = \; & d_n^{1-\frac{\hat{\gamma}}{\gamma}}\left(1+\frac{D_n}{\sqrt{k}}+\frac{E_n}{d_n^{\hat{\gamma}}\sqrt{k}}\right)^{-\frac{1}{\gamma}} =: s_n. \end{aligned}$$

Note that D_n and E_n are $O_p(1)$ due to (e), and also that $d_n^{\hat{\gamma}}\sqrt{k} \to \infty$ (for $\gamma < 0$ see (5.4)). Hence, we have

$$\left(1 + \frac{D_n}{\sqrt{k}} + \frac{E_n}{d_n^{\hat{\gamma}}\sqrt{k}}\right)^{-\frac{1}{\gamma}} \xrightarrow{p} 1.$$

We also have

$$d_n^{1-\frac{\hat{\gamma}}{\gamma}} = e^{-\frac{\log d_n}{\sqrt{k}} \cdot \frac{\Gamma_n}{\gamma}} \xrightarrow{p} 1.$$

Consequently, $s_n \xrightarrow{p} 1$, and thus (5.5) holds for $\gamma \neq 0$.

For $\gamma = 0$, the proof of (5.5) goes as follows. By definition $(1 + \gamma x)^{-1/\gamma} = e^{-x}$ in this case. Note then

$$\begin{aligned} & d_n \exp\left(-\frac{\hat{b} - b}{a} - \frac{\hat{a}}{a}\frac{d_n^{\hat{\gamma}} - 1}{\hat{\gamma}}\right) \\ = \ & \exp\left(-\frac{B_n}{\sqrt{k}} - \left(1 + \frac{A_n}{\sqrt{k}}\right)\left(\frac{d_n^{\hat{\gamma}} - 1}{\hat{\gamma}} - \log d_n\right) - \frac{A_n}{\sqrt{k}} \log d_n\right), \end{aligned}$$

Clearly, if

$$\frac{d_n^{\hat{\gamma}} - 1}{\hat{\gamma}} - \log d_n \xrightarrow{p} 0, \tag{5.7}$$

holds, then (5.5) follows immediately from (e). To show (5.7), we observe that

$$\begin{aligned} & \left|\frac{d_n^{\hat{\gamma}} - 1}{\hat{\gamma}} - \log d_n\right| = \left|\int_1^{d_n} \frac{s^{\hat{\gamma}} - 1}{s} ds\right| = \left|\hat{\gamma} \int_1^{d_n} \int_1^s u^{\hat{\gamma}} \frac{1}{u} du \frac{1}{s} ds\right| \\ \leq \ & |\hat{\gamma}| \, d_n^{|\hat{\gamma}|} \log^2 d_n = \left|\hat{\gamma}\sqrt{k}\right| \frac{\log^2 d_n}{\sqrt{k}} e^{\frac{\log d_n}{\sqrt{k}} \left|\hat{\gamma}\sqrt{k}\right|}. \end{aligned}$$

The last expression tends to zero in probability under the conditions (c) and (e). This completes the proof.

Proof of Theorem 3.2. Note that

$$\frac{1}{p} P\left(\hat{Q}_p \triangle Q_p\right) \leq \frac{\left|\bar{F}_1(\hat{x}_{\hat{p}_1}) - \bar{F}_1(x_{p_1})\right|}{p} + \frac{\left|\bar{F}_2(\hat{y}_{\hat{p}_2}) - \bar{F}_2(y_{p_2})\right|}{p}.$$

We consider only the first term on the right, since the second can be handled similarly:

$$\begin{aligned} & \frac{1}{p}\left|\bar{F}_1(\hat{x}_{\hat{p}_1}) - \bar{F}_1(x_{p_1})\right| \\ \leq \ & \frac{1}{p}\left|\bar{F}_1(\hat{x}_{\hat{p}_1}) - \bar{F}_1(x_{\hat{p}_1})\right| + \frac{1}{p}\left|\bar{F}_1(x_{\hat{p}_1}) - \bar{F}_1(x_{p_1})\right|. \end{aligned} \tag{5.8}$$

Note that the second term of (5.8) can be dealt with as follows:

$$
\begin{aligned}
& \frac{1}{p}\left|\bar{F}_1(x_{\hat{p}_1}) - \bar{F}_1(x_{p_1})\right| = \frac{1}{p}\left|\hat{p}_1 - p_1\right| \\
\leq \quad & \frac{1}{p}\left|\frac{p}{\tilde{l}(1,c)} - \frac{p}{l(1,c)}\right| + \frac{1}{p}\left|\frac{p}{l(1,c)} - p_1\right| \\
= \quad & \frac{\left|\tilde{l}(1,c) - l(1,c)\right|}{l(1,c)\tilde{l}(1,c)} + \left|\frac{1}{l(1,c)} - \frac{p_1}{p}\right|.
\end{aligned} \tag{5.9}
$$

Following Theorem 1 of Chapter 2 (consistency of $\hat{l}(x,y)$) in Huang (1992) and the homogeneity of l listed in Section 3, the first term of (5.9) tends to zero in probability. Since F is in the bivariate domain of attraction, by the argument given in Section 3.1, the second term in (5.9) also tends to zero in probability.

Thus, it remains to show that

$$\frac{1}{p}\left|\bar{F}_1(\hat{x}_{\hat{p}_1}) - \bar{F}_1(x_{\hat{p}_1})\right| \xrightarrow{p} 0.$$

Note that since $p = \hat{p}_1\tilde{l}(1,c)$ and $\tilde{l}(1,c) \xrightarrow{p} l(1,c)$, it suffices to show

$$\frac{1}{\hat{p}_1}\left|\bar{F}_1(\hat{x}_{\hat{p}_1}) - \bar{F}_1(x_{\hat{p}_1})\right| \xrightarrow{p} 0.$$

This clearly follows from the result in univariate case and Remark 3.2, since

$$\frac{\hat{p}_1}{p} = \frac{1}{\tilde{l}(1,c)} \xrightarrow{p} \frac{1}{l(1,c)} \in (0,\infty).$$

This completes the proof.

Acknowledgements We are grateful to Ana Ferreira and Laurens de Haan for various helpful comments, in particular concerning the bootstrap procedures. We also like to thank Deyuan Li for performing the "domain of attraction checks".

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of extremes, theory and applications.* Wiley, New York.

[2] Coles, S.G. (2001). *An introduction to statistical modelling of extreme values.* Springer, New York.

[3] Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, **17**, 1833-1855.

[4] Dietrich, D., de Haan, L. and Hüsler, J. (2002). Testing extreme value conditions. *Extremes*, **5**, 71-85.

[5] Drees, H., de Haan, L. and Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *J. Statist. Plann. Inference*, **136**, 3498-3538.

[6] Einmahl, J.H.J., de Haan, L. and Li, D. (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Ann. Statist.*, **34**.

[7] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling extremal events, for insurance and finance.* Springer, New York.

[8] Ferreira, A., de Haan, L. and Peng, L. (2003). On optimizing the estimation of high quantiles of a probability distribution. *Statistics*, **37**, 401-434.

[9] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, **24**, 180-190.

[10] Fraga Alves, M.I., de Haan, L. and Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory. *Math. Methods of Stat.*, **12**, 155-176.

[11] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, **44**, 423-453.

[12] de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction.* Springer, New York.

[13] de Haan, L. and Huang, X. (1995). Large quantile estimation in a multivariate setting. *J. Multivariate Anal.*, **53**, 247-263.

[14] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163-1174.

[15] Huang, X. (1992). Statistics of bivariate extreme values. Ph.D. thesis 22, Tinbergen Institute, Erasmus University, Rotterdam.

[16] Li, D. (2004). *On extreme value approximation to tails of distribution functions.* Ph.D. thesis 342, Tinbergen Institute, Erasmus University, Rotterdam.

[17] Peng, L. (1998). *Second order condition and extreme value theory.* Ph.D. thesis 178, Tinbergen Institute, Erasmus University, Rotterdam.