# Thresholding Events of Extreme in Simultaneous Monitoring of Multiple Risks

John H. J. EINMAHL, Jun LI, and Regina Y. LIU

This article develops a threshold system for monitoring airline performance. This threshold system divides the sample space into regions with increasing levels of risk and allows instant assessments of risk level of any observed airline performance. Of particular concern is the performance with extreme risk. In this article, a multivariate extreme value theory approach is used to establish thresholds for signaling varying levels of extremeness in the context of simultaneous monitoring of multiple risk measures. The threshold system is justified in terms of multivariate extreme quantiles, and its sample estimator is shown to be consistent. This threshold system applies to general extreme risk management. Finally, a simulation and comparison study demonstrates the good performance of the proposed multivariate extreme quantile estimator. Supplemental materials providing technical details are available online.

KEY WORDS: Aviation safety; Extreme quantile; Multiple risks; Multivariate extreme quantile; Multivariate extreme value theory; Statistics of extremes; Threshold system.

## 1. INTRODUCTION

The Federal Aviation Administration (FAA) is responsible for monitoring and regulating aviation safety. The rapid increase in air traffic volume, coupled with aging aircraft, has led the FAA to increase surveillance and at the same time search for more effective analyses of its massive streaming surveillance data. A particular need is a monitoring scheme equipped with an effective threshold system. This system should provide instant assessments of airline performances of varying degrees of risk and signal those that appear to be extreme. The goal of the present work is to develop such a threshold system for the simultaneous monitoring of multiple risk or performance measures. More specifically, we apply extreme value theory (EVT) to derive multivariate extreme quantiles for the formulation of threshold systems. We also develop inference for these quantiles.

When monitoring a single risk measure, the threshold point is simply the $(1-p)$th quantile of the underlying distribution for a prescribed exceedance probability $p$. When $p$ is not too small, the usual sample quantile provides a satisfactory solution. However this is not the case when $p$ is very small, where rare events are of interest to us. EVT is particularly useful for making inferences about rare events. It has been applied successfully to many real life applications, including calculating heights of sea dikes in The Netherlands in studies of flood prevention (e.g. for $p = 10^{-4}$ per year), estimating the so-called "value-at-risk" and the related stress testing for equity portfolios, or determining insurance premiums. Other applications include sports statistics (e.g., a study of athletic records in Einmahl and Magnus 2008), meteorology (e.g., a study of precipitation return levels in Cooley, Naveau, and Nychka 2007), and engineering (e.g., a study of pitting corrosion in Fougères, Holm, and Rootzén 2006). Extensive discussions have been provided by Embrechts, Klüppelberg, and Mikosch (1997), Coles (2001), Beirlant et al. (2004), and de Haan and Ferreira (2006).

The monitoring of multiple risks presents some new difficulties. The first of these is that there may be different notions of *multivariate* quantiles for different monitoring purposes. In this article we propose defining a $(1-p)$th quantile region 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 quantile region is suitable for thresholding in risk assessment in the multivariate setting, because any observation that falls outside the proposed quantile region would imply that at least one of its component variates exceeds a certain allowable threshold. To broaden the applicability of our threshold system, we further 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 relative importance of the exceedance in individual component variables. This added flexibility is particularly useful in our application of monitoring airline performance. Note that once the weights are fixed, our quantile regions are defined uniquely.

The second difficulty in monitoring multiple risks is that here the definition of consistency requires a modification of the usual definition of consistency, because of the very small value of $p$ (see Remark A.3 and Theorems A.1 and A.2 in the Appendix). In fact, this difficulty already occurs in the univariate case. When $p$ is extremely small, we provide an estimator for the proposed quantile region and show its consistency.

The article is organized as follows. In the remaining part of Section 1 we describe the application to aviation safety and the related data. In Section 2 we briefly review EVT in the univariate setting and discuss estimators for extreme quantiles. In Section 3 we review the relevant EVT in the multivariate setting,

John H. J. Einmahl is Professor, Department of Econometrics & OR and CentER, Tilburg University, 5000 LE Tilburg, The Netherlands (E-mail: *j.h.j.einmahl@uvt.nl*). Jun Li is Assistant Professor, Department of Statistics, University of California, Riverside, CA 92521 (E-mail: *jun.li@ucr.edu*). Regina Y. Liu is Professor, Department of Statistics & Biostatistics, Rutgers University, Piscataway, NJ 08854 (E-mail: *rliu@stat.rutgers.edu*). This research was supported in part by grants from the National Science Foundation, and the Federal Aviation Administration. The discussion on aviation safety in this article reflects the views of the authors, who are solely responsible for the accuracy of the results presented herein, and does not necessarily reflect the official view or policy of the FAA. The data set used in this article has been partially masked to protect confidentiality. The authors thank the editor and referees for their thoughtful comments, which helped improve the article. They also thank Ana Ferreira and Laurens de Haan for providing many helpful suggestions, particularly concerning the bootstrap procedures, and Deyuan Li for performing the "domain of attraction" checks.


Journal of the American Statistical Association
September 2009, Vol. 104, No. 487, Applications and Case Studies
DOI: 10.1198/jasa.2009.ap08329

propose a definition for multivariate extreme quantiles, and provide corresponding estimators. We also show that these estimators are consistent. In Section 4 we apply the proposed notions of extreme quantiles to establish a threshold system for monitoring two aviation performance measures. Specifically, this threshold system divides the sample space into regions with increasing levels of risk. In Section 5 we present a simulation and comparison study to demonstrate the good performance of our approach. We provide consistency theorems and an extension of the bivariate threshold system to higher dimensions in the Appendix. The proofs of theorems are available online as supplemental materials.

Although our application here involves aviation risk assessment, the proposed thresholding procedure based on extreme quantiles should have broad applicability in other fields as well. One example of this is the development of an alarm system for signaling events of varying degrees of extremeness when monitoring several financial markets or investment assets simultaneously.

### 1.1 Data and Problem Statements

This work is motivated by the need for a threshold system for flagging events of extreme risk in an aviation monitoring scheme, which can be useful for regulating agencies, such as the FAA. The FAA regularly conducts surveillance inspections on all air carriers and carefully monitors and analyzes the findings. To increase the efficiency of their monitoring scheme, the FAA hopes to embed a threshold system in the scheme that can assign appropriate levels of potential risk to inspection outcomes. In aviation safety analysis, the regions corresponding to an increasing level of risk generally are termed *informational*, *expected*, *advisory*, and *concern*:

- *concern* (colored red) corresponds to the worst 0.15% of all possible performances.
- *advisory* (colored yellow) corresponds to the worst 1% of all possible performances but not the worst 0.15%.
- *informational* (colored green) corresponds to the best 5% of all possible performances.
- *expected* (colored blue) corresponds to the remainder of the sample space. It represents observations considered to meet the FAA's expectation under normal circumstances.

This threshold system allows the FAA to label the inspection results in terms of the severity of potential flaws and to quickly assess the safety performance of each carrier.

Many airline performance measures are considered important risk measures in aviation safety. The threshold system developed in the present work allows for simultaneous monitoring of any number of risk measures. To simplify the exposition, we focus here on two key airline performance measures: *incident rate* (IR) and *operational unfavorable ratio* (OU). The precise definitions of *incident* and *operation control* are available from the Flight Standards Information Management System at *http://fsims.faa.gov*. There an *incident* is defined as "an occurrence involving one or more aircraft in which a hazard or a potential hazard to safety is involved but not classified as an accident due to the degree of injury and/or extent of damage." This definition covers a broad range of events and may include

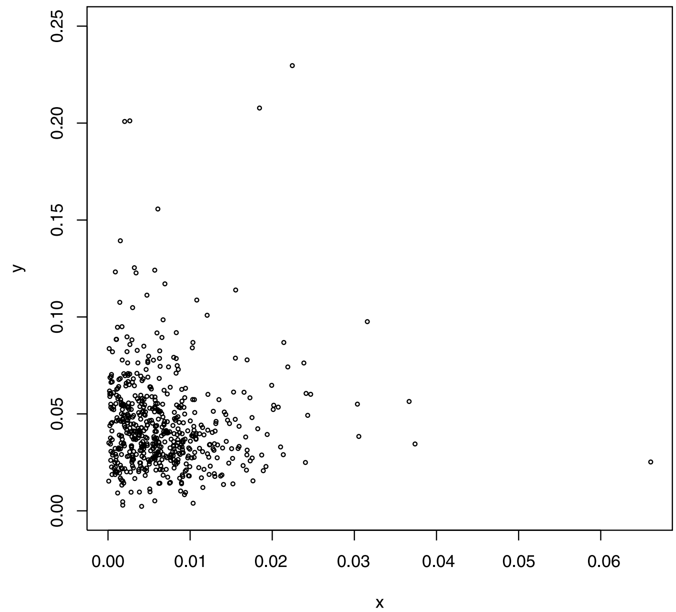


Figure 1. Scatterplot for air carrier risk measures (IR, OU).

damage to an aircraft (other than an accident) or pilot deviations. The variable IR is the number of incidents normalized by the fleet size of each air carrier. The variable OU is the number of operation control violations normalized by the total number of operations. Examples of operational control include crew assignment, load, and flight planning. Obviously, incidents can affect operation safety and vice versa; thus the variables IR and OU are not independent of each other.

We focus on a data set collected by the FAA from 10 air carriers of similar service type and fleet size over a 57-month period from July 1993 to March 1998 (see the scatterplot in Figure 1). Each data point represents a monthly observation of (IR, OU) from a given carrier. Both IR and OU are measures of nonconformance, for which a higher value is associated with a more severe potential flaw. Note that the results from autocorrelation plots and Durbin–Watson tests on all air carriers do not indicate any time-dependence of the data. This observation, along with the fact that the carriers are of similar fleet size and service type, allows us to consider the data set as a bivariate iid sample.

The specific task before us is to use our multivariate extreme quantile approach to identify the different potential risk regions desired by the FAA. In our approach, the *concern* region comprises the points lying outside the $(1 - 0.0015)$th quantile region, and the *advisory* region comprises the points lying outside the $(1 - 0.01)$th quantile region but inside the $(1 - 0.0015)$th quantile region. There are a total of 570 data points. Note that $570 \times 0.0015 = 0.855$, which implies that on the average there is less than one observation in the *concern* region. Even for the *concern* and *advisory* regions combined, the possible expected number of observations is still quite small $(570 \times 0.01 = 5.7)$. This setting of very few or no occurrences is ideal for applications of multivariate EVT.

Another aviation application, which we do not pursue in depth in this article, is to choose for an airport runway a threshold point beyond which runway crossings could be allowed. Due to the recent explosive growth in air traffic, this shortage of runway capacity has become the main cause of delays and congestion. While the construction of additional runways is being sought, the FAA may consider implementing the so-called "land and hold-short operations" (LAHSO) on aircraft landings to help ease air traffic. LAHSO would require that all aircraft landings be completed before a predetermined *hold-short point* on the runway. The advantage of implementing LAHSO is to free up a certain portion of the runway to allow for other usage and in turn reduce air traffic congestion. To establish an acceptable land-and-hold-short point on the runway, safety requirements mandate that the portion of the runway from its touchdown to the hold-short point constitute an available safe landing distance. For example, this may require that the probability that the full stop of a landing aircraft occurs beyond the hold-short point be no more than 1 in 10 million. This amounts to determining the $(1 - 0.0000001)$th quantile of the distribution of landing distance for all aircraft. Usually a data set comprises the landing distances of about 10,000 aircraft on a given airport runway. Because $10{,}000 \times 0.0000001 = 0.001 \ll 1$, this is a setting with no occurrence. The univariate extreme quantile estimator discussed later would be well suited for this application.

## 2. MONITORING ONE RISK: UNIVARIATE EXTREME QUANTILE

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

$$x_p = \inf\{x \in \mathbb{R} : \mathbb{P}(X_1 > x) \le p\}.$$

In general, the nonparametric estimator of $x_p$ is simply the corresponding sample quantile, that is, $\tilde{x}_p = \inf\{x \in \mathbb{R}: \sum_{i=1}^n I_{\{X_i > x\}}/n \le p\}$. However if $p$ is small, then the sample may contain insufficient observations to make this estimate useful in practice. For example, with a sample of 1,000 observations, the $(1 - 0.0001)$th quantile would not be well estimated by the foregoing formula. As we discuss later, EVT is useful for the inference related to such extreme quantiles of a probability distribution.

Statistical inference generally involves the central limit theorem, which characterizes the limiting distribution of the sample mean. In EVT, our focus is on the sample maximum rather than the mean. Specifically, we search for a sequence of positive numbers $\{a_n; n \ge 1\}$ and another sequence of numbers $\{b_n; n \ge 1\}$, such that

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

for all $x \in \mathbb{R}$ at which the limiting distribution function $G$ is continuous. Here $G$ is a nondegenerate distribution function. If such sequences $a_n$ and $b_n$ exist, then $F$ 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

$$G(x) = G_\gamma(x) = \exp\bigl(-(1+\gamma x)^{-1/\gamma}\bigr), \qquad 1 + \gamma x > 0,\ \gamma \in \mathbb{R} \tag{2.2}$$

[by convention, $(1+\gamma x)^{-1/\gamma} = e^{-x}$ for $\gamma = 0$]. These distributions are referred to as extreme value distributions.

The parameter $\gamma$, the extreme value index, 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 with corresponding $\gamma = 1$, a uniform distribution on the interval $[0, 1]$ has a finite endpoint with corresponding $\gamma = -1$, and a normal distribution is attracted by the Gumbel distribution with corresponding $\gamma = 0$.

Clearly, the parameter $\gamma$ determines $G$. To estimate $\gamma$, we define

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

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

The estimator $\hat{\gamma}_n^+$, proposed by Hill (1975), is generally referred to as the Hill estimator. It has been shown that $\hat{\gamma}_n^+$ is consistent and, under additional assumptions, asymptotically normal when $\gamma > 0$. Dekkers, Einmahl, and de Haan (1989) constructed the moment estimator

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

and showed that it is consistent and asymptotically normal for a general $\gamma \in \mathbb{R}$. Other generalizations of the Hill estimator for general $\gamma$ include those of Smith (1987) and Beirlant, Vynckier, and Teugels (1996).

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

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

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

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

Because the $(1-p)$th quantile of $F$, $x_p$, satisfies $1 - F(x_p) = p$, the foregoing 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

$$\hat{b}_{n/k} = X_{n-k:n} \qquad \text{and} \qquad \hat{a}_{n/k} = X_{n-k:n} M_n^{(1)} (1 - \hat{\gamma}_n^-). \tag{2.7}$$

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

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

Because the expressions (2.3)–(2.8) all involve $k$, the properties of the estimators $\hat{x}_p$ clearly depend on the choice of $k$. The estimator in (2.8) was presented by Dekkers, Einmahl, and de Haan (1989), but with no discussion on the choice of $k$.

## 2.1 The Choice of $k$

The value $k$ can be viewed as the effective sample size for tail extrapolation. 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. The importance of choosing a suitable $k$ here can be easily illustrated using the LAHSO project as an example. The larger landing distance values in the data set should be more relevant for the inference for the extreme landing pattern and thus have a greater effect on the determination of the hold-short point. The question then is how many (namely $k$) such large landing distances should be included when determining the hold-short point.

One commonly used heuristic approach for choosing $k$ in practice is to plot the estimated quantile $\hat{x}_p$ versus $k$, and then choose a $k$ that corresponds to the first stable part of the plot. This visual approach is simple but lacks precise statistical justification. Moreover, it may not always be easy to identify the first stable part of the plot.

To overcome this problem, we look for the theoretically optimal $k$ by minimizing the mean squared error (MSE) of $\hat{x}_p$, defined as

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

However this optimal choice of $k$ clearly depends on the unknown $x_p$. Thus we need to consider a data-driven approach to determining $k$, which hopefully is asymptotically equivalent (in probability) to the optimal one. Several methods for finding such a $k$ have been proposed in the literature. Gomes and Oliveira (2001) conducted a detailed investigation and comparison of these methods and found the double-bootstrap technique introduced by Danielsson et al. (2001) for estimating a positive extreme value index to be the "most appealing" approach. Ferreira, de Haan, and Peng (2003) adapted this technique to the case of extreme quantile estimation. Here we review this latter procedure and use it in our approach.

The objective MSE in (2.9) is replaced by a somewhat similar expression that contains no unknown parameters and thus can be computed directly from the given data. This analog is obtained by replacing $x_p$ with an estimator differing from that in (2.8). Following this idea, Ferreira, de Haan, and Peng (2003) defined, for $1 < k < n$, $\hat{\gamma}_{n,1}(k) = \hat{\gamma}_n$ and $\hat{a}_{n/k,1} = \hat{a}_{n/k}$ and proposed two alternative estimators, $\hat{\gamma}_{n,2}(k)$ and $\hat{a}_{n/k,2}$, which in turn yield the following two estimators for $x_p$:

$$\hat{x}_{n,j}(k) = X_{n-k:n} + \hat{a}_{n/k,j} \frac{\left(\frac{k}{np}\right)^{\hat{\gamma}_{n,j}(k)} - 1}{\hat{\gamma}_{n,j}(k)}, \qquad j = 1, 2. \tag{2.10}$$

Note that $\hat{x}_{n,1}(k)$ here is the same as $\hat{x}_p$ in (2.8) and that $\hat{x}_{n,2}(k)$ is an alternative estimator for $x_p$. Now we replace $MSE(n,k)$ in (2.9) with

$$E(\hat{x}_{n,1}(k) - \hat{x}_{n,2}(k))^2. \tag{2.11}$$

The double-bootstrap procedure can be used to determine the optimal $k$ in an asymptotic version of (2.9) as follows:

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 (2.10) based on the two bootstrap samples obtained in steps 1 and 2.
4. Repeat steps 1–3 independently, sufficiently many (say $B$) times. Calculate
$$\widehat{MSE}^*(n_i,k) = \frac{1}{B}\sum_{j=1}^{B}\left(\hat{x}_{n_i,1}^{*(j)}(k) - \hat{x}_{n_i,2}^{*(j)}(k)\right)^2, \qquad i = 1, 2, \tag{2.12}$$
where $\hat{x}_{n_i,1}^{*(j)}(k)$ and $\hat{x}_{n_i,2}^{*(j)}(k)$ are $\hat{x}_{n_i,1}(k)$ and $\hat{x}_{n_i,2}(k)$ from the $j$th bootstrap sample.
5. Find a $\hat{k}_i$ that minimizes $\widehat{MSE}^*(n_i,k)$, $i = 1, 2$ ($\hat{k}_i$ is away from 1 or $n_i$).
6. The optimal choice of $k$ in the estimator $\hat{x}_{n,1}(k)$ is then given by
$$\hat{k}_0 = \frac{\hat{k}_1^2}{\hat{k}_2} g(\hat{\gamma}_n, \hat{\rho}), \tag{2.13}$$
where, if $\hat{\gamma}_n > 0$,
$$g(\hat{\gamma}_n, \hat{\rho}) = \left(\frac{\hat{\rho}^2}{(1-\hat{\rho})^2}\right)^{1/(1-2\hat{\rho})}. \tag{2.14}$$
Ferreira, de Haan, and Peng (2003) have provided more background on formulas (2.13) and (2.14) and the expression for $g$ in the case where $\hat{\gamma}_n \le 0$.
To proceed with the case of $\hat{\gamma}_n > 0$, we use the estimator $\hat{\rho} = \hat{\rho}(k)$ of Fraga Alves, de Haan, and Lin (2003). We 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 be sufficiently large.

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

## 3. MONITORING MULTIPLE DEPENDENT RISKS: MULTIVARIATE EXTREME QUANTILE

We now consider an application of EVT in the multivariate case to establish a threshold system for the simultaneous monitoring of multiple measurements that are possibly dependent. To streamline the exposition, here we present only the arguments for the two-dimensional case. We defer the treatment of the general multidimensional case to Appendix A.2.

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from an unknown continuous distribution function $F$, with the corresponding probability measure $P$. Write $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$ the marginal distributions of $F$. Let $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ and $Y_{1:n} \le Y_{2:n} \le \cdots \le Y_{n:n}$ denote the order statistics of $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$. Similar to the univariate EVT, here $F$ is assumed to belong to the domain of attraction of an extreme value distribution; in other words, there exist sequences $\{a_{1n} > 0; n \ge 1\}$, $\{b_{1n}; n \ge 1\}$, $\{a_{2n} > 0; n \ge 1\}$, and $\{b_{2n}; n \ge 1\}$ 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 nondegenerate 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 for some $\gamma_1, \gamma_2 \in \mathbb{R}$, where $1 + \gamma_1 x > 0$ and $1 + \gamma_2 y > 0$,

$$G_1(x) := G(x, \infty) = \exp\bigl(-(1 + \gamma_1 x)^{-1/\gamma_1}\bigr) \qquad \text{and}$$
$$G_2(y) := G(\infty, y) = \exp\bigl(-(1 + \gamma_2 y)^{-1/\gamma_2}\bigr).$$

Along with the quantiles from the two marginal distributions, the tail dependence structure between the two component variables is also essential for deriving extreme quantiles in the bivariate case. We briefly describe this tail dependence structure. Let $C$ denote the distribution function of the pair $(1 - F_1(X_1), 1 - F_2(Y_1))$. Note that

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

in view of (3.1), 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 $x, y \ge 0$. Two key properties of $l$ are as follows:

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

(ii) $\max(x, y) \le l(x, y) \le x + y$, where the equality on the left 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*).

For a more thorough discussion on $l$ and general multivariate EVT, see de Haan and Ferreira (2006), part II.

### 3.1 Defining Multivariate Extreme Quantiles for Simultaneous Thresholding

In simultaneous monitoring of multiple measurements, our task is to identify proper threshold points for which the exceedance probabilities are within certain predetermined values. For monitoring bivariate measurements $(X, Y)$ from $F$, our task is to find threshold points $x$ and $y$ such that for a predetermined value $p$,

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

Obviously, there are infinitely many choices of $(x, y)$ that satisfy this condition. Different applications may also impose additional constraints on condition (3.3). One possible constraint, which is also required in our application in Section 4, is that

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

where the positive constant $c$ represents the different weights assigned to the two marginal tail probabilities. The value $c$ generally is chosen in advance to satisfy some practical constraints on the relative importance attached to the marginal variables. For example, $c = 1$ implies that exceedances of either variable are considered equally important. If $c$ is chosen to be $>1$ (as is the case in our application in Sec. 4), then the exceedance in $Y$ is considered more important or critical.

Henceforth we call the region $(-\infty, x] \times (-\infty, y]$, where $x$ and $y$ satisfy (3.3) and (3.4), as the $(1-p)$th quantile region, and refer to $(x, y)$ as its corresponding threshold point. Whenever there is no possibility of confusion, we use the term $(1 - p)$th quantile for both the region and its threshold point.

Note that for a very small $p$, (3.2) implies that

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

where $p_1 = 1 - F_1(x) = \mathbb{P}(X > x)$ and $p_2 = 1 - F_2(y) = \mathbb{P}(Y > y)$. Because $cp_1 = p_2$ [see (3.4)], $p \approx l(p_1, cp_1) = p_1 l(1, c)$, and thus

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

The foregoing discussion demonstrates that estimating a bivariate extreme quantile essentially involves two steps, estimating the marginal quantiles and then estimating $l(1, c)$. The first step can be addressed in the same way as in the univariate case discussed in Section 2, although we now need to estimate both $p_1$ and $p_2$. We address the second step next.

### 3.2 Estimating the Tail Dependence Function and the Choice of $k$

Based on the definition of $l(x, y)$ in (3.2), Huang (1992) proposed the following empirical tail dependence function of $F$ (see also Einmahl, de Haan, and Li 2006):

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

This function has been shown to be consistent and asymptotically normal. Following Peng (1998, sec. 5.4), we define and focus on the alternative estimator of $l(x, y)$,

$$\tilde{l}_{n,k}(x, y) = \hat{l}_{n,k}(2x, 2y) - \hat{l}_{n,k}(x, y), \quad (3.8)$$

because it can be adapted for bootstrap procedures. Obviously, the homogeneity property of $l$ implies that $\tilde{l}_{n,k}(x, y)$ is a consistent estimator of $l(x, y)$. As in the univariate extreme quantile estimation discussed in Section 2, we find the optimal $k$ by minimizing the MSE of $\tilde{l}_{n,k}(x, y)$, that is,

$$MSE(n, k) = E(\tilde{l}_{n,k}(x, y) - l(x, y))^2. \quad (3.9)$$

Clearly, here the optimal choice of $k$ depends on the unknown $l(x, y)$. Mimicking the approach to deriving univariate extreme quantiles, we may circumvent this problem by replacing (3.9) with an auxiliary statistic. Here Peng (1998) considered replacing (3.9) by

$$E(\tilde{l}_{n,k}(x, y) - \hat{l}_{n,k}(x, y))^2, \quad (3.10)$$

and then derived a double bootstrap procedure to find the optimal $k$ for estimating $l(x, y)$. In our setting, we consider $x = 1$ and $y = c$. This bootstrap procedure is similar to that presented in Section 2 for estimating extreme quantiles. We omit the details except to mention that step 6 there yields the optimal choice of $k$, similar to (2.13),

$$\hat{k}_0 = \frac{\hat{k}_1^2}{\hat{k}_2} g(\hat{\rho}), \quad (3.11)$$

where

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

(see Peng 1998). Here $\hat{\rho}$ can be derived following the approach of Fraga Alves, de Haan, and Lin (2003, p. 156), as

$$\hat{\rho} = \frac{1}{\log 2} \big( \log(\hat{l}_{n,k}(1, 1) - 1/2\hat{l}_{n,k}(2, 2)) - \log(2\hat{l}_{n,k}(1/2, 1/2) - \hat{l}_{n,k}(1, 1)) \big).$$

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

Step a. Obtain the estimate $\tilde{l}(1, c)$ [as given in (3.8)] for $l(1, c)$ by using the optimal $k$ obtained from the bootstrap procedure in (3.11).

Step b. Estimate the marginal tail probabilities $p_1$ and $p_2$, following (3.6), 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 estimates for the marginal quantiles $\hat{x}_{\hat{p}_1}$ and $\hat{y}_{\hat{p}_2}$. Here the optimal $k$ should be the one in (2.13), derived from the bootstrap procedure given in Section 2.

The $(\hat{x}_{\hat{p}_1}, \hat{y}_{\hat{p}_2})$ thus obtained is our proposed estimator for the quantile $(x, y)$.

### 3.3 Consistency

Let $Q_p = (-\infty, x] \times (-\infty, y]$ such that $F(x, y) = 1 - p$ and $c(1 - F_1(x)) = 1 - F_2(y)$ for some predetermined value $c \in (0, \infty)$. Denote the estimator of $Q_p$ by $\hat{Q}_p$, that is,

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

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 $F_2$ given in step c, with $\hat{p}_1 = \frac{p}{\tilde{l}(1,c)}$ and $\hat{p}_2 = c\hat{p}_1 = \frac{cp}{\tilde{l}(1,c)}$. Here the optimal $k$ is determined separately in estimating $l(1, c)$, $x_{\hat{p}_1}$ or $y_{\hat{p}_2}$. Henceforth, we denote these by $k$, $k_1$, and $k_2$.

Before establishing the consistency of $(\hat{x}_{\hat{p}_1}, \hat{y}_{\hat{p}_2})$, we note that to have a useful asymptotic model for our extreme quantile, our $p$ should depend on $n$ and should tend to 0 as $n \to \infty$. This assumption is meaningful, because it reflects the fact that we are working on the boundary of the sample. Subsequently, this suggests that an appropriate definition of consistency of extreme quantiles should be in terms of the ratio tending to 1, rather than the difference tending to 0.

Finally, we are ready to state the consistency of our estimator. Under proper conditions, including $np = O(1)$, $\frac{k}{n}, \frac{k_1}{n}, \frac{k_2}{n} \to 0$, $k, k_1, k_2 \to \infty$, as $n \to \infty$, we have

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

where $\triangle$ denotes the symmetric difference. This immediately implies that

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

More technical details are provided in the Appendix.

## 4. APPLICATION: SIMULTANEOUS THRESHOLDING OF TWO RISKS

We now apply the threshold system derived in Section 3 to carry out the project described in Section 1.1. Recall that our threshold system should divide the sample space into four regions with increasing risk levels: *informational*, *expected*, *advisory*, *concern*, with the performance measures *incident rate* (IR) and *operational unfavorable ratio* (OU) as the monitoring subjects. We have a total of 570 data points from 10 air carriers; each point represents a monthly observation of (IR, OU) from a given carrier, as shown in Figure 1. In the aviation industry, the OU is considered "twice as important" as the IR.

We first discuss the construction of the *concern* and *advisory* regions. We construct the *informational* region (and thus also the *expected* region) at the end of this section using empirical process theory, because on average sufficiently many ($570 \times 0.05 = 28.5$) observations fall in that region.

### 4.1 Concern and Advisory Regions: Extreme Quantiles

Because higher observed values imply worse performance, we interpret a multivariate performance measure as being flawed or at risk if any of its component measures exceeds a prescribed threshold. As discussed in Section 3, this interpretation gives rise to the quantile regions with the form $(-\infty, x] \times (-\infty, y]$ such that $F(x, y) = 1 - p$. The constraint that OU is twice as important as IR can be translated into the expression

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

if we denote IR as $X$ and OU as $Y$ in the setting for (3.4). This setting exactly fits the framework discussed in Section 3 with $c = 2$. Our task now is to simply find $x$ and $y$ that can satisfy the conditions $1 - F(x, y) = p$ and $2(1 - F_1(x)) = 1 - F_2(y)$ for $p = 0.0015$ for the *concern* region and for $p = 0.01$ for *advisory* region.

Before attempting to solve the foregoing problem, we first need to verify that the assumptions for bivariate EVT hold for the data set. In other words, we need to check whether $F$ is in the domain of attraction of an extreme value distribution. To begin, we use the tests proposed by Dietrich, de Haan, and Hüsler (2002) and Drees, de Haan, and Li (2006) to check whether each of the two marginal distributions $F_1$ and $F_2$ is in the domain of attraction of a univariate extreme value distribution. Both tests indicate that this is in fact so. Next, we check whether the dependence structure of $F$ ensures that $F$ is in the domain of attraction. To this end, we apply the test proposed by Einmahl, de Haan, and Li (2006) to our data. Again, we find that $F$ satisfies the assumption. Finally, we apply the procedure outlined in Section 3 to our data. To streamline the exposition, here we describe each step in the procedure only for $p = 0.0015$; the same procedure applies to $p = 0.01$.

We begin by estimating the tail dependence function $l(1, 2)$. We first apply the bootstrap procedure described in Section 3.2 to obtain the optimal $k$ for estimating $l(1, 2)$. We choose $n_1 = n^{0.95}$ and $B = 10{,}000$ (see Gomes and Oliveira 2001 for the rationale behind this choice). To avoid the few nonconvergence situations, we use a multistage bootstrap procedure, in which a bootstrap sample of size $m = 200$ is drawn for each of the 50 ($= r$) replications. With this multistage bootstrap procedure, we obtain 50 pairs of $\hat{k}_1$ and $\hat{k}_2$ by minimizing the bootstrap versions of (3.10). The bootstrap procedure is known to work well only if $\hat{k}_2 \le \hat{k}_1 \le \frac{n_1}{n_2}\hat{k}_2$. Only 37 pairs satisfy this constraint and are retained. Then we only need determine $\hat{\rho}$ [see (3.11)] to obtain the optimal $\hat{k}_0$. Based on the plot of $\hat{\rho}$ versus $k$ shown in Figure 2, we choose 1.635 (indicated by the horizontal line in the plot) as our final $\hat{\rho}$. Using this $\hat{\rho}$, together with the 37 pairs of $\hat{k}_1$ and $\hat{k}_2$, we obtain 37 approximate optimal $k$ for estimating $\tilde{l}(1, 2)$. Plugging these choices of $k$ into (3.8), we get 37 estimates of $\tilde{l}(1, 2)$. Our final choice of $\tilde{l}(1, 2)$ is the average of these 37 estimates, 2.702. Figure 3 plots $\tilde{l}(1, 2)$ versus $k$ with the horizontal line at 2.702.

Next, we plug the final estimate $\tilde{l}(1, 2) = 2.702$ into (3.6) to obtain the estimates for the two marginal tail probabilities $\hat{p}_1 = 0.00056\ (= 0.0015/2.702)$ and $\hat{p}_2 = 0.00111\ (= 2\hat{p}_1)$. After determining these two tail probabilities, we follow the procedure described in Section 2 to obtain the two corresponding univariate extreme quantiles and then the bivariate $(1 - 0.0015)$th quantile.

To proceed, we examine Figure 4, which plots the estimated $\gamma$ for both $X$ and $Y$ by different choices of $k$. The plots clearly show that both marginal distributions have positive $\gamma$. This observation allows us to apply (2.14) to choose the optimal $k$ below.

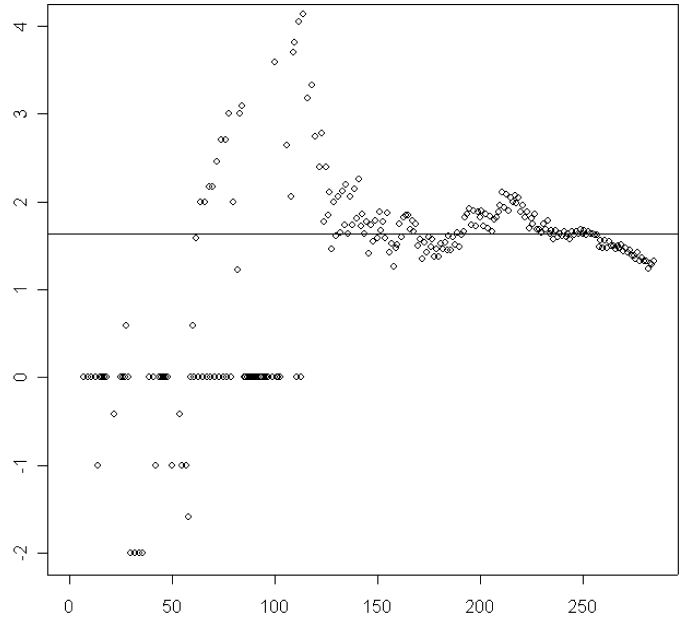


Figure 2. $\hat{\rho}$ versus $k$. (The horizontal line corresponds to $\hat{\rho} = 1.635$.)

For each marginal, we apply the (multistage) bootstrap procedure outlined in Section 2 to determine the optimal $k$ for the univariate quantile estimate. Again we chose $n_1 = n^{0.95}$ and $B = 10{,}000$ ($m = 200$ and $r = 50$). In each of the 50 replications, we need to find the $\hat{k}_i$ that minimizes $\widehat{MSE}^*(n_i, k) = \frac{1}{m}\sum_{j=1}^{m}(\hat{x}^{*(j)}_{n_i,1}(k) - \hat{x}^{*(j)}_{n_i,2}(k))^2$, $i = 1, 2$. Figure 5 plots $\widehat{MSE}^*(n_i, k)$ versus $k$ from one replication based on OU data, for $i = 1, 2$. Both plots show that $\widehat{MSE}^*$ achieves its global minimum at either end of the range of $k$. This means $\hat{k}_i$ is either very small or very large (close to $n_i$). Because neither of these is a practical estimate, we add some constraints to the range of the possible values of $k_i$ by focusing only on those

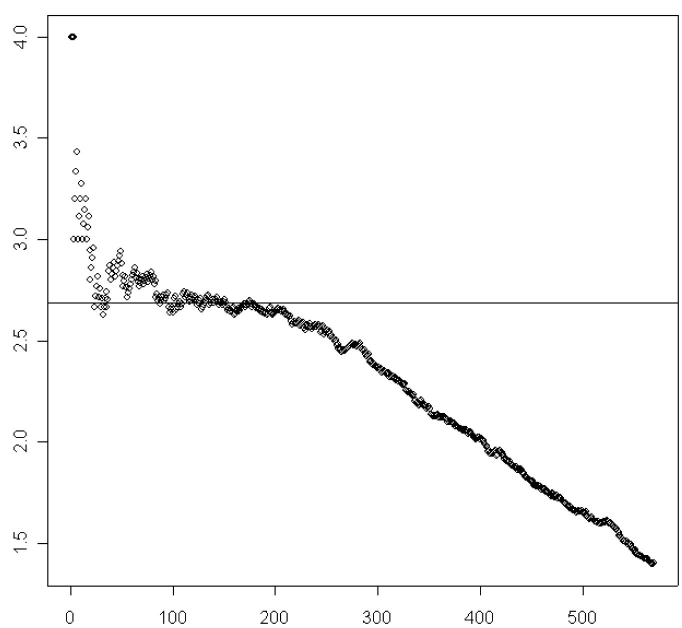


Figure 3. $\tilde{l}(1, 2)$ versus $k$. [The horizontal line corresponds to $\tilde{l}(1, 2) = 2.702$.]

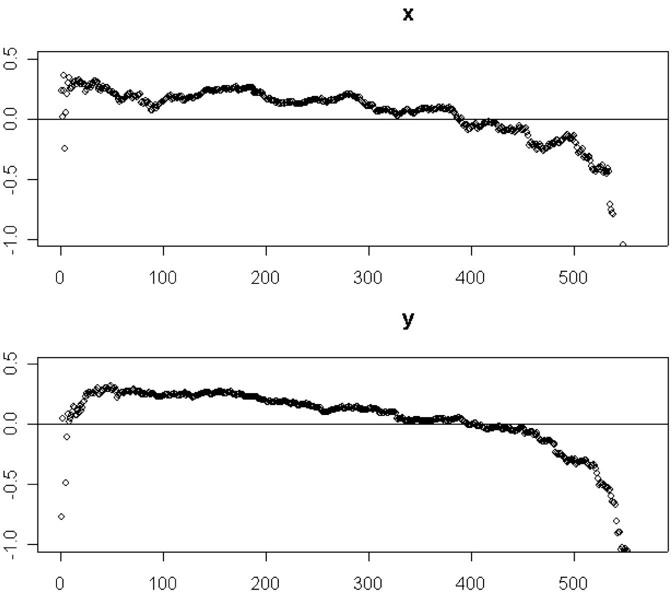


Figure 4. $\hat{\gamma}$ versus $k$.

$k_i$'s that yield a local minimum instead. Under this constraint, we find local minimums in the ranges [20, 240] in the upper panel of Figure 5 and [10, 175] in the lower panel. Thus we obtain $\hat{k}_i$, $i = 1, 2$, by minimizing $\widehat{MSE}^*(n_i, k)$ in those intervals. We apply the same procedure to obtain $\hat{k}_i$, $i = 1, 2$, for IR. Under the constraint $\hat{k}_2 \le \hat{k}_1 \le \frac{n_1}{n_2}\hat{k}_2$, we can keep only 22 pairs of $(\hat{k}_1, \hat{k}_2)$ for IR (vs 32 pairs for OU). We use these pairs together with $\hat{\rho}$ [obtained from (2.14)] to derive $\hat{k}_0$ in (2.13). We plot $\hat{\rho}$ versus $k$ for both IR and OU in Figure 6. In this figure, the horizontal lines indicate the final choices of $\hat{\rho}$: $-0.356$ for IR and $-0.456$ for OU. After plugging these values, along with the choices of $(\hat{k}_1, \hat{k}_2)$, into (2.13) and then into (2.8), we obtain 22 estimates for the marginal quantile of IR with an average of 0.064 and 32 estimates for the marginal quantile of OU with an average 0.252. Thus (0.064, 0.252) is our estimate for the bivariate $(1 - 0.0015)$th quantile. Figure 7 plots the estimated marginal quantiles versus different choices of $k$. The horizontal lines correspond to the final estimates, 0.064 and 0.252.

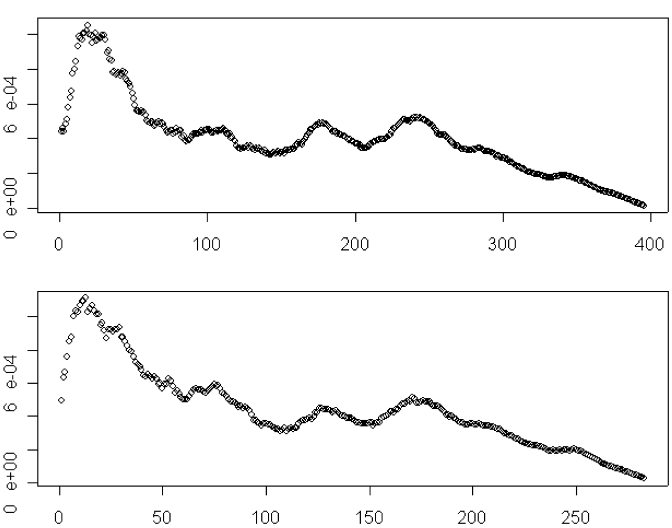

Figure 5. $\widehat{MSE}^*(n_i, k)$ versus $k$. (The upper plot is for bootstrap sample size $n_1$; the lower plot, for bootstrap sample size $n_2$.)

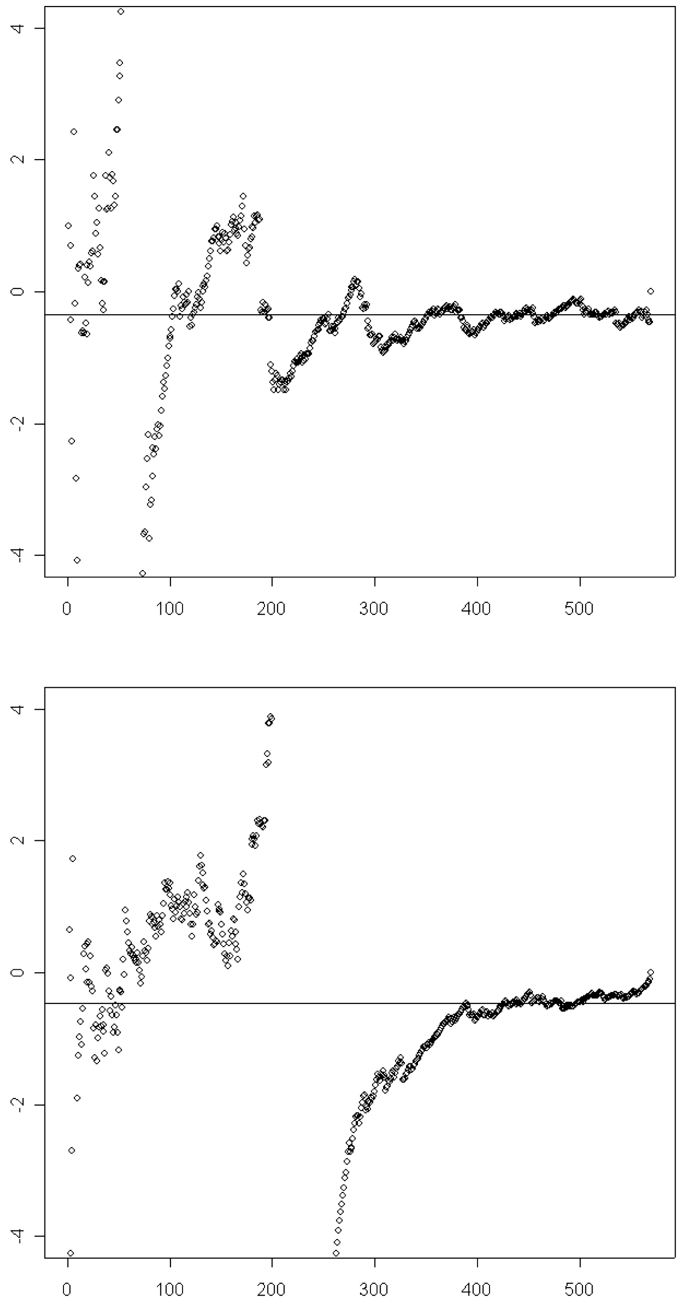

Figure 6. $\hat{\rho}$ versus $k$ for IR and OU. (The horizontal lines indicate $\hat{\rho} = -0.356$ and $-0.456$.)

Following this same procedure, we obtain (0.036, 0.150) as an estimate of the bivariate $(1 - 0.01)$th quantile.

## 4.2 Informational Region: Nonextreme Quantile

Finally, we discuss the *informational* region, which constitutes the best 5% of all performances. Because on average this region contains sufficient observations, we can use the usual empirical process approach instead of EVT. Obviously, both components in this region should assume low values, and the region is of the form $(-\infty, x] \times (-\infty, y]$, with $F(x, y) = \tilde{p}$. Let $F_1^{-1}$ and $F_2^{-1}$ denote the left-continuous quan-

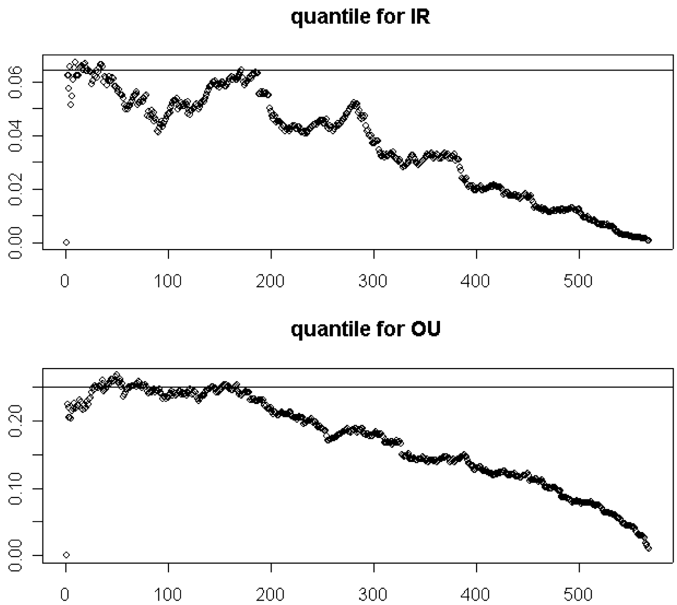


Figure 7. Extreme quantile estimates for IR and OU versus $k$. The horizontal lines correspond to the final estimates, 0.064 and 0.252.

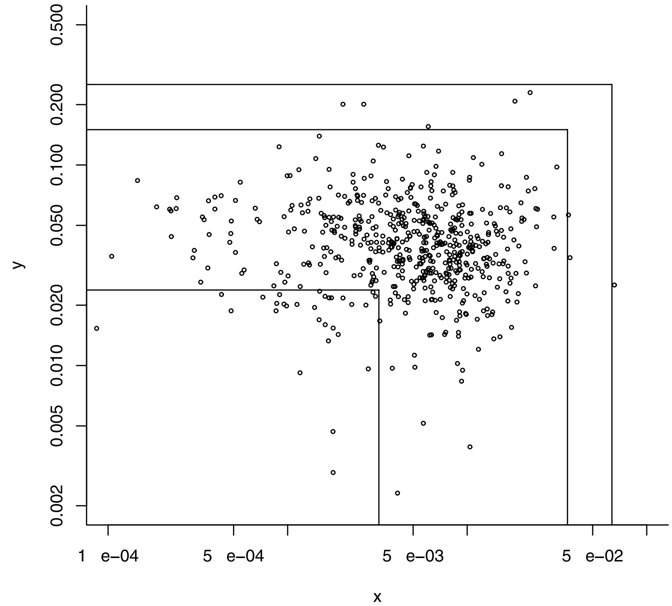


Figure 8. Threshold system: the four designated regions (scatterplot on the log scale).

tile functions corresponding to $F_1$ and $F_2$. The constraint that OU is "twice as important" as IR is then translated into $x = F_1^{-1}(2t), y = F_2^{-1}(t)$ for some $t \in (0, 1)$. For a given $\tilde{p}$, let $t_0$ satisfy $F(F_1^{-1}(2t_0), F_2^{-1}(t_0)) = \tilde{p}$. We can express the *informational* region as

$$O_{\tilde{p}} = (-\infty, F_1^{-1}(2t_0)] \times (-\infty, F_2^{-1}(t_0)],$$

and estimate it by

$$\hat{O}_{\tilde{p}} = \big(-\infty, \tfrac{1}{2}(X_{2n\hat{t}:n} + X_{2n\hat{t}+1:n})\big] \times \big(-\infty, \tfrac{1}{2}(Y_{n\hat{t}:n} + Y_{n\hat{t}+1:n})\big],$$

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 to our data with $\tilde{p} = 0.05$, we get the values (0.0032, 0.0238), corresponding to IR and OU.

Following an empirical process approach, we immediately obtain the following consistency result for any continuous $F$:

$$P(\hat{O}_{\tilde{p}} \triangle O_{\tilde{p}}) \xrightarrow{p} 0.$$

## 4.3 Final Solution

The estimated threshold regions derived in Sections 4.1 and 4.2 are shown in Figure 8, in which both coordinates are presented on a logarithmic scale for better viewing. In terms of the original scale, the right upper corners of the three nested rectangles correspond to the estimated quantiles (0.0032, 0.0238), (0.036, 0.150), and (0.064, 0.252). The upper right region corresponds to *concern*, the next upper right region corresponds to *advisory*, the lowest rectangle corresponds to the *informational* region, and the region between *informational* and *advisory* corresponds to *expected*. This is a case study of EVT in action.

# 5. SIMULATION AND COMPARISON STUDY

Finally, we conduct a simulation to study the finite-sample performance of our approach. We take $n = 1{,}000$, $c = 2$ (as in the case of our application), and $p = 0.001$ and 0.003. For each case we run 2,000 simulations. The data are simulated from two distributions. The first is a Cauchy distribution restricted to only the first quadrant, with density $(2/\pi)(1 + x^2 + y^2)^{-3/2}, x, y > 0$. The corresponding tail dependence function satisfies $l(1, 2) = \sqrt{5}$. The other distribution is the bivariate Gumbel logistic model with distribution function $\exp(-(x^r + y^r)^{1/r}), x, y > 0; r \ge 1$. To make this distribution sufficiently resemble our FAA data in the application, we choose $r$ so that $l(1, 2) = 2.702$ and transform both marginals to Pareto distributions with the extreme value index 0.25 (see Figure 4), for which the marginal distribution functions become $1 - 1/x^4, x \ge 1$.

We compare our estimator with the nonparametric estimator (NP) based on the usual marginal order statistics. The comparison is based on the MSE

$$E\left(\frac{1 - P(\hat{Q}_p)}{p} - 1\right)^2.$$

The simulation results, reported in Table 1, clearly show that our estimator consistently outperforms the NP estimator. Note that the improvement from using our EVT approach is substantially greater in the case with $p = 0.001$ than in the case with $p = 0.003$. This is as expected, because a smaller $p$ value makes it more difficult to extrapolate the tail behavior, which is exactly the strong point of EVT. Note also that the mean squared errors of our estimator are quite good given the difficult nature of the estimation problem. We stress that for either $p = 0.001$ or $p = 0.003$ we are essentially dealing with multivariate quantiles on the boundary of the sample. The estimation of these

Table 1. Empirical MSE for the EVT and NP approaches under Cauchy and Gumbel distributions

| | $p = 0.001$ | | $p = 0.003$ | |
|---|---|---|---|---|
| Method | Cauchy | Gumbel | Cauchy | Gumbel |
| EVT | 0.47 | 0.64 | 0.24 | 0.28 |
| NP | 2.10 | 2.47 | 0.34 | 0.32 |

extreme quantiles is statistically more difficult than the estimation of the mean, and thus the same level of accuracy should not be expected. Overall, our approach performs well.

## APPENDIX

### A.1 Consistency of the Quantile Estimators

We show that the proposed 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). (To streamline the presentation, we use the same notation $p$ to indicate the tail probability in both univariate and multivariate settings. When needed, an index $j$ will be added to indicate the $j$th marginal.)

*Theorem A.1.* Assume the following:

(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 $\lim_{t\to\infty} A(t) = 0$, constant sign near $\infty$, and

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

for all $x$ with $1 + \gamma x > 0$ and some $\eta < 0$, where $H_{\gamma,\eta}(x) = \frac{1}{\eta}(\frac{x^{\gamma+\eta}-1}{\gamma+\eta} - \frac{x^\gamma - 1}{\gamma})$.
(e) $A_n = \sqrt{k}(\frac{\hat{a}_{n/k}}{a_{n/k}} - 1) = O_p(1)$, $B_n = \sqrt{k}(\frac{\hat{b}_{n/k} - b_{n/k}}{a_{n/k}}) = O_p(1)$, and $\Gamma_n = \sqrt{k}(\hat{\gamma}_n - \gamma) = O_p(1)$.

Then we have

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

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

*Remark A.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 also can be easily shown that $\bar{F}(\hat{x}_{\hat{p}})/\hat{p} \xrightarrow{p} 1$.

The next theorem is an extension of Theorem A.1 to a bivariate setting.

*Theorem A.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 A.1. Then we have

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

A somewhat related work is the study of multivariate quantile curves of de Haan and Huang (1995). Another approach, considered by Joe, Smith, and Weissman (1992), modeled *parametrically* the tail dependence function.

*Remark A.3.* Observe that $np = O(1)$ and thus $p \to 0$ (very fast). This implies that the usual consistency statement $P(\hat{Q}_p \triangle Q_p) \xrightarrow{p} 0$ is inappropriate. Thus we consider the consistency in terms of a ratio here as well as in Theorem A.1. Actually, this consistency statement is very precise; it states that the ratio of two quantities, which are of the order $O_p(1/n)$ and $O(1/n)$, tends to 1 in probability.

*Remark A.4.* One important novelty of our approach is that $k$, $k_1$, and $k_2$ can be chosen "independently." Thus is desirable because their optimal values can be very different, as they depend on $C$, $F_1$, and $F_2$, respectively. The aforementioned work of de Haan and Huang (1995) and almost all developments in multivariate EVT to date require that $k = k_1 = k_2$.

### A.2 In the Setting of $\mathbb{R}^d$

The results of Theorem A.2 can be easily generalized to random vectors in higher dimension. Here we highlight the changes needed for such a generalization. Assume that the sample consists of $n$ independent copies of $\mathbf{V} = (V_1, \ldots, V_d)$, a random vector in $\mathbb{R}^d$ with distribution function $F$. Denote the sample by $\mathbf{V}_1, \ldots, \mathbf{V}_n$ and the $j$th order statistic of the $m$th component variable of the sample points by $V_{j:n,m}$. Assume also that $F$ satisfies the generalization to dimension $d$ of the domain of attraction condition (3.1) with the corresponding $d$ extreme value indexes $\gamma_1, \ldots, \gamma_d$. We then have

$$l(x_1, \ldots, x_d) = -\log G\left(\frac{x_1^{-\gamma_1} - 1}{\gamma_1}, \ldots, \frac{x_d^{-\gamma_d} - 1}{\gamma_d}\right),$$

where $l$ again is homogeneous. Similar to the case of $\mathbb{R}^2$ in (3.7), $l$ can be estimated with

$$\hat{l}_{n,k}(x_1, \ldots, x_d) = k^{-1} \sum_{j=1}^n I_{[V_{j,1} \ge V_{n-[kx_1]+1:n,1} \text{ or } \cdots \text{ or } V_{j,d} \ge V_{n-[kx_d]+1:n,d}]},$$

and with the corresponding $\tilde{l}_{n,k}$; compare (3.8). It is easy to show that both estimators are consistent (see, e.g., exercise 7.1 in de Haan and Ferreira 2006).

For thresholding purposes, we would need to search for the multivariate quantile $(x_1, \ldots, x_d)$ such that

$$\mathbb{P}(V_1 > x_1 \text{ or } \cdots \text{ or } V_d > x_d) = p,$$

and also that for given positive constants $c_2, \ldots, c_d$,

$$c_j \mathbb{P}(V_1 > x_1) = \mathbb{P}(V_j > x_j), \qquad j = 2, \ldots, d.$$

Denote the resulting solution by $(x_{1,p_1}, \ldots, x_{d,p_d})$. We then have, as in (3.6),

$$p_1 \approx \frac{p}{l(1, c_2, \ldots, c_d)}, \qquad p_j \approx \frac{c_j p}{l(1, c_2, \ldots, c_d)}, \qquad j = 2, \ldots, d.$$

Proceeding similarly as for the bivariate case, we can now define the subspace determined by the threshold point $(x_{1,p_1}, \ldots, x_{d,p_d})$ as

$$Q_p(d) = \big(-\infty, x_{1,p_1}\big] \times \cdots \times \big(-\infty, x_{d,p_d}\big],$$

and define its corresponding estimate as

$$\hat{Q}_p(d) = \big(-\infty, \hat{x}_{1,\hat{p}_1}\big] \times \cdots \times \big(-\infty, \hat{x}_{d,\hat{p}_d}\big].$$

Finally, arguments similar to those in the proof of Theorem A.2 allow us to conclude that, under the $d$-variate version of the conditions of Theorem A.2,

$$\frac{P(\hat{Q}_p(d) \triangle Q_p(d))}{p} \xrightarrow{p} 0.$$

## SUPPLEMENTAL MATERIALS

The proofs of Theorems A.1 and A.2 are available online at *http://center.uvt.nl/staff/einmahl/AppELL.pdf*.

*[Received June 2008. Revised January 2009.]*

## REFERENCES


Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004), *Statistics of Extremes, Theory and Applications*, New York: Wiley.

Beirlant, J., Vynckier, P., and Teugels, J. (1996), "Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics," *Journal of the American Statistical Association*, 91, 1659–1667.

Coles, S. G. (2001), *An Introduction to Statistical Modelling of Extreme Values*, New York: Springer.

Cooley, D., Naveau, P., and Nychka, D. (2007), "Bayesian Spatial Modeling of Extreme Precipitation Return Levels," *Journal of the American Statistical Association*, 102, 824–840.

Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001), "Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation," *Journal of Multivariate Analysis*, 76, 226–248.

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

de Haan, L., and Huang, X. (1995), "Large Quantile Estimation in a Multivariate Setting," *Journal of Multivariate Analysis*, 53, 247–263.

Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989), "A Moment Estimator for the Index of an Extreme-Value Distribution," *The Annals of Statistics*, 17, 1833–1855.

Dietrich, D., de Haan, L., and Hüsler, J. (2002), "Testing Extreme Value Conditions," *Extremes*, 5, 71–85.

Drees, H., de Haan, L., and Li, D. (2006), "Approximations to the Tail Empirical Distribution Function With Application to Testing Extreme Value Conditions," *Journal of Statistical Planning and Inference*, 136, 3498–3538.

Einmahl, J. H. J., and Magnus, J. R. (2008), "Records in Athletics Through Extreme-Value Theory," *Journal of the American Statistical Association*, 103, 1382–1391.

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," *The Annals of Statistics*, 34, 1987–2014.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events, for Insurance and Finance*, New York: Springer.

Ferreira, A., de Haan, L., and Peng, L. (2003), "On Optimizing the Estimation of High Quantiles of a Probability Distribution," *Statistics*, 37, 401–434.

Fisher, R. A., and Tippett, L. H. C. (1928), "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Mathematical Proceedings of the Cambridge Philosophical Society*, 24, 180–190.

Fougères, A., Holm, S., and Rootzén, H. (2006), "Pitting Corrosion: Comparison of Treatments With Extreme Value Distributed Responses," *Technometrics*, 48, 262–272.

Fraga Alves, M. I., de Haan, L., and Lin, T. (2003), "Estimation of the Parameter Controlling the Speed of Convergence in Extreme Theory," *Mathematical Methods of Statistics*, 12, 155–176.

Gnedenko, B. V. (1943), "Sur la distribution limite du terme maximum d'une série aléatoire," *Annals of Mathematics*, 44, 423–453.

Gomes, M. I., and Oliveira, O. (2001), "The Bootstrap Methodology in Statistics of Extremes—Choice of the Optimal Sample Fraction," *Extremes*, 4, 331–358.

Hill, B. M. (1975), "A Simple General Approach to Inference About the Tail of a Distribution," *The Annals of Statistics*, 3, 1163–1174.

Huang, X. (1992), "Statistics of Bivariate Extreme Values," Ph.D. Thesis 22, Tinbergen Institute, Erasmus University, Rotterdam.

Joe, H., Smith, R., and Weissman, I. (1992), "Bivariate Threshold Methods for Extremes," *Journal of Royal Statistical Society, Ser. B*, 54, 171–183.

Peng, L. (1998), "Second Order Condition and Extreme Value Theory," Ph.D. Thesis 178, Tinbergen Institute, Erasmus University, Rotterdam.

Smith, R. L. (1987), "Estimating Tails of Probability Distributions," *The Annals of Statistics*, 15, 1174–1207.