Estimation concerning risk under extreme value conditions

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, Prof. Dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op woensdag 24 oktober 2012 om 10.15 uur door

Juan-Juan Cai
geboren op 23 january 1985 te Xianyou, China.
Promotores: Prof. dr. John H.J. Einmahl
           Prof. dr. Laurens de Haan

Overige Commissie Leden: Dr. Feike C. Drost
                          Prof. dr. Anne-Laure Fougères
                          Prof. dr. Roger J.A. Laeven
                          Prof. dr. Thomas Mikosch
                          Prof. dr. Bas J.M. Werker
Acknowledgements

It has been nearly three years since I first came to Tilburg University and started my work on this thesis. These three years meant a lot to me: the presented thesis certainly counts and the people who I want to thank here made this part of my life special.

Extreme events do happen in reality and consequences might be very positive. I consider that to have supervisors like John (Einmahl) and Laurens (de Haan), one needs to be extremely lucky. The first time I met Laurens was in the summer of 2009 in Shanghai. On the train to Suzhou we were talking about the book, *Extreme value theory: an introduction*, which later turned out to be the Bible of my Ph.D study. And I remember that I had questions starting from page 2. From that moment, I have learned the patience of Laurens who also explains mathematical issues in an elegant and beautiful way that usually inspires me. John has influenced me to a great extent with his rigorous attitude toward science, in-depth thinking, extremely responsible for both research and teaching and also his humorous and considerate personality. Besides their incredible knowledge and experience, I have also benefited greatly from the time that we sat next to each other and worked together. Both John and Laurens have spent large amount of time teaching me and working with me. This can happen everywhere, in John’s office, at a garden in Alcochete, at mensa, and even on a train. It has been so joyful and precious for me. I am deeply indebted to their support and help which go beyond the research.

The recommendation and encouragement from my master thesis advisor, Prof. Han-Ying Liang, was the first step toward this Ph.D study. I am grateful for him. Also, I would like to thank (Zhou) Chen not only for being a great co-author of two papers in this thesis but also for all his warm help throughout my
Ph.D study. He has generously shared with me his opinions and suggestions on research work and abroad life.

It is a great honor for me to have Dr. Feike C. Drost, Prof. Anne-Laure Fougères, Prof. Roger J.A. Laeven, Prof. Thomas Mikosch and Prof. Bas J.M. Werker as my Ph.D committee. I am grateful for their helpful comments and suggestions given during my pre-defense.

I have visited University of Lisbon for several times and was financially supported by the research grant ENES-Project. Therefore, I would like to thank the hospitality of the Faculty of Science. I also take this opportunity to thank Fatima for her kindness and encouragement that was particularly important for me.

Job searching was a difficult stage for me. Besides the constant support of my supervisors, I have also been helped by many others: Regina Liu and Zhou Chen with their references and Bas, Ramon, and Roger with their feedbacks and opinions on the mock interviews and presentations. I thank Andrea, Otilia, Meltem and Paopao (Chaoyang) for their help as well.

The Department of EOR in Tilburg University offers a great working environment for Ph.D students. I would like to thank Sami Umut Can and Mitja Stadje for being my seminar discussants and for the nice teaching experience we three shared together, this also applies to Serhan. My thanks also go to Korine for her most efficient work.

My dear friends, without you, I would probably still have written this thesis however the three years life would have been completely different. I have had great office time with Maria thanks to her warm attitude, and Luc and Jan thanks to their google-like knowledge on English and Latex. The three of you have helped me out on many practical problems. Thanks! The cheerful lunch group, the afternoon coffee breaks (maybe too often and too long), the squash tournaments, the Friday after-work drinks, the Chinese-Italian dinners, the Czech mom-made Slivovica, the Easter-Karaoke, Roman lessons, Tilburg Chinese Ph.D dinners, European Cup,..., the list is endless. I thank all of you for the time we
spent and the fun we had together: Rasa, Ying, Chiara, Chiara, Marco, Jan, Jarda, Umut, Wangdun, Jikan, Zongxin, Huaxiang, Maning, Kebing, Niemiao, Fangfang, Ting, Michele, Geng, Dominik. To Gaia, Yan and Sara, I am extremely thankful to have you girls in my life. Gaia asked me about homesick the first day we met. Because of them, this has never been a problem. I went to Gaia with all types of questions I encountered and she has always been reliable. Yan, with her inner peace, could magically calm me down when I was panic. Thanks to Sara’s passion for almost everything (no matter it suits her or not), I experienced novel things which I had never tried before. The way we care and support each other is more family-like. The happiness and tears we shared together make our friendship unique and eternal. Thank you, girls, for everything!

Dear Ruli, Dao-Liuye, Xiaoyu, Daolu, Shimei, thanks for your support. Long-distance friendship is not easy and hence means even more.

Feng, thanks for encouraging me to pursue this Ph.D at the first place. Thanks for being patient and always supportive. Having you next to me in this journey has made it much easier for me. I am happy that we grow so much together. The love of baba and mama has always been there in my whole life and encourages me to face all the difficulties. My deepest thanks also go to my beloved nainai, gege, saozi, Zhiyun and Hanyu. At the end, only love really matters!

Tilburg, The Netherlands
August 2012
# Contents

Acknowledgements

## 1 Introduction
1.1 Univariate extreme risk
1.2 Multivariate extreme risk

## 2 Estimation of extreme risk regions under multivariate regular variation
2.1 Introduction
2.2 Main Result
2.3 Simulation study
2.4 Application
2.5 Proofs

## 3 Bias Correction in Extreme Value Statistics with Index around Zero
3.1 Introduction
3.2 Tail expansion and the choice of $k$
  3.2.1 Tail expansion: the first, second and third order conditions
  3.2.2 The choice of $k$
3.3 Bias correction for the PWM estimators on the extreme value index, high quantiles and endpoint
  3.3.1 Bias correction for the extreme value index estimator
  3.3.2 Bias correction for the quantile estimator
  3.3.3 Bias correction for the endpoint estimator
3.4 Simulations
  3.4.1 Simulations on $\gamma$ estimation
  3.4.2 Simulations on the high quantile and endpoint estimation
3.5 Application to the still water level
3.6 Appendix
4 Estimation of marginal expected shortfall: the mean when a related variable is extreme 67
4.1 Introduction ........................................ 68
4.2 Main Results ........................................ 70
  4.2.1 $X$ Positive .................................. 71
  4.2.2 $X$ Real ....................................... 73
4.3 Simulation Study .................................... 74
4.4 Application ......................................... 77
  4.4.1 Discussion on Serial Dependence .......... 80
4.5 Proofs .............................................. 81

Bibliography 97
Chapter 1

Introduction

Extreme risks are associated with highly unusual events. By nature, the historical and available data corresponding to those events are scarce if exist, which makes the statistical inference difficult. Extreme value theory is particularly useful to tackle problems of this type. The development of the theory dates back to early 20th century, see Fisher and Tippet (1928). Since then it has become a rapidly growing branch in statistics and probability, with applications in a large variety of domains. One of the key applications is to answer questions from meteorology and environmental science, such as problems related to floods, heavy rainfall, wind storms, extreme temperature and earthquakes; see de Haan (1990); Coles and Tawn (1991); Coles and Walshaw (1994); de Haan and de Ronde (1998); Buishand et al. (2008). It has also been used in risk management in finance and insurance; especially since the financial crisis started in 2007, the heavy tail phenomena have received significant attention and recognition; see Embrechts et al. (1997); Klüppelberg and Mikosch (1997); Donnelly and Embrechts (2010); Embrechts et al. (2009). Applications can also be found in many other fields, for instance, the aviation safety (Einmahl et al. (2009)) and internet auctions (de Haan et al. (2009)).

Extremes are not always negative or unfavorable. Interesting examples are maximum life span of humans (Aarssen and de Haan (1994)) and ultimate athletic world records (Einmahl and Magnus (2008)).
For a comprehensive study on extreme value theory, we recommend the monographs of de Haan and Ferreira (2006); Resnick (1987, 2007) and Beirlant et al. (2004).

In this introduction, we briefly present some basic facts on both univariate and multivariate extreme theory and the contents of the thesis.

1.1 Univariate extreme risk

Based on a random sample $X_1, \ldots, X_n$ from distribution $F$, univariate extreme risk modeling is to answer two types of questions:

- Given an extremely small $p$, how to find the $(1 - p)$-th quantile $x_p = \inf \{ x : 1 - F(x) \leq p \}$?
- Given a large $x_0$, what is the probability of $X$ exceeding $x_0$: $P(X > x_0)$?

People often encounter problems where $p < 1/n$ or $x > \max\{X_1, \ldots, X_n\}$. In other words, we need to study the tail of the distribution. Suppose there exist two sequences of real numbers, $a_n > 0$ and $b_n$, such that, as $n \to \infty$,

$$\frac{\max\{X_1, \ldots, X_n\} - b_n}{a_n} \xrightarrow{d} Y,$$

where $Y$ has a non-degenerate distribution $G$, or equivalently,

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

for any continuity point $x$ of $G$. We call $G$ an extreme value distribution and $F$ is in the max-domain of attraction of $G$. The tail behavior of $F$ is then very much captured by $G$. Further, up to a location and scale change, the distribution of $G$ is determined only by one parameter. As shown in Fisher and Tippet (1928) and Gnedenko (1943),

$$G_\gamma(x) = \exp\left(-\frac{1}{\gamma x}\right), \quad 1 + \gamma x > 0, \quad \gamma > 0,$$
with $\gamma$ real and where for $\gamma = 0$, $(1 + \gamma x)^{-1/\gamma} = \exp(-x)$. The parameter $\gamma$ is called extreme value index. It characterizes the heaviness of the tail of the distribution. According to the sign of $\gamma$, the distributions can be distinguished into three categories.

- For $\gamma > 0$, the distribution has a heavy right tail as the right endpoint is infinite and the moments of order greater than $1/\gamma$ do not exist. Examples of distributions in such domain of attraction are Cauchy, Student and Pareto distributions.

- For $\gamma = 0$, the distribution has a light right tail. The right endpoint can be either finite or infinite and moments of any order exist. Examples are normal and Gamma distributions.

- For $\gamma < 0$, the distribution has a finite endpoint. The uniform distribution is one of the examples.

The estimation of $\gamma$ is crucial when applying extreme value statistics. The mostly used estimators are the Hill (1975) estimator for $\gamma$ positive, the moment estimator (Dekkers et al. (1989)) for $\gamma$ real, the maximum likelihood estimator (Smith (1987); Zhou (2010)) for $\gamma > -1$, and the probability weighted moment (PWM) estimator (Hosking and Wallis (1987)) for $\gamma < 1$.

- In Chapter 3, we study the PWM estimators that are often applied in meteorology and environmental science. We develop bias correction procedures for the PWM estimators of $\gamma$, high quantile and, for $\gamma < 0$, the endpoint of the distribution. Advantages of the estimators are shown in a simulation and comparison study. We also present an environmental application on estimating “once per 10,000 years” still water level at Hoek van Holland, The Netherlands.

To conclude this section, we give an idea on how to construct estimators of the tail quantile and tail probability as raised in the beginning of this section. We consider heavy tail situation, so $\gamma > 0$. The estimators are initially studied in Weissman (1978) and Hill (1975). We start with a necessary and sufficient
condition for $F$ being in the max-domain of attraction of a positive $\gamma$, which is that
\[
\lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad x > 0.
\] (1.1.1)

For a large $k < n$, choosing the $(n - k)$-th order statistic, $X_{n-k,n}$, as $t$ and $x = \frac{x_0}{X_{n-k,n}}$ in (1.1.1), the estimator for $P(X > x_0)$ is then given by
\[
\frac{k}{n} \left( \frac{X_{n-k,n}}{x_0} \right)^{1/\gamma}.
\]

By solving $\frac{k}{n} \left( \frac{X_{n-k,n}}{x_p} \right)^{1/\gamma} = p$ for $x_p$, we obtain the quantile estimator as
\[
\hat{x}_p = X_{n-k,n} \left( \frac{k}{np} \right)^{\gamma}.
\]

The choice of $k$ for the estimation is a difficult and important issue; it balances the bias and variance of the estimators; see our discussion in Section 3.2.2.

### 1.2 Multivariate extreme risk

Unlike in the univariate case, there is no unique way to define tail quantiles or threshold exceedances for a multivariate distribution. A proper modeling depends on the specific question at hand. Hence there are various types of interesting problems. In Chapters 2 and 4 of this thesis, we study two issues concerning multivariate extreme risk.

- In Chapter 2, we estimate extreme risk regions of a random vector $Z$, which can be seen as multivariate quantiles. For a given small probability $p$, we define the extreme risk region as a set of the form $Q = \{ z \in \mathbb{R}^d : f(z) < \beta \}$, with $f$ the joint density, and $\beta$ a small number such that $P(Z \in Q) = p$.

- In Chapter 4, we estimate, for a given small $p$, $E(X|Y > y_p)$ with $y_p$ the $(1 - p)$-th quantile of $Y$. Consider $X$ and $Y$ as the loss return on the equity of a financial institution ($X$) and that of the entire market ($Y$); then the
quantity under investigation is called marginal expected shortfall, which is an important factor for measuring the systemic risk of financial institutions.

Since both problems ask for inference in the tail of the sample, that is, when $p = O(1/n)$ with $n$ the sample size, multivariate extreme value theory is exploited. A multivariate distribution is fully determined by the marginal distributions and the dependence structure of the components. As the marginal distributions can be dealt with using univariate techniques, the study of the tail dependence is essential in multivariate extreme value theory.

For simplicity, we consider a bivariate random vector $(X, Y)$ with distribution $F$ and marginal distributions $F_1$ and $F_2$. $F$ is said to be in the max-domain attraction of $G$, a (bivariate) extreme value distribution if there exist sequences of numbers $a_n, c_n > 0$ and $b_n, d_n$ real, such that for all continuity points $(x, y)$ of $G$,

$$\lim_{n \to \infty} F_n(a_n x + b_n, c_n y + d_n) = G(x, y).$$  

(1.2.1)

As a result, $F_1$ and $F_2$ are in the univariate max-domain of attraction and the limit, 

$$\lim_{t \to \infty} P(1 - F_1(X) < x/t \text{ or } 1 - F_2(Y) < y/t) =: l(x, y),$$  

(1.2.2)

exists for all $(x, y) \in [0, \infty)^2 \backslash \{(\infty, \infty)\}$. The limit $l(x, y)$ is called tail dependence function. It carries the information about the tail dependence between $X$ and $Y$. There are also other ways to study tail dependence, for instance, via the exponent measure or spectral measure; see Chapter 2. For a general review of tail dependence, consult de Haan and Ferreira (2006, Chapter 6) and Beirlant et al. (2004, Chapter 8).

When applying multivariate extreme value theory, the setup of the model is quite flexible. For the estimation of the extreme risk regions in Chapter 2, we assume multivariate regular variation, that is, the marginal distributions are all corresponding to one $\gamma$ whereas in Chapter 4, the threshold component $Y$ is not necessary to be in the max-domain of attraction. In Chapter 2, we prove a refined form of consistency of the extreme risk regions estimator, given a random sample of multivariate regularly varying random vectors. In Chapter 4, the asymptotic
normality of the estimator of the marginal expected shortfall is established for a wide nonparametric class of bivariate distributions. Good finite sample performance of the two estimators is shown in detailed simulation and comparison studies. Both procedures are applied to financial data.
Chapter 2

Estimation of extreme risk regions under multivariate regular variation

[Based on joint work with John H.J. Einmahl and Laurens de Haan
Estimation of extreme risk regions under multivariate regular variation,

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

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

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

$$\lim_{t \to \infty} \frac{P(||X|| \geq tx, X/||X|| \in A)}{P(||X|| \geq t)} = x^{-\alpha} \Psi(A), \quad (2.1.1)$$

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

$$\lim_{t \to \infty} \frac{P(||X|| \geq tx)}{P(||X|| \geq t)} = x^{-\alpha}, \text{ for } x > 0, \quad (2.1.2)$$

and there exists a measure $\nu$ such that

$$\lim_{t \to \infty} \frac{P(X \in tB)}{P(||X|| \geq t)} = \nu(B) < \infty, \quad (2.1.3)$$

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

$$\nu(aB) = a^{-\alpha} \nu(B). \quad (2.1.4)$$

Clearly, on $\{z \in \mathbb{R}^d : ||z|| \geq 1\}$, $\nu$ is a probability measure. The limit relation in (2.1.3) is a multivariate analogue of the “peaks-over-threshold” or “generalized Pareto limit” method in one-dimensional extreme value theory. Particular cases of (2.1.1) are distributions in the sum-domain of attraction of $\alpha$-stable distributions and heavy tailed elliptical distributions such as multivariate $t$-distributions (see Hashorva (2006)).
We require the convergence in (2.1.2) and (2.1.3) at the density level:
(a) Suppose that the distribution of $X$ has a continuous and positive density $f$ and that for some positive function $q$ and some positive function $V$ regularly varying at infinity with negative index $-\alpha$, we have

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

(2.1.5)

and

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

(2.1.6)

Then $q$ is continuous on $\mathbb{R}^d \setminus \{0\}$ and $q(az) = a^{-d-\alpha}q(z)$ for all $a > 0$ and $z \neq 0$. Throughout we can and will take $V(t) = P(||X|| > t)$ (see Lemma 1, Section 5). From Lemma 1 it follows that doing so (2.1.3) holds with $\nu(B) = \int_B q(z)dz$.

The extreme region will be of the form

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

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

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

Having an estimate of $Q$ can be important in various settings. It can be used as an alarm system in risk management: if a new observation falls in the estimated $Q$ it is a signal of extreme risk. See Einmahl et al. (2009) for an application to aviation safety along these lines. In a financial or insurance setting, points on the boundary of the estimate of $Q$ can be used for stress testing. The estimate of $Q$ can also be used to rank extreme observations (see Remark 3, Section 2).
For the “central” part of the distribution, i.e. $\beta$ is fixed (and “not too small”), non-parametric estimation of density level sets has been studied in depth in the literature. Two approaches are used, the plug-in approach using density estimation, see Baille et al. (2001) and Rigollet and Vert (2009), and the excess mass approach, see Müller and Sawitzki (1991); Polonik (1995); Tsybakov (1997). Our estimation problem and (hence) our approach are quite different from these.

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

### 2.2 Main Result

Consider a random sample $X_1, X_2, \ldots, X_n$ with $X_i \stackrel{d}{=} X$, for $i = 1, \ldots, n$; their common probability measure on $\mathbb{R}^d$ is denoted with $P$. Write $R_i$ for the radius $||X_i||$ and $W_i$ for the direction $X_i/||X_i||$ of $X_i$. We wish to estimate an extreme risk region of the form

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

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

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

It will turn out that $Q_n$ can be approximated by a properly inflated version of $S$. In fact, it follows from (2.1.6) that the risk regions are asymptotically homothetic as a function of $p$, for small values of $p$. Define $H(s) = 1 - V(s) = P(R \leq s)$ and $U(t) = H^{-1}(1 - \frac{1}{t})$. Note that $U$ is regularly varying at infinity with index $1/\alpha$.

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

$$\frac{P(Q_n \triangle \tilde{Q}_n)}{p} \to 0 \quad (2.2.1)$$

($\triangle$ denotes “symmetric difference”) and is based on the above limit relations. The region $\tilde{Q}_n$ can therefore be estimated using extreme value theory. The first step is to establish an approximation of $\beta = \beta(p)$. Let
(b) \( k = k_n(< n) \) be a sequence of positive integers such that \( k \to \infty \) and \( k/n \to 0 \). The region \( Q_n \) is approximated by

\[
Q_n = \left\{ z : f(z) \leq \left( \frac{np}{k \nu(S)} \right)^{\frac{d}{n}} \frac{1}{\frac{n}{k} \left( U(\frac{z}{k}) \right)^{\frac{d}{n}}} \right\}.
\]

Next we approximate \( \tilde{Q}_n \) by a further region \( \tilde{Q}_n \) defined in terms of the limit density \( q \) rather than \( f \):

\[
\tilde{Q}_n = U\left( \frac{n}{k} \right) \left( \frac{k \nu(S)}{np} \right)^{\frac{d}{n}} S.
\]  

(2.2.2)

Indeed, \( S \) and this approximation of \( Q_n \) are homothetic.

Write

\[
B_{r,A} = \{ z : ||z|| \geq r, z/||z|| \in A \},
\]

for a Borel set \( A \) on \( \Theta \). Clearly, \( B_{r,A} = rB_{1,A} \) and hence \( \nu(B_{r,A}) = r^{-\alpha} \nu(B_{1,A}) \). The relation between the spectral measure \( \Psi \) and \( \nu \) is (cf. (2.1.1) and (2.1.3))

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

for a Borel set \( A \subset \Theta \). Recall that the spectral measure is a probability measure. The existence of a density \( q \) of \( \nu \) implies the existence of a density \( \psi \) of \( \Psi \), i.e.,

\[
\Psi(A) = \int_A \psi(w) d\lambda(w)
\]

where \( \lambda \) is the Hausdorff measure (surface area) on \( \Theta \) and

\[
q(rw) = \alpha r^{-\alpha-d} \psi(w).
\]

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

\[
S = \left\{ z = rw : r \geq (\alpha \psi(w))^{\frac{1}{\alpha+d}}, w \in \Theta \right\}
\]

and hence

\[
\nu(S) = \alpha^{-\frac{\alpha}{\alpha+d}} \int_{\Theta} (\psi(w))^{\frac{\alpha}{\alpha+d}} d\lambda(w).
\]

To estimate \( \tilde{Q}_n \), we need estimators for \( U(n/k), \alpha, S \) and \( \nu(S) \). From the above expressions for \( S \) and \( \nu(S) \), we see that this means that we have to estimate \( U(n/k), \alpha \)
and $\psi$. First we define
\[
\hat{U} \left( \frac{n}{k} \right) = R_{n-k:n}
\]
(the \((n-k)\)-th order statistic of the \(R_i, i = 1, \ldots, n\)). Since the tail of the distribution function of \(R_i\) is regularly varying with index \(-\alpha\), we can use one of the well-known estimators of the extreme value index \(1/\alpha\), based on the \(R_i, i = 1, \ldots, n\); see, e.g., Hill (1975); Smith (1987), and Dekkers et al. (1989). It remains to estimate $\psi$. Let \(K : [0, 1] \rightarrow [0, 1]\) be a continuous and non-increasing (kernel) function with \(K(0) = 1\) and \(K(1) = 0\). For \(w \in \Theta\) define an estimator of $\psi(w)$ by
\[
\hat{\psi}_n(w) = \frac{c(h, K)}{k} \sum_{i=1}^{n} K \left( \frac{1 - w^T W_i}{h} \right) 1_{[R_i > R_{n-k:n}]},
\]
with \(0 < h < 1\) and
\[
c(h, K) = \left( \int_{C_w(h)} K \left( \frac{1 - v^T w}{h} \right) d\lambda(v) \right)^{-1}, \quad C_w(h) = \{v \in \Theta : v^T v \geq 1 - h\};
\]

For estimating $Q_n$ it suffices to estimate $\tilde{Q}_n$, see (2.2.1). Hence, in view of (2.2.2), we define
\[
\hat{Q}_n = \hat{U} \left( \frac{n}{k} \right) \left( \frac{k \nu(S)}{np} \right)^{1/\hat{\alpha}} \hat{S}, \quad (2.2.3)
\]
with
\[
\hat{S} = \left\{ z = r w : r \geq \left( \frac{\hat{\psi}_n(w)}{n+\hat{\alpha}} \right)^{\frac{1}{n+\hat{\alpha}}}, w \in \Theta \right\},
\]
and
\[
\nu(S) = \hat{\alpha}^{-\frac{\hat{\alpha}}{n+\hat{\alpha}}} \int_{\Theta} (\hat{\psi}_n(w)) \frac{d}{n+\hat{\alpha}} d\lambda(w).
\]
In the definition of the set \(S\), the choice of the value 1 was not motivated. We could have taken any number \(c > 0\) instead. Such an alternative definition of \(S\) would lead to exactly the same estimator \(\tilde{Q}_n\), which shows that the value 1 plays no role.

Assume
\[
(c) \quad \lim_{t \to \infty} \frac{U(t)}{t^{1/\alpha}} = c, \quad \text{for some } c \in (0, \infty).
\]
Note that this simple condition is weaker than the usual second order condition with negative second order parameter $\rho$ (see, e.g., Theorem 4.3.8 in de Haan and Ferreira (2006)); indeed, there exist functions $U$ with $\rho = 0$ that satisfy condition (c).

**Theorem 2.2.1.** Let $p \to 0$ as $n \to \infty$. Assume conditions (a), (b), (c) hold and that $\hat{\alpha}$ is such that $\sqrt{k}(\hat{\alpha} - \alpha) = O_p(1)$. Also assume that $(\log np)/\sqrt{k} \to 0$, $h \to 0$, and $k/(c(h, K) \log k) \to \infty$, as $n \to \infty$. Then we have

$$\frac{P(\hat{Q}_n \triangle Q_n)}{p} \overset{p}{\to} 0, \quad \text{as } n \to \infty,$$

and hence

$$\frac{P(\hat{Q}_n)}{p} \overset{p}{\to} 1.$$

**Remark 1** The tuning parameter $k$ is used in the estimators of $\alpha, U(n/k)$ and $\psi$. It is important to be able to choose three different values for $k$, denoted with $k_\alpha, k_U$ and $k_\psi$, respectively. (Note that “good” values of $k_\alpha$ and $k_U$ are determined by the tail of $H$ - the distribution function of $R_1$ - whereas a good $k_\psi$ is determined by the conditional distribution of $W_1$, given that $R_1 > r$, for large $r$.) If we adapt the conditions of the theorem, in particular if (b) holds for $k_\alpha, k_U, k_\psi$, and if $(\log np)/\sqrt{k_\alpha} \to 0$, $k_\psi/(c(h, K) \log k_\psi) \to \infty$, and $(\log k_U)/\sqrt{k_\alpha} \to 0$, then (2.2.4) remains true for the generalized estimator that allows for the aforementioned different $k$-values. We will use this generalized estimator in the simulation study and the real data application.

The actual choice of these $k$-values is a notorious problem in extreme value theory. A solution of this problem is far beyond the scope of the present paper. We will only give heuristic guidelines here. First consider the estimation of $\alpha$. Plot $\hat{\alpha}$ as a function of $k$. Now find the first stable, i.e. approximately constant, region in the graph of this function. This vertical level is the final estimate of $\alpha$. It is also possible to use (complicated) asymptotically optimal procedures, see, e.g., Danielsson et al. (2001). Once the estimate $\hat{\alpha}$ is fixed, we plot $\hat{U}(\frac{n}{k}) (\frac{k}{n})^{1/\hat{\alpha}}$ against $k$ and we search for the first stable part in this graph. The vertical level is now the estimate of the constant $c$ in condition (c). Observe that $\hat{U}(\frac{n}{k}) (\frac{k}{n})^{1/\hat{\alpha}}$ is a building block of $\hat{Q}_n$, so we do not need to estimate $U(\frac{n}{k})$ separately. Also observe that we do not (need to) determine $k_\alpha$ and $k_U$, but only a region of good values. Finally, using again the already fixed $\hat{\alpha}$, we plot $\hat{\nu}(S)$ as a function of $k$ and again we search for the first stable region; we take $k_\psi$ to be the midpoint of this region of $k$-values.
Remark 2 The class of multivariate regularly varying distributions is quite large. It contains, e.g., all elliptical distributions with a heavy tailed radial distribution and all distributions in the domain of a sum-attraction of a multivariate (non-normal) stable distribution. It seems natural, however, to try to extend the assumption of multivariate regular variation to the case of non-equal tail indices \( \alpha \). It is an important feature of the present model that all directions are equally important: the marginal distributions do not play a special role. An extension to non-equal tail indices would be possible in principle, but it will be of limited value since it only works if marginal transformations lead to the present model. Also note that basically all linear combinations of the components inherit the lowest of the marginal tail indices: the tail index is not a smooth function of the direction (if it is not constant). Moreover, the statistical theory that will be needed will be challenging and will lead to a new and different project.

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

2.3 Simulation study

In this section a detailed simulation study is performed in order to investigate the finite sample performance of our estimator (with \( 1/\alpha \) estimated using the moment estimator of Dekkers et al. (1989) and with \( K(u) = 1 - u \)). We consider five multivariate distributions.

- The bivariate Cauchy distribution with density

\[
f(x, y) = \frac{1}{2\pi(1 + x^2 + y^2)^{3/2}}, \quad (x, y) \in \mathbb{R}^2.
\]  

(2.3.1)

This is a very heavy tailed density, with \( \alpha = 1 \) and \( \psi(w) = 1/(2\pi) \), for \( w \in \Theta \).

- The trivariate Cauchy distribution with density

\[
f(x, y, z) = \frac{1}{\pi^2(1 + x^2 + y^2 + z^2)^3}, \quad (x, y, z) \in \mathbb{R}^3.
\]  

(2.3.2)
This is also a very heavy tailed density, with \( \alpha = 1 \) and \( \psi(w) = 1/(4\pi) \), for \( w \in \Theta \).

- A bivariate elliptical distribution with density \((r_0 \approx 1.2481)\)
  \[
  f(x, y) = \begin{cases} 
  \frac{3}{10\pi} r_0^3 (1 + r_0^2)^{-3/2}, & x^2/4 + y^2 < r_0^2, \\
  3(0(x^2+y^2)^2-32x^2y^2) & x^2/4 + y^2 \geq r_0^2.
  \end{cases}
  \]
  \(\text{(2.3.3)}\)

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

- A bivariate “clover” distribution with density \((r_0 \approx 1.2481)\)
  \[
  f(x, y) = \begin{cases} 
  \frac{3}{10\pi} r_0^3 (1 + r_0^2)^{-3/2} \left( 5 + \frac{3(x^2+y^2)^2-32x^2y^2}{r_0(x^2+y^2)^{3/2}} \right), & x^2+y^2 < r_0^2, \\
  10\pi (1+(x^2+y^2)^2)^{3/2}, & x^2+y^2 \geq r_0^2.
  \end{cases}
  \]
  \(\text{(2.3.4)}\)

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

- A bivariate asymmetric shifted distribution with density \((r_0 \approx 1.2331, \tilde{r}(x, y) := r_0 \vee ((x + y)^2 + y^2)^{1/2})\)
  \[
  f(x, y) = \begin{cases} 
  \frac{\tilde{r}^2(x, y)}{6\pi (1+r^2(x,y))^{3/4}} \left[ \frac{3 + \frac{x+y}{\tilde{r}(x, y)}}{3 + \frac{(x+y)^2-3(x+y)y^2}{\tilde{r}^2(x, y)}} \right], & y \geq 0, \\
  \frac{\tilde{r}^2(x, y)}{6\pi (1+r^2(x,y))^{3/4}} \left[ \frac{3 + \frac{x+y}{\tilde{r}(x, y)}}{3 + \frac{(x+y)^2-3(x+y)y^2}{\tilde{r}^2(x, y)}} \right], & y < 0.
  \end{cases}
  \]
  \(\text{(2.3.5)}\)

This distribution is not symmetric and the “center” is not the origin, but \((-5, 0)\): 
\( \alpha = 1 \) and \( \psi(w_1, w_2) = \frac{1}{6\pi}(3 + w_1), \) if \( w_2 \geq 0 \), and \( \psi(w_1, w_2) = \frac{1}{6\pi}(3 + 4w_1^2 - 3w_1) \), if \( w_2 < 0 \), \( w = (w_1, w_2) \in \Theta \).

First we simulated single data sets of size 5000 of the bivariate Cauchy distribution, the elliptical distribution in (2.3.3), the clover distribution in (2.3.4), and the asymmetric shifted distribution in (2.3.5). We computed the true and estimated risk regions for \( p = 1/2000, 1/5000 \) or \( 1/10,000 \). This is depicted in Figure 2.1. We see that the estimated regions are relatively close to the true risk regions. It is interesting to note that the \( p \)-value (see Remark 3) of the largest observation for the Cauchy sample is \( 0.0000209 \), which is about \( 1/n \). This shows that this observation is a typical one. (Looking at the data only, one might want to conclude that this observation is an outlier.) Also note that for the bivariate Cauchy distribution, for, e.g., \( p = 1/10,000 \), the density \( f \) at the boundary of the true risk region is less than \( 10^{-12} \). This emphasizes that we
are estimating in an “almost empty” part of the plane and that a fully nonparametric procedure could not work here.

\[ \text{Cauchy Density, } n=5000, \ p=1/2000, 1/5000, 1/10000 \]

\[ \text{Elliptical Density, } n=5000, \ p=1/2000, 1/10000 \]

\[ \text{Clover Density, } n=5000, \ p=1/2000, 1/10000 \]

\[ \text{Asymmetric Shifted Density, } n=5000, \ p=1/2000, 1/10000 \]

**Figure 2.1:** True and estimated risk regions based on one sample of size 5000 from the bivariate Cauchy distribution, the elliptical distribution in (2.3.3), the clover distribution in (2.3.4), and the asymmetric shifted distribution in (2.3.5).

In addition, we simulated one sample of the bivariate distribution with independent \( t_3 \)-components. This distribution does not satisfy condition (a), since the spectral measure is discrete and concentrated on the intersection of the coordinate axes with the unit circle. We also simulated one sample of a bivariate “logarithmic” distribution
with $\alpha = 1$ and uniform spectral measure, but where the radial distribution satisfies $U(t)/(t \log t)$ tends to a constant and hence $U(t)/t \to \infty$ as $t \to \infty$, i.e. this distribution does not satisfy condition (c). Although both distributions do not satisfy our conditions, we see nevertheless satisfactory behavior of the estimator in Figure 2.2. In the left panel the estimated region has about the right size and the difficult shape is approximated reasonably well; in the right panel, we see that both the shape and the size are approximated quite well.

Figure 2.2: True and estimated risk regions based on one sample of size 5000 from the bivariate distribution with independent $t_3$-components and the logarithmic distribution.

After this visual assessment of our estimator based on one sample at a time, we now investigate its performance based on 100 simulated samples of size 5000. We will compare our estimator (denoted EVT) to a non-parametric and to a more parametric estimator. The non-parametric estimator is only defined in case $p = 1/n$ and tries to mimic the largest order statistic as an estimator of the $(1 - 1/n)$-th quantile in the univariate case. It aims at elliptical level sets. It is defined as follows. First calculate the smallest ellipsoid containing half of the data, the so-called MVE. Then inflate this ellipsoid, such that the “largest” observation lies on its boundary. Now the region outside this ellipsoid is the estimator.
For $d = 2$, the more parametric estimator is defined similarly to $\hat{Q}_n$ in (2.2.3), but (only) the estimation of $(\nu(S))^{1/\alpha} S$ is done parametrically. Therefore, this estimator has the same size as $\hat{Q}_n$, but a different shape. (Note that the fully parametric estimator based on multivariate normality would have a very bad performance.)

Take the $k$ observations with radius $R_i > R_{n-k,n}$ and consider the transformed data $(R_i / R_{n-k,n}, W_i)$. In line with the limit result in (2.1.1), assume that these data have a “distribution” $(\cdot)^{-\alpha} \Psi$, where $\Psi$ depends on a parameter $\rho$. To be precise we assume for the density

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

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

The following table shows for the three different estimators the median of the 100 relative errors $P(\hat{Q}_n \triangle Q_n)/p$ for $p = 1/5000$ (p1) and 1/10,000 (p2).

<table>
<thead>
<tr>
<th>Density</th>
<th>EVT p1</th>
<th>Par p1</th>
<th>NP p1</th>
<th>EVT p2</th>
<th>Par p2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biv. Cauchy</td>
<td>0.28</td>
<td>0.29</td>
<td>0.72</td>
<td>0.31</td>
<td>0.32</td>
</tr>
<tr>
<td>Triv. Cauchy</td>
<td>0.22</td>
<td>–</td>
<td>0.54</td>
<td>0.24</td>
<td>–</td>
</tr>
<tr>
<td>Elliptical</td>
<td>0.36</td>
<td>0.51</td>
<td>0.80</td>
<td>0.39</td>
<td>0.54</td>
</tr>
<tr>
<td>Clover</td>
<td>0.44</td>
<td>0.57</td>
<td>0.58</td>
<td>0.49</td>
<td>0.61</td>
</tr>
<tr>
<td>Asymm. shifted</td>
<td>0.26</td>
<td>0.27</td>
<td>0.61</td>
<td>0.30</td>
<td>0.32</td>
</tr>
</tbody>
</table>

**Table 2.1:** Median of the relative errors $P(\hat{Q}_n \triangle Q_n)/p$ of the three estimators, for $p = 1/5000$ (p1) and 1/10,000 (p2).

In Figure 2.3 boxplots are shown of the relative error $P(\hat{Q}_n \triangle Q_n)/p$ for $p = 1/5000$ (p1) and 1/10,000 (p2). From this table and figure we see a good performance of our estimator. Its behavior does not change much if $p$ changes from 1/5000 to 1/10,000. The parametric estimator performs reasonably well, but it is outperformed by our estimator, in particular for the elliptical and clover densities. Recall that this estimator can be seen as a modification of our estimator, since it uses the same estimated inflation factor, but the shape is estimated differently. We see a moderate performance of the non-parametric estimator; also, it cannot be adapted to $p = 1/10,000$. Given that the estimation of these extreme risk regions is a statistically difficult problem, we see decent behavior of the three estimation methods. Obviously the parametric and the non-parametric estimator do not perform well if the parametric part of the model is not adequate or if the shape of the region is not elliptical, respectively. The EVT estimator,
Figure 2.3: Boxplots of $P(\hat{Q}_n \triangle Q_n)/p$ for the here proposed estimator and for the parametric and the nonparametric estimator, based on 100 simulated data sets of size 5000 from the five presented densities for $p = 1/5000$ (p1) and $1/10,000$ (p2).
presented in this paper, does not suffer from these shortcomings and performs well for many multivariate distributions.

2.4 Application

In this section an application of our method to foreign exchange rate data is presented. The data are the daily exchange rates of Yen-Dollar and Pound-Dollar from January 4, 1999 to July 31, 2009. Consider the daily log returns given by $X_{i,j} = \log \left( \frac{Y_{i+1,j}}{Y_{i,j}} \right)$, with $i = 1, \ldots, 2664$, $j = 1, 2$, and $Y_{i,1}$ is the daily exchange rate of the Yen to the Dollar and $Y_{i,2}$ of the Pound to the Dollar. First, we check the equality of the extreme value indices (the reciprocals of the tail indices) of the right and left tails of both marginal distributions and that of the radius. This yields 5 extreme value indices; the 5 estimates in increasing order are: 0.141, 0.191, 0.223, 0.242, 0.256. Hence the maximal difference is 0.115. Based on the asymptotic normality of the moment estimator of the extreme value index, we compute an approximate upper bound for the maximal difference of the 5 estimators under the null hypothesis of equality: 0.264. Hence there is no evidence that the 5 extreme value indices are different. Other exchange rate data sets share this property. There are also economic arguments supporting this claim. Therefore we estimate $\alpha$ based on the radius and find $\hat{\alpha} = 3.90$. As a next step, we estimate the density $\psi$ of the spectral measure. The estimate $\hat{\psi}_n$ is depicted in Figure 4; it is almost periodic with period $\pi$. This yields that the estimated extreme risk region is not like a circle, but more like an ellipse. The location of the maxima of $\hat{\psi}_n$ correspond to the major axis of the region. We estimate the extreme risk regions for $p = 1/2000, 1/5000$ and $1/10,000$, see Figure 5. For risk management of financial institutions in the U.S. it is important to know which extreme exchange rate returns w.r.t. the Pound and
the Yen can occur and which returns essentially never occur. Our estimate answers
this question. More specifically, points on the boundary of the estimated extreme risk
region can be used as multivariate stress test scenarios. A scenario on the intersection
of the major axis of the ellipse-like extreme risk region and the boundary of the region
corresponds to a larger shock than a scenario on the intersection of the minor axis of the
extreme risk region and its boundary, but our method shows that their “extremeness”
is about the same.

![Figure 2.5: Estimated extreme risk regions of exchange rate returns.](image)

### 2.5 Proofs

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

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

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

(2.5.1)
Moreover
\[
\lim_{t \to \infty} \frac{\mathbb{P}(X \in tB)}{V(t)} = \int_B q(z) dz, \tag{2.5.2}
\]
for any Borel set \(B\) bounded away from the origin. Define \(q_t(z) = t(U(t))^d f(U(t)z)\). Then
\[
\lim_{t \to \infty} \sup_{||z|| \geq \varepsilon} |q_t(z) - lq(z)| = 0. \tag{2.5.3}
\]

Let \(\tilde{h}\) be the density of \(H\), then
\[
\lim_{t \to \infty} \frac{\tilde{h}(t)}{t^{-1}V(t)} = \frac{\alpha}{l}. \tag{2.5.4}
\]

**Proof** For any \(||z|| \geq \varepsilon > 0\) (cf. Theorem 2.1 in de Haan and Resnick (1987)),
\[
\begin{align*}
&\left| \frac{f(tz)}{t^{-d}V(t)} - q(z) \right| \\
= &\left| \frac{f\left(\frac{t||z||}{||z||}z\right)}{t||z||^{-d}V(t||z||)} - \frac{f(t||z||z)}{t||z||^{-d}V(t||z||)} \right| \\
\leq &\left| \frac{f(t||z||z)}{t||z||^{-d}V(t||z||)} - q\left(\frac{z}{||z||}\right) \right| + \left| \frac{f(t||z||z)}{t||z||^{-d}V(t||z||)} - \frac{f(t||z||z)}{t||z||^{-d}V(t||z||)} \right|
\end{align*}
\]
Then (2.5.1) follows from condition (a).

Let a Borel set \(B\) be such that \(B \subset \{||z|| \geq \gamma\}\), for some \(\gamma > 0\). Then for \(z \in B\) and sufficiently large \(t\), \(f(tz)/t^{-d}V(t)\) is bounded by \(q(||z||^{-1}z)||z||^{-d/2-d}\). Hence (2.5.2) holds by Lebesgue’s dominated convergence theorem.

We have from (2.5.2), as \(t \to \infty\),
\[
tV(U(t)) = \frac{V(U(t))}{\mathbb{P}(R \geq U(t))} \to l.
\]
Hence (2.5.1) implies, uniformly for \(||z|| \geq \varepsilon\),
\[
q_t(z) = tV(U(t)) \frac{f(U(t)z)}{(U(t))^{-d}V(U(t))} \to lq(z).
\]
Note that
\[ 1 - H(t) = \mathbb{P}(R > t) = \int_t^\infty \int_{\Theta} f(rw)d\lambda(w) r^{d-1}dr. \]

By taking derivatives, (2.5.1), and the homogeneity of \( q \), we obtain
\[ \lim_{t \to \infty} \frac{\tilde{h}(t)}{t^{-1}V(t)} = \int_{\Theta} q(w)d\lambda(w) = \alpha \int_{\{||z|| \geq 1\}} q(z)dz = \alpha/l. \]
□

We now see that (2.1.5) and (2.1.6) hold with \( V = 1 - H \). From now on we will make the choice \( V = 1 - H \) and hence \( l = 1 \). Note that with this choice the relations (2.1.3) (with \( \nu(B) = \int_B q(z)dz \)) and (2.1.4) readily follow from (2.5.2).

**Corollary 1** For all Borel sets \( B \) with positive distance from the origin,
\[ \lim_{t \to \infty} tP(U(t)B) = \nu(B), \tag{2.5.5} \]
and
\[ \lim_{n \to \infty} \frac{\nu(S)}{p} P\left(U\left(\frac{n}{k}\right)\left(\frac{kn(S)}{np}\right)^\frac{1}{n} B\right) = \nu(B). \tag{2.5.6} \]

**Proof** From \( \mathbb{P}(R \geq U(t)) = 1/t \) and (2.1.3) we obtain (2.5.5). It follows from (c) that
\[ \frac{U\left(\frac{\nu(S)}{p}\right)}{U\left(\frac{n}{k}\right)\left(\frac{kn(S)}{np}\right)^\frac{1}{n}} \to 1. \tag{2.5.7} \]
This yields (2.5.6). □

**Lemma 2** For each \( \varepsilon > 0 \) there exists a \( \delta > 0 \) and \( t_0 > 0 \) such that for \( t > t_0 \)
\[ \left\{ z : \frac{f(tz)}{t^{-d}V(t)} \leq \varepsilon \right\} \subset \left\{ z : ||z|| > \delta \right\}. \]
Proof It is sufficient to prove \( \{ z : ||z|| \leq \delta \} \subset \{ z : f(tz)/(t^{-d}V(t)) > \varepsilon \} \). First, by (2.1.6) and the continuity of \( q \), for some \( c_1 > 0 \), there exists \( s_0 > 0 \) such that for \( s > s_0 \)

\[
\inf_{w \in \Theta} \frac{f(sw)}{s^{-d}V(s)} \geq c_1
\]

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

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

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

(i) \( t||z|| > s_0 \), then

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

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

\[
\frac{f(tz)}{t^{-d}V(t)} > \varepsilon.
\]

\( \square \)

Lemma 3 For \( \varepsilon > 0 \) and large \( n \),

\[
\tilde{Q}_n \subset U \left( \frac{\nu(S)}{p} \right) \{ z : q(z) \leq 1 + \varepsilon \}
\]

and

\[
\tilde{Q}_n \supset U \left( \frac{\nu(S)}{p} \right) \{ z : q(z) \leq 1 - \varepsilon \}.
\]

Proof Recall that \( \tilde{Q}_n = \left\{ z : f(z) \leq \left( \frac{np}{k
u(S)} \right)^{1+\frac{d}{n}} \cdot \frac{1}{U^{1+\frac{d}{n}}(U^{1+\frac{d}{n}})^d} \right\} \). It follows from (2.5.7) that for \( n \) large enough and \( \varepsilon > 0 \),

\[
\tilde{Q}_n = U \left( \frac{\nu(S)}{p} \right) \left\{ z : f \left( U \left( \frac{\nu(S)}{p} \right) z \right) \leq \left( \frac{np}{k
u(S)} \right)^{1+\frac{d}{n}} \cdot \frac{1}{U^{1+\frac{d}{n}}(U^{1+\frac{d}{n}})^d} \right\}
\]

\[
= U \left( \frac{\nu(S)}{p} \right) \left\{ z : q_{\nu(S)/p}(z) \leq \left( \frac{np}{k
u(S)} \right)^{\frac{d}{n}} \cdot \left( U \left( \frac{\nu(S)}{p} \right) \right)^{-d} \cdot \left( U \left( \frac{\nu(S)}{p} \right) \right)^d \right\}.
\]
\[
\subset U \left( \frac{\nu(S)}{p} \right) \left\{ z : q_{\nu(S)/p}(z) \leq 1 + \varepsilon_1 \right\}.
\]

Now Lemma 2 implies \( \{ z : q_{\nu(S)/p}(z) \leq 1 + \varepsilon_1 \} \subset \{ z : ||z|| > \delta \} \), hence we have by (2.5.3)
\[
\bar{Q}_n \subset U \left( \frac{\nu(S)}{p} \right) \{ z : q(z) \leq 1 + \varepsilon_1 \}.
\]

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

**Lemma 4** For \( \varepsilon > 0 \) and large \( n \),
\[
\tilde{Q}_n \subset U \left( \frac{\nu(S)}{p} \right) \{ z : q(z) \leq 1 + \varepsilon \}
\]
and
\[
\tilde{Q}_n \supset U \left( \frac{\nu(S)}{p} \right) \{ z : q(z) \leq 1 - \varepsilon \}.
\]

**Proof** Recall that \( \tilde{Q}_n = U \left( \frac{n}{k} \right) \left( \frac{\nu(S)}{np} \right)^{\frac{1}{p}} \left\{ z : q(z) \leq 1 \right\} \).

Put \( T_n = U \left( \frac{\nu(S)}{p} \right)^{-1} U \left( \frac{n}{k} \right) \left( \frac{\nu(S)}{np} \right)^{\frac{1}{p}} \), then
\[
\bar{Q}_n = U \left( \frac{\nu(S)}{p} \right) \{ T_n z : q(z) \leq 1 \} = U \left( \frac{\nu(S)}{p} \right) \{ T_n z : q(T_n z) \leq T_n^{-d-\alpha} \}
\]
\[
= U \left( \frac{\nu(S)}{p} \right) \{ z : q(z) \leq T_n^{-d-\alpha} \}.
\]

Since \( T_n \to 1 \) as \( n \to \infty \) by (2.5.7), the result follows.

**Proposition 1** We have
\[
\lim_{n \to \infty} \frac{P(Q_n \triangle \tilde{Q}_n)}{p} = 0.
\]

**Proof** Note that \( P(Q_n \triangle \tilde{Q}_n) \leq P(Q_n \triangle \tilde{Q}_n) + P(\tilde{Q}_n \triangle \tilde{Q}_n) \). Observe that \( Q_n \subset \bar{Q}_n \) or \( \bar{Q}_n \subset Q_n \), hence \( P(Q_n \triangle \tilde{Q}_n) \leq |p - P(\tilde{Q}_n)| \). By Lemma 3 and Corollary 1, for any \( \varepsilon > 0 \) and large \( n \)
\[
\frac{\nu(S)}{p} P(Q_n) \leq \frac{\nu(S)}{p} P \left( U \left( \frac{\nu(S)}{p} \right) \{ z : q(z) \leq 1 + \varepsilon \} \right)
\]
Chapter 2. Estimation of extreme risk regions

\[ \nu(\{ z : q(z) \leq 1 + \varepsilon \}) \]
\[ = \nu(\{ z : q(z(1 + \varepsilon)^{1/(d + \alpha)}) \leq 1 \}) \]
\[ = \nu(\{ (1 + \varepsilon)^{-1/(d + \alpha)} z : q(z) \leq 1 \}) \]
\[ = (1 + \varepsilon)^{\alpha/(d + \alpha)} \nu(S). \]

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

Similarly we have \( \liminf_{n \to \infty} \frac{P(\bar{Q}_n)}{p} \geq (1 - \varepsilon)^{\alpha/(2 + \alpha)}. \) Hence \( \lim_{n \to \infty} \frac{P(\bar{Q}_n)}{p} = 1, \) i.e. \( \lim_{n \to \infty} \frac{P(\bar{Q}_n \bigtriangleup \tilde{Q}_n)}{p} = 0. \)

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

\[ \frac{\nu(S)}{p} P(\bar{Q}_n \bigtriangleup \tilde{Q}_n) \leq \frac{\nu(S)}{p} P\left( U\left( \frac{\nu(S)}{p} \right) \{ z : 1 - \varepsilon \leq q(z) \leq 1 + \varepsilon \} \right) \]
\[ \Rightarrow \nu(\{ z : 1 - \varepsilon \leq q(z) \leq 1 + \varepsilon \}) \]
\[ = \nu(S) \left( (1 + \varepsilon)^{\alpha/(d + \alpha)} - (1 - \varepsilon)^{\alpha/(d + \alpha)} \right) \]

Hence \( \lim_{n \to \infty} \frac{P(\bar{Q}_n \bigtriangleup \tilde{Q}_n)}{p} = 0. \)

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

**Proposition 2** As \( n \to \infty, \)

\[ \sup_{w \in \Theta} \left| \frac{\hat{\psi}_n(w) - \psi(w)}{P} \right| \to 0. \]

**Proof** It is easy to see that, for any \( \eta > 0, \) there exists a function

\[ K^* = \sum_{j=1}^{m} \alpha_j 1_{[r_{j-1}, r_j)], \]
Chapter 2. Estimation of extreme risk regions

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

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

Write \(U_i = 1 - H(R_i), i = 1, \ldots, n\), and denote the corresponding order statistics with \(U_{i:n}\). Let \(P\) be the probability measure on \(\Theta \times (0,1)\) corresponding to \((W_1, U_1)\) and let \(\tilde{P}_n\) be the empirical measure of the \((W_i, U_i) i = 1, \ldots, n\). Define

\[
\psi^*_{n,j}(w) = \frac{nc(h,K)}{k} \tilde{P}_n(D_{w,j} \times (0, U_{k:n})),
\]

with \(D_{w,j} = \{ v \in \Theta : 1 - hr_j < w^T v \leq 1 - hr_{j-1} \}\). Observe that \(\psi^*_n(w) = \sum_{j=1}^m \alpha_j \psi^*_{n,j}(w)\).

Also write

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

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

\[
P \left( \sup_{w \in \Theta} \left| \psi_n(w) - \sum_{j=1}^m \alpha_j \psi_{n,j}(w) \right| \geq 2\varepsilon \right) \leq 2\varepsilon, \quad (2.5.8)
\]

\[
P \left( \sup_{w \in \Theta} \left| \sum_{j=1}^m \alpha_j (\psi_{n,j}(w) - c(h,K)\Psi(D_{w,j})) \right| \geq 2\varepsilon \right) \leq \varepsilon, \quad (2.5.9)
\]

\[
\sup_{w \in \Theta} \left| c(h,K) \sum_{j=1}^m \alpha_j \Psi(D_{w,j}) - \psi(w) \right| \leq \varepsilon. \quad (2.5.10)
\]

For \(w \in \Theta\) and \(\delta \in (0,1)\), write \(C_\delta = \{ C_w(a) : w \in \Theta, a \leq \delta \}\). Note that, as \(n \to \infty\),

\[
\sup_{C \in C_1, 0 < s \leq 2} \frac{1}{\lambda(C)} \left| \frac{n}{k} \tilde{P}(C \times (0, sk/n]) - s \Psi(C) \right| \to 0. \quad (2.5.11)
\]

This readily follows from

\[
\frac{n}{k} \tilde{P}(C \times (0, sk/n]) = \frac{n}{k} \mathbb{P} \left( W \in C, R \geq U \left( \frac{n}{sk} \right) \right)
\]
Chapter 2. *Estimation of extreme risk regions*  

\[
= \frac{n}{k} \int_{U(\frac{r}{n})} \int_C \frac{f(rw)}{r^{-d}V(r)} d\lambda(w)r^{-1}V(r)dr
\]

and (2.5.1) and (2.5.4).

Now we prove (2.5.8). It is easy to show that

\[
c(h, K) = \left( \frac{2\pi^{(d-1)/2}}{(d-1)/2} \right) \left( 1 - \frac{1}{h} \right) \left( 1 - \frac{t^2}{h^2} \right)^{(d-3)/2} \right)^{-1}
\]

and hence

\[
\limsup_{h \downarrow 0} c(h, K) \lambda(C_w(h)) < \infty. \tag{2.5.12}
\]

We have

\[
\left| \hat{\psi}_n(w) - \psi^*_n(w) \right| = \frac{c(h, K)}{k} \sum_{i=1}^n \left| K \left( \frac{1 - w^T W_i}{h} \right) - K^* \left( \frac{1 - w^T W_i}{h} \right) \right| 1_{[R_i > R_n - k, n]}
\]

\[
\leq \frac{c(h, K)}{k} \sum_{i=1}^n \eta 1_{[W_i \in C_w(h), R_i > R_n - k, n]}
\]

\[
\leq \eta \frac{nc(h, K)}{k} \tilde{P}(C_w(h) \times (0, U_{k,n}])
\]

\[
+ \eta \frac{nc(h, K)}{k} \left| (\tilde{P}_n - \tilde{P}) (D_w, j \times (0, U_{k,n}]) \right|. \tag{2.5.14}
\]

By (2.5.11), for \( \eta \) small enough the first term is less than \( \varepsilon \), with probability tending to one, uniformly in \( w \in \Theta \). Also,

\[
\left| \psi^*_n(w) - \sum_{j=1}^m \alpha_j \psi_{n,j}(w) \right|
\]

\[
\leq \sum_{j=1}^m \alpha_j \left| \psi^*_n(w) - \psi_{n,j}(w) \right|
\]

\[
\leq \sum_{j=1}^m \alpha_j \frac{nc(h, K)}{k} \left| (\tilde{P}_n - \tilde{P}) (D_{w,j} \times (0, U_{k,n}] \right|. \tag{2.5.15}
\]
From (2.5.15), (2.5.14) and (2.5.12), we see that for a proof of (2.5.8) it remains to show that

$$\frac{n}{k\lambda(C_w(h))} \sup_{w \in \Theta} \sup_{0 < a \leq 1} \left| \left( \tilde{P}_n - \tilde{P} \right) (C_w(ah) \times (0, U_{kn}) \right| \to 0.$$ 

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

$$C_w(ah) \in C_{w_l}(ch), \text{ for some } l.$$ 

Hence for $\varepsilon_1 > 0$,

$$\mathbb{P} \left( \frac{n}{k\lambda(C_w(h))} \sup_{w \in \Theta} \sup_{0 < a \leq 1} \left| \left( \tilde{P}_n - \tilde{P} \right) (C_w(ah) \times (0, U_{kn}) \right| \geq \varepsilon_1 \right) \leq \mathbb{P} \left( \max_{1 \leq l \leq l_h} \sup_{C \in C_h} \sup_{0 < s \leq 2} \left| (\tilde{P}_n - \tilde{P})(C \times (0, sk/n) \right| \geq \varepsilon_1 k/n \lambda(C_w(h)) \right) + \mathbb{P}(U_{kn} > 2k/n) \leq \sum_{l=1}^{l_h} \mathbb{P} \left( \sup_{C \in C_h} \sup_{0 < s \leq 2} \left| (\tilde{P}_n - \tilde{P})(C \times (0, sk/n) \right| \geq \varepsilon_1 k/n \lambda(C_w(h)) \right) + \mathbb{P}(U_{kn} > 2k/n).$$

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

$$\mathbb{P} \left( \sup_{C \in C_h} \sup_{0 < s \leq 2} \left| (\tilde{P}_n - \tilde{P})(C \times (0, sk/n) \right| \geq b \right) \leq \sum_{r=\lceil \mu + b/3 \rceil}^{\lfloor \mu - b/3 \rfloor} \mathbb{P} \left( \sup_{C \in C_h} \sup_{0 < s \leq 2} \left| (\tilde{P}_n - \tilde{P})(C \times (0, sk/n) \right| \geq b \left| N = r \right) \right) \mathbb{P}(N = r) + \mathbb{P}(\left| N - \mu \right| \geq b/3)$$
\[
\leq \sum_{r=\lfloor \mu + b/3 \rfloor}^{r=\lfloor \mu - b/3 \rfloor} \mathbb{P} \left( \sup_{C \subset \mathcal{C}_h} \sup_{0 < s \leq 2} n \left( \tilde{P}_n - \frac{N}{\mu} \tilde{P} \right) (C \times (0, sk/n]) \bigg| b/2 \bigg| N = r \right) \mathbb{P}(N = r)
\]

\[
+ \sum_{r=\lfloor \mu - b/3 \rfloor}^{r=\lfloor \mu + b/3 \rfloor} \mathbb{P} \left( \sup_{C \subset \mathcal{C}_h} \sup_{0 < s \leq 2} n \left( \frac{(N - \mu)}{\mu} \tilde{P} (C \times (0, sk/n]) \bigg| b/2 \bigg| N = r \right) \mathbb{P}(N = r)
\]

\[
+ \mathbb{P}(\{|N - \mu| \geq b/3\})
\]

\[
\leq \sum_{r=\lfloor \mu + b/3 \rfloor}^{r=\lfloor \mu - b/3 \rfloor} \mathbb{P} \left( \sup_{C \subset \mathcal{C}_h} \sup_{0 < s \leq 2} r \left( \tilde{P}_{c,r} - \tilde{P}_c \right) (C \times (0, sk/n]) \bigg| b/2 \bigg| N = r \right) \mathbb{P}(N = r)
\]

\[
+ \mathbb{P}(\{|N - \mu| \geq b/3\}) + \mathbb{P}(\{|N - \mu| \geq b/3\})
\]

(2.5.16)

Note that the first probability of the second sum in the right side of (2.5.16) is equal to 0. From Bennett’s inequality (cf. Shorack and Wellner (1986) p. 851) it follows that for some constant \( c_1 \)

\[
\mathbb{P}(\{|N - \mu| \geq b/3\}) \leq 2 \exp \left( -\epsilon_1^2 c_1 \frac{k}{c(h, K)} \right).
\]

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

\[
\sum_{l=1}^{l_h} \mathbb{P}(\{|N - \mu| \geq b/3\}) = O \left( c(h, K) \exp \left( -\epsilon_1^2 c_1 \frac{k}{c(h, K)} \right) \right) = o(1).
\]

To complete the proof of (2.5.8), we need to consider the first sum in the right side of (2.5.16). For the first probability in there we use Corollary 2.9 in Alexander (1984), a good probability bound for empirical processes on VC classes. We obtain as an upper bound

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

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

\[
\sum_{l=1}^{l_h} \sum_{r=\lfloor \mu - b/3 \rfloor}^{r=\lfloor \mu + b/3 \rfloor} \mathbb{P} \left( \sup_{C \subset \mathcal{C}_h} \sup_{0 < s \leq 2} r \left( \tilde{P}_{c,r} - \tilde{P}_c \right) (C \times (0, sk/n]) \bigg| b/2 \bigg| N = r \right) \mathbb{P}(N = r)
\]
\[ \leq 16 \sum_{l=1}^{l_h} \sum_{r=[\mu-b/3]}^{\lfloor \mu+b/3 \rfloor} \exp \left( -\varepsilon_1^2 c_2 \frac{k}{c(h,K)} \right) \mathbb{P}(N = r) \]
\[ \leq 16 \sum_{l=1}^{l_h} \exp \left( -\varepsilon_1^2 c_2 \frac{k}{c(h,K)} \right) = o(1). \]

Next we show (2.5.9). From (2.5.12) and (2.5.11), we obtain for \( \varepsilon_2 > 0 \) small enough,
\[ \sup_{w \in \Theta} \left| \sum_{j=1}^{m} \alpha_j (\psi_{n,j}(w) - c(h,K)\Psi(D_{w,j})) \right| \]
\[ = \sup_{w \in \Theta} \left| \sum_{j=1}^{m} \alpha_j c(h,K) \left( n/k \tilde{P} (D_{w,j} \times (0,U_{k:n}) - \Psi(D_{w,j})) \right) \right| \]
\[ \leq \varepsilon_2 \sum_{j=1}^{m} \alpha_j c(h,K) \lambda(C_w(h)) + \sup_{w \in \Theta} \left| \sum_{j=1}^{m} \alpha_j c(h,K) (nU_{k:n}/k - 1) \Psi(D_{w,j}) \right| \]
\[ \leq \varepsilon + \frac{n}{k} U_{k:n} - 1 \sum_{j=1}^{m} \alpha_j c(h,K) \lambda(C_w(h)) \sup_{w \in \Theta} \psi(w) < 2\varepsilon, \]
with probability tending to one.

It remains to prove (2.5.10). It is readily seen that \( \int_{C_w(h)} K^* \left( \frac{1-w^T \nu}{h} \right) d\lambda(\nu) = \sum_{j=1}^{m} \alpha_j \lambda(D_{w,j}) \). Hence for \( \varepsilon_3 > 0 \) small enough
\[ \sup_{w \in \Theta} \left| c(h,K) \sum_{j=1}^{m} \alpha_j \Psi(D_{w,j}) - \psi(w) \right| \]
\[ \leq \sup_{w \in \Theta} \psi(w) \left| c(h,K) \sum_{j=1}^{m} \alpha_j \lambda(D_{w,j}) - 1 \right| + \varepsilon_3 c(h,K) \sum_{j=1}^{m} \alpha_j \lambda(D_{w,j}) \]
\[ \leq \sup_{w \in \Theta} \psi(w) \left| \frac{\int_{C_w(h)} K^* \left( \frac{1-w^T \nu}{h} \right) d\lambda(\nu)}{\int_{C_w(h)} K \left( \frac{1-w^T \nu}{h} \right) d\lambda(\nu)} - 1 \right| + \varepsilon_3 c(h,K) \lambda(C_w(h)) \sum_{j=1}^{m} \alpha_j \]
\[ \leq \eta c(h,K) \lambda(C_w(h)) \sup_{w \in \Theta} \psi(w) + \varepsilon_3 c(h,K) \lambda(C_w(h)) \sum_{j=1}^{m} \alpha_j \]
\[ \leq \varepsilon. \] \[ \square \]
From Proposition 2 and the consistency of $\hat{\alpha}$, we obtain immediately, as $n \to \infty$,

$$\hat{\nu}(S) \xrightarrow{p} \nu(S)$$

and, for $\varepsilon > 0$,

$$P((1 + \varepsilon)S \subset \hat{S} \subset (1 - \varepsilon)S) \to 1.$$  (2.5.17)

**Proposition 3** As $n \to \infty$,

$$P\left(\tilde{Q}_n \triangle \hat{Q}_n\right) \xrightarrow{p} 0.$$

**Proof** Note that as $n \to \infty$, we have

$$\tilde{U}\left(\begin{pmatrix} n \\ k \end{pmatrix}\right) / U\left(\begin{pmatrix} n \\ k \end{pmatrix}\right) \xrightarrow{p} 1,$$

$$\left(\hat{\nu}(S)\right)^{\frac{1}{\hat{\alpha}}} \xrightarrow{p} (\nu(S))^{\frac{1}{\hat{\alpha}}},$$

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

Combining these three limit relations we obtain

$$\tilde{U}\left(\begin{pmatrix} n \\ k \end{pmatrix}\right) \left(\frac{\nu(S)}{np}\right)^{\frac{1}{\alpha}} \xrightarrow{p} 1.$$

This and (2.5.17) yields that with probability tending to one, as $n \to \infty$,

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

Then,

$$P\left(\tilde{Q}_n \triangle \hat{Q}_n\right) \leq \frac{1}{p} P\left(U\left(\begin{pmatrix} n \\ k \end{pmatrix}\right) \left(\frac{\nu(S)}{np}\right)^{\frac{1}{\alpha}} \left(1 - \varepsilon\right)^2 S \setminus \left(1 + \varepsilon\right)^2 S\right),$$

and, by (2.5.6), the latter expression tends to

$$\nu((1 - \varepsilon)^2 S \setminus (1 + \varepsilon)^2 S) / \nu(S).$$
\[ \frac{\nu((1 - \varepsilon)^2 S)/\nu(S) - \nu((1 + \varepsilon)^2 S))/\nu(S)}{(S)} = (1 - \varepsilon)^{-2\alpha} - (1 + \varepsilon)^{-2\alpha}, \]

which in turn tends to 0, as \( \varepsilon \downarrow 0. \)

**Proof of Theorem 1** The result follows from Propositions 1 and 3.

**Acknowledgements** We thank two referees for many insightful comments that led to an improved version of the paper. We are grateful to Kees Koedijk, Roger Laeven, Ronald Mahieu, and Chen Zhou for discussions of the financial application.
Chapter 3

Bias Correction in Extreme Value Statistics with Index around Zero

[Based on joint work with Laurens de Haan and Chen Zhou]

Abstract Applying extreme value statistics in meteorology and environmental science requires accurate estimators on extreme value indices that can be around zero. Without having prior knowledge on the sign of the extreme value indices, the probability weighted moment (PWM) estimator is a favorable candidate. As most other estimators on the extreme value index, the PWM estimator bears an asymptotic bias. In this paper, we develop a bias correction procedure for the PWM estimator. Moreover, we provide bias-corrected PWM estimators for high quantiles and, when the extreme value index is negative, the endpoint of a distribution. The choice of $k$, the number of high order statistics used for estimation, is crucial in applications. The asymptotically unbiased PWM estimators allows the choice of higher level $k$, which results in a lower asymptotic variance. Moreover, since the bias-corrected PWM estimators can be applied for a wider range of $k$ compared to the original PWM estimator, one gets more flexibility in choosing $k$ for finite sample applications. All advantages become apparent in simulations and an environmental application on estimating “once per 10,000 years” still water level at Hoek van Holland, The Netherlands.
3.1 Introduction

Extreme value statistics has been widely applied in modeling and analyzing rare events in meteorology, environmental science, finance, among other fields. An example in application is estimating high quantiles with extremely low tail probabilities. The major difficulty in application is to produce accurate estimates of the extreme value index, denoted as $\gamma$ which characterizes the shape of a distribution function in the tail region. The case $\gamma > 0$ corresponds to the “heavy-tailed” distributions, while the case $\gamma < 0$ corresponds to distributions with finite right endpoint. Empirical literature has documented that random variables investigated in meteorology and environmental science exhibit generally extreme value indices around zero,\(^1\) including $\gamma = 0$. In other words, when analyzing meteorological and environmental variables, scientists tend not to make assumption on the distribution function as either heavy-tailed ($\gamma > 0$) or with finite endpoint ($\gamma < 0$). Estimators of a general $\gamma$ without having prior knowledge on its sign are necessary for such a situation.

Following the usual peak-over-threshold (POT) approach in extreme value analysis, with observations from a distribution that is in the max-domain of attraction, the excesses above a high threshold follows approximately the generalized Pareto distribution (GPD); see, e.g. de Haan and Ferreira (2006, Chap 3). The extreme value index is the shape parameter in the limiting GPD. Hosking and Wallis (1987) introduced the probability weighted moment (PWM) estimator for estimating the shape parameter in the GPD without having prior knowledge on the sign of the shape parameter. A modified version of the PWM estimator is then suitable for handling estimation of the extreme value index that is around zero, when the observations are from the max-domain of attraction. The estimator is widely used in meteorology and environmental applications. With independent and identically distributed (i.i.d.) observations $X_1, X_2, \cdots, X_n$ ranked as $X_{n,1} \leq \cdots \leq X_{n,n}$, we define the PWM estimator as

$$\hat{\gamma}_{pwm} := \frac{I_1 - 4I_2}{I_1 - 2I_2},$$

(3.1.1)

\(^1\)For example, for hourly surge level on the English east coast, Coles and Tawn (1991) find no significant evidence against $\gamma = 0$; for hourly maximum wind speed in Sheffield, UK, Coles and Walshaw (1994) estimate its $\gamma$ at -0.12; for daily rainfall in the south-west of England, Coles and Tawn (1996) get the $\gamma$ estimate at 0.066; for wave height and still water level on the Dutch coast, de Haan and de Ronde (1998) provide $\gamma$ estimates of the two at -0.0074 and -0.12 respectively; for daily rainfall in North Holland, The Netherlands, Buishand et al. (2008) obtain a $\gamma$ estimate at 0.1082.
where the probability weighted moments are given by

\[
I_q = \frac{1}{k} \sum_{i=1}^{k} \left( \frac{i}{k} \right)^{q-1} (X_{n,n-i+1} - X_{n,n-k}),
\]

(3.1.2)

for \( q = 1, 2, \cdots \). Here \( k := k(n) \) is an intermediate sequence such that \( k \to \infty \) and \( k/n \to 0 \) as \( n \to \infty \).

Similar to other estimators of the extreme value index, under a suitable second order condition (see Section 3.2.1) with \( \gamma < 1/2 \), \( \hat{\gamma}_{pwm} \) has the following asymptotic expansion,

\[
\hat{\gamma}_{pwm} - \gamma \approx \frac{N_\gamma}{\sqrt{k}} + A(n/k) \frac{(1 - \gamma)(2 - \gamma)}{(1 - \gamma - \rho)(2 - \gamma - \rho)},
\]

(3.1.3)

where \( N_\gamma \) is a mean zero normal random variable, \( A \) is a positive or negative function with \( A(t) \to 0 \) as \( t \to \infty \), and \( \rho \) is a parameter governing the second order behavior of the tail. From (3.1.3) we see that \( \sqrt{k}(\hat{\gamma}_{pwm} - \gamma) \xrightarrow{d} N_\gamma \) provided that \( \sqrt{k}A(n/k) \to 0 \) as \( n \to \infty \). The imposed condition requires a rather slow speed of convergence for the PWM estimator. When \( \sqrt{k}A(n/k) \to \lambda, \lambda \neq 0 \), a bias appears as

\[
\lambda \frac{(1 - \gamma)(2 - \gamma)}{(1 - \gamma - \rho)(2 - \gamma - \rho)}.
\]

Since the bias is an explicit function, the aim of this paper is to estimate the bias and subtract it from the original PWM estimator, thus creating a more efficient estimator.

Our bias correction procedure has the following advantages.

Firstly, for the application of the bias-corrected PWM estimator, larger values of \( k \) can be used than that for the original PWM estimator. This results in a lower asymptotic variance, which is at the level of \( 1/k \), cf. Remark 3.3.1.

Secondly, in applications the choice of \( k \) becomes less crucial since for a much wider range of \( k \)-values, the estimation stays at a stable level.

Thirdly, the bias-corrected PWM estimator is asymptotically unbiased in the sense that the asymptotic normal distribution has zero mean. All these features become apparent in the simulations and application.

Bias correction has been studied extensively in the literature of estimating the extreme value index. Most studies focus on the \( \gamma \) positive case, with a few papers focusing on the \( \gamma \) negative case. (i) The first generation of bias correction methods is based on an indirect approach which uses weighted combination of different estimators in order to cancel out the bias terms;\(^2\) see Peng (1998), Caeiro and Gomes (2002), Gomes

\(^2\)This approach is sometimes called the “generalized jackknife estimators” of extreme value index.
and Martins (2002, 2004), Gomes et al. (2000, 2002, 2004, 2005, 2007). (ii) The second generation of bias correction methods considers subtracting the bias term in the asymptotic distribution directly; see Caeiro et al. (2005), Gomes and Pestana (2007a), Gomes et al. (2007, 2008) and Caeiro and Gomes (2008). (iii) Another category of methods comes from the maximum likelihood procedure: by applying a maximum likelihood procedure to a second order expansion of an explicit tail distribution function, the bias term in the extreme value index estimator is reduced to a lower level. This has been done by Feuerverger and Hall (1999) in the $\gamma$ positive case. We remark that for all aforementioned bias correction methods, $\gamma$ is assumed to be positive. Some recent studies focus on the $\gamma$ negative case following the maximum likelihood method; see Li and Peng (2009) and Li et al. (2011).

To our best knowledge, no bias correction method has been studied for the general case when $\gamma$ is not restricted to be positive or negative. The present paper fills this gap because the PWM estimator does not require prior knowledge on the sign of $\gamma$.

We further provide a bias correction procedure for the PWM estimators of high quantiles and, when $\gamma$ is negative, of the endpoint of a distribution. Simulations confirm that the bias-corrected estimators exhibit a superior performance compared to most existing estimators, except in the endpoint estimation. We conduct a real data application on the still water level at Hoek van Holland, The Netherlands.

3.2 Tail expansion and the choice of $k$

3.2.1 Tail expansion: the first, second and third order conditions

Suppose that $F$ is the common distribution function of $X_1, \cdots, X_n$. Write $U := (1/(1-F))^{-}$, where $^{-}$ denotes the left-continuous inverse function. Then the necessary and sufficient condition for $F$ being in the max-domain of attraction is

$$
\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{3.2.1}
$$

for $x > 0$, where $a(t)$ is a positive function, called the (first order) scale function, and the parameter $\gamma$ is the extreme value index.
Under the first order condition (3.2.1), existing estimators of $\gamma$, such as the Hill estimator (Hill (1975)), the moment estimator (Dekkers et al. (1989)) and the maximum likelihood estimator (Smith (1987)), have been proved to be consistent. In order to possess asymptotic normalities, suitable second order conditions of $U$ are typically required. We use the generalized second order condition which characterizes the speed of convergence in (3.2.1) as follows (de Haan and Stadtmüller (1996)): there exists a second order scale function $A(t)$ with constant sign near infinity satisfying $A(t) \to 0$ as $t \to \infty$ and a second order index $\rho < 0$,\(^3\) such that for all $x > 0$,

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

(3.2.2)

As a consequence, $|A|$ is a regularly varying function with index $\rho$.

The direct approach on bias correction employs estimators on $A(n/k)$ as well as $\rho$. Similar to the logic that the second order condition ensures asymptotic normalities of estimators for first order parameters, to obtain the asymptotic properties of the estimators for the second order parameters, we need to impose a third order condition which characterizes the speed of convergence in the second order condition as follows. Suppose there exists a third order scale function $B(t)$ with constant sign near infinity satisfying $B(t) \to 0$ as $t \to \infty$ and a third order index $\rho' < 0$, such that for all $x > 0$,

$$\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} - \frac{x^{\gamma+\rho - 1}}{\rho(\gamma+\rho)} = x^{\gamma+\rho+\rho'} - \frac{1}{\gamma + \rho + \rho'}. $$

(3.2.3)

As a consequence, $|B|$ is a regularly varying function with index $\rho'$. The third order condition has been developed in Fraga Alves et al. (2003). In the literature on bias correction for positive $\gamma$, a similar but simpler third order condition applies.

### 3.2.2 The choice of $k$

With a bias-corrected estimator of the extreme value index, it is possible to choose a high value of $k$, but the choice can not be too high. As pointed out in Caeiro and Gomes (2008), under the third order condition, for an intermediate series $k$ such that $\sqrt{k} |A(n/k)| \to \infty$ and $\sqrt{k} (A^2(n/k) + |A(n/k)B(n/k)|) \to \lambda'$ at a finite level as

\(^3\)In a general setup of the second order condition, it is possible to have a second order index $\rho$ zero. Nevertheless, for bias correction studies, it is usually assumed that $\rho < 0$. We follow such an assumption. Similar argument holds for the third order index below.
Chapter 3. Bias Correction of PWM Estimators

$n \to \infty$, the “bias-corrected estimators” in Caeiro et al. (2005) again have a bias. This applies in general to bias subtraction procedures for any extreme value index estimator. Therefore, to obtain a fully bias-corrected estimator, the highest $k$ one can choose in the original $\gamma$ estimator, denoted as $k_\gamma$, should satisfy
\[ \sqrt{k_\gamma} \left( A^2 \left( n/k_\gamma \right) + |A(n/k_\gamma)B(n/k_\gamma)| \right) \to 0. \] We do choose such a $k$ level in the bias correction procedure of the PWM estimator, i.e. $k_\gamma$ satisfies the condition that, as $n \to \infty$,
\[ \left\{ \begin{array}{l} \sqrt{k_\gamma} |A(n/k_\gamma)| \to \infty, \\ \sqrt{k_\gamma} \left( A^2 \left( n/k_\gamma \right) + |A(n/k_\gamma)B(n/k_\gamma)| \right) = O(1), \end{array} \right. \] (3.2.4)

We remark that this condition is not too restrictive: considering the regular variation property of $|A|$ and $|B|$, condition (3.2.4) is equivalent to choosing a $k$ of order $n^\kappa$, with $\kappa \in \left( \frac{2\rho}{2\rho-1}, \frac{2(\rho+\max(\rho,\rho'))}{2(\rho+\max(\rho,\rho'))-1} \right)$. Notice that in the literature on the optimal choice of $k$ for estimators of the extreme value index, the optimal choice $k_{opt}$ is of order $n^\kappa$ with $\kappa = \frac{2\rho}{2\rho-1}$, cf. Theorem 1 in Danielsson et al. (2001). Thus the choice of $k$ in our bias correction procedure is higher and less restrictive.

When estimating the bias term, estimators of the second order index $\rho$ are employed, which also involves a (different) choice of $k$. An accurate estimate of the second order index $\rho$ is in general difficult to achieve according to extensive simulations in the literature, see, e.g., Gomes et al. (2002). The recommendation in the literature is that for the $\rho$ estimator, the choice of $k$, $k_\rho$, should be at a higher level compared to $k_\gamma$.\footnote{We remark that in the maximum likelihood approach on bias reduction, the maximum likelihood procedure is applied to all parameters including the $\gamma$ and $\rho$ parameter simultaneously. The choice of $k$ for $\gamma$ thus follows the optimal choice for the $\rho$ estimator. This leads to less flexibility in obtaining the bias-reduced estimator of $\gamma$ as pointed out in Li and Peng (2009).}

More precisely, we choose a $k_\rho$ such that as $n \to \infty$,
\[ \left\{ \begin{array}{l} \sqrt{k_\rho} |A(n/k_\rho)| \to \infty, \\ \sqrt{k_\rho} \left( A^2 \left( n/k_\rho \right) + |A(n/k_\rho)B(n/k_\rho)| \right) = O(1), \\ \frac{k_\rho}{k_\gamma} \to 0. \end{array} \right. \] (3.2.5)
3.3 Bias correction for the PWM estimators on the extreme value index, high quantiles and endpoint

3.3.1 Bias correction for the extreme value index estimator

Recall that we have i.i.d. observations $X_1, X_2, \cdots, X_n$ with a common distribution function $F$. The following lemma gives the asymptotic behavior of the PWMs, defined in (3.1.2). It can be obtained by applying the expansion of excesses above a high threshold as in Drees (1998, Theorem 2.1); see the proof in Appendix.

Lemma 3.3.1. Suppose the third order condition (3.2.3) holds with $\gamma < 1/2$ and $\rho, \rho' < 0$. Let $k$ be an intermediate sequence such that $k \rightarrow \infty$, $k/n \rightarrow 0$ and $\sqrt{k} \left( A^2(n/k) + |A(n/k)B(n/k)| \right) = O(1)$ as $n \rightarrow \infty$. There exists a probability space where versions of $X_n$ and a sequence of standard Brownian motions $W_n$, $n = 1, 2, \ldots$, are defined and for $q = 1, 2, \ldots$,

$$\frac{q(q - \gamma)I_q}{a(n/k)} = 1 + \frac{1}{\sqrt{k}}L_q + A(n/k)B_q + \varepsilon_{n,k}, \quad (3.3.1)$$

where

$$L_q = q(q - \gamma) \int_0^1 s^{q-1}(s^{-\gamma-1}W_n(s) - W_n(1)) ds,$$

$$B_q = q(q - \gamma) \int_0^1 s^{q-1}\frac{s^{q-\gamma-\rho} - 1}{\rho(\gamma + \rho)} ds = \frac{q - \gamma}{\rho(q - \gamma - \rho)},$$

and $\varepsilon_{n,k} = O_p(A^2(n/k) + |A(n/k)B(n/k)|)$. Moreover, for any positive integers $q$ and $r$,

$$\text{Cov}(W_n(1), L_q) = \gamma, \quad (3.3.2)$$

and

$$\text{Cov}(L_q, L_r) = \frac{qr}{q + r - 1 - 2\gamma} + \gamma^2. \quad (3.3.3)$$

Under the condition (3.2.5), the $\varepsilon_{n,k}$ term in (3.3.1) is of a lower order compared to the other two terms. Throughout, $\varepsilon_{n,k}$ denotes any term at the level of $O_p(A^2(n/k) + |A(n/k)B(n/k)|)$. They are not necessarily equal.
The PWM estimator of the extreme value index $\gamma$ is constructed by combining two different PWMs. With two different positive integers $q$ and $r$, from Lemma 3.3.1, we get that, as $n \to \infty$,

\[
\frac{qI_q}{rI_r} \overset{P}{\to} \frac{r - \gamma}{q - \gamma}.
\]

For $\gamma < \min(q, r)$, this leads to a consistent PWM estimator of $\gamma$ as

\[
\hat{\gamma}_{q,r} = \frac{q^2I_q - r^2I_r}{qI_q - rI_r}.
\]

The estimator $\hat{\gamma}_{2,1}$ coincides with $\hat{\gamma}_{PWM}$ defined in (3.1.1), which is a modified version of the classic PWM estimator initially proposed by Hosking and Wallis (1987). The original estimator, designed for observations from the GPD has no threshold, or one can say it has threshold zero, whereas we have threshold $X_{n,n-k}$ due to the fact that we deal with observations in the domain of attraction. We are going to correct for the bias of the estimator $\hat{\gamma}_{2,1}$.

The asymptotic expansion of the PWM estimator can be derived from (3.3.1) as

\[
\hat{\gamma}_{q,r} - \gamma = -\frac{1}{\sqrt{k}} \frac{(q-\gamma)(r-\gamma)}{q-r} (L_q - L_r) + A(n/k) \frac{(q-\gamma)(r-\gamma)}{(q-\gamma-\rho)(r-\gamma-\rho)} + \varepsilon_{n,k},
\]

by applying Cramér’s delta method. Thus, we omit the details of the derivation.

We see from the asymptotic expansion that in order to correct the bias term in the PWM estimator, $\hat{\gamma}_{2,1}$, it is necessary to construct estimators for the second order index $\rho$ as well as the second order scale function $A(n/k)$. We construct the estimators starting from comparing two PWM estimators. For simplicity, we compare $\hat{\gamma}_{q,r}$ and $\hat{\gamma}_{p,r}$, where $p$, $q$ and $r$ are three different integers. From the asymptotic property in (3.3.4), we get that

\[
\hat{\gamma}_{q,r} - \hat{\gamma}_{p,r} = \frac{1}{\sqrt{k}} N_{p,q,r} + A(n/k) \frac{\rho(r-\gamma)(p-q)}{(p-\gamma-\rho)(q-\gamma-\rho)(r-\gamma-\rho)} + \varepsilon_{n,k}.
\]

Here the stochastic term

\[
N_{p,q,r} := (r-\gamma) \left( \frac{p-\gamma}{p-r} L_p - \frac{q-\gamma}{q-r} L_q + \frac{(p-q)(r-\gamma)}{(p-r)(q-r)} L_r \right)
\]

is a normally distributed random variable with mean zero. We take specific values for $p, q, r$ in (3.3.5): $p = 4$, $q = 3$, where $r$ is first set to 1 and then 2. When choosing $k_{\rho}$.
as in (3.2.5), relation (3.3.5) implies
\[
\frac{\hat{\gamma}_{3.1}(k_p) - \hat{\gamma}_{4.1}(k_p)}{\sqrt{n/k}} \to (1 - \gamma, 2 - \gamma - \rho),
\]
as \(n \to \infty\). Thus, a consistent estimator of \(\rho\) can be given as
\[
\hat{\rho}(k_p) := 1 - \hat{\gamma}_{2.1}(k_p) - \frac{1}{\sqrt{n/k}} \frac{\gamma_{3.1}(k_p) - \gamma_{4.1}(k_p)}{\gamma_{3.2}(k_p) - \gamma_{4.2}(k_p)},
\]
(3.3.6)

As discussed in Section 3.2.2, the intermediate sequence \(k\) we will use in the bias-corrected PWM estimator, \(k_{\gamma}\), satisfies the condition (3.2.4), while \(k_\rho\) is at a higher level, in the sense that \(k_\rho/k_{\gamma} \to \infty\) as \(n \to \infty\). It ensures that the asymptotic distribution of \(\rho\) will not contaminate the bias subtraction procedure. From Lemma 3.6.2, we get that
\[
\sqrt{k_n} A(n/k_{\gamma})(\hat{\rho}(k_p) - \rho) = o_p(1).
\]
(3.3.7)

From now on, we just write \(k\) for \(k_{\gamma}\) satisfying (3.2.4).

The estimator of \(A(n/k)\) is based on the expansion of \(\hat{\gamma}_{2.1} - \hat{\gamma}_{3.1}\) as in (3.3.5),
\[
\hat{\gamma}_{2.1}(k) - \hat{\gamma}_{3.1}(k) = \frac{1}{\sqrt{k}} N_{3.2.1} + A(n/k) \frac{\rho(1 - \gamma)}{(1 - \gamma - \rho)(2 - \gamma - \rho)(3 - \gamma - \rho)} + \epsilon, \quad \text{(3.3.8)}
\]
This suggests the following estimator
\[
\hat{A}(n/k) := (\hat{\gamma}_{2.1}(k) - \hat{\gamma}_{3.1}(k)) \left(1 - \frac{1}{\hat{\gamma}_{2.1}(k) - \hat{\rho}(k_p)} - \frac{2 - \gamma - \rho}{2 - \gamma_{2.1}(k) - \hat{\rho}(k_p)}\right).
\]
(3.3.9)

Using these estimators of the second order index and scale function, we can now directly subtract the bias term from the PWM estimator \(\hat{\gamma}_{2.1}\) to obtain the bias-corrected estimator \(\hat{\gamma}_{ub}\) given as
\[
\hat{\gamma}_{ub} := \hat{\gamma}_{2.1}(k) - \hat{A}(n/k) \frac{(1 - \hat{\gamma}_{2.1}(k)) (2 - \hat{\gamma}_{2.1}(k))}{(1 - \hat{\gamma}_{2.1}(k) - \hat{\rho}(k_p)) (2 - \gamma_{2.1}(k) - \hat{\rho}(k_p))}.
\]
(3.3.10)

The following theorem gives the asymptotic normality of \(\hat{\gamma}_{ub}\).

\textit{Theorem 3.3.1.} Write \(a_0 = \frac{1}{2\sigma}(1 - \gamma)(2 - \gamma)(3 - \gamma)\) and \(a_1 = \gamma + \rho - 1\). Under the conditions (3.2.3), (3.2.4) and (3.2.5) with \(\gamma < 1/2\) and \(\rho, \rho' < 0\), as \(n \to \infty\),
\[
\sqrt{k} (\hat{\gamma}_{ub} - \gamma) \overset{d}{\to} N(0, \sigma_{ub}^2(\gamma, \rho)),
\]
(3.3.11)
where $\sigma^2_{ub}(\gamma, \rho) = a^2_0 \left( 16\gamma^2 + \frac{a^2_1}{1-2\gamma} - \frac{4a_1(a_1-1)}{1-\gamma} + \frac{22(a_1-1)^2-6}{3-2\gamma} - \frac{12(a_1-1)(a_1-2)}{2-\gamma} + \frac{9(a_1-2)^2}{5-2\gamma} \right)$.

**Remark 3.3.1.** From (3.3.4), we have
\[
\sqrt{k_{pwm}}(\hat{\gamma}_{pwm} - \gamma) \xrightarrow{d} N(b, \sigma^2_{pwm}(\gamma)),
\]
where $k_{pwm}$ is an intermediate sequence such that $\lim_{n \to \infty} \sqrt{k_{pwm}} A(n/k_{pwm})$ exists and is finite. In Table 3.1, we report the ratio of $\sigma^2_{ub}(\gamma, \rho)$ and $\sigma^2_{pwm}(\gamma)$ for various values of $\gamma$ and $\rho$. Although $\sigma^2_{ub}(\gamma, \rho)$ is larger than $\sigma^2_{pwm}(\gamma)$, the convergence speed of the bias-corrected estimator is faster than that of the PWM estimator, since $\lim_{n \to \infty} k_{pwm}/k = 0$. This guarantees that $\hat{\gamma}_{ub}$ has a lower level of root mean square error, which is supported by the simulation study in Section 3.4.

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$-0.4$</th>
<th>$-0.3$</th>
<th>$-0.2$</th>
<th>$-0.1$</th>
<th>$0$</th>
<th>$0.1$</th>
<th>$0.2$</th>
<th>$0.3$</th>
<th>$0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-2$</td>
<td>2.5</td>
<td>2.5</td>
<td>2.4</td>
<td>2.3</td>
<td>2.2</td>
<td>2.2</td>
<td>2.3</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>$-1.5$</td>
<td>3.9</td>
<td>3.7</td>
<td>3.5</td>
<td>3.3</td>
<td>3.1</td>
<td>2.9</td>
<td>2.8</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>$-1$</td>
<td>7.6</td>
<td>7.1</td>
<td>6.6</td>
<td>6.0</td>
<td>5.4</td>
<td>4.8</td>
<td>4.3</td>
<td>3.9</td>
<td>3.6</td>
</tr>
<tr>
<td>$-0.5$</td>
<td>27.3</td>
<td>25.1</td>
<td>22.7</td>
<td>20.0</td>
<td>17.1</td>
<td>14.2</td>
<td>11.4</td>
<td>9.1</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Note: The reported values are $\sigma^2_{ub}(\gamma, \rho)/\sigma^2_{2,1}(\gamma)$, where $\sigma^2_{ub}(\gamma, \rho)$ indicates the variance of the limit distribution in (3.3.11) and $\sigma^2_{2,1}(\gamma)$ indicates that of $\sqrt{k}(\hat{\gamma}_{2,1} - \gamma)$ in (3.3.4). Notice that the convergence speed, $\sqrt{k}$, might be different for the two estimators. The bias-corrected estimator allows a higher level of $k$.

**Remark 3.3.2.** Similar to the PWM estimator $\hat{\gamma}_{2,1}$, this bias-corrected PWM estimator can be applied to any $\gamma$ that is below $1/2$ and is shift and scale invariant. Therefore, it is particularly useful in dealing with the case that the extreme value index is around zero.

### 3.3.2 Bias correction for the quantile estimator

Estimating high quantiles is one of the major interests in applications of extreme value statistics. The high quantile estimator has been introduced in Weissman (1978). Under the second order condition (3.2.2), the estimator has a bias term that can be corrected under the framework of the third order condition (3.2.3). To our best knowledge, bias
correction for high quantile estimators has been studied for the $\gamma$ positive case only; see e.g. Matthys et al. (2004), Gomes and Figueiredo (2006), Gomes and Pestana (2007b), Beirlant et al. (2008) and Li et al. (2010).

For a low probability level $p = p(n)$ depending on $n$ such that $\lim_{n \to \infty} p(n) = 0$, we are interested in estimating the high quantile $x_p$ defined by $x_p := \sup \{ x \mid F(x) < 1 - p \}$, or equivalently, $x_p = U(1/p)$. The high quantile estimator in de Haan and Rootzén (1993) is based on the first order expansion of $U$ in (3.2.1) with taking $tx = 1/p$ and $t = n/k$. That is

$$\hat{x}_p = \hat{U}(n/k) + \hat{a}(n/k) \frac{d_{\hat{n}} - \gamma}{\gamma},$$

where $\hat{U}(n/k) := X_{n,n-k}$, $d_n = \frac{k}{np}$ and $\hat{a}(n/k) := \frac{2I_{11}I_2 I_{11} - 2I_{2} I_{11}}{I_{11}^2 - I_{12}^2}$, see de Haan and Ferreira (2006). Notice that the asymptotic bias of $\hat{a}(n/k)$ given in de Haan and Ferreira (2006) is incorrect. For a correct asymptotic expansion of $\hat{a}(n/k)$, see (3.6.5). Based on that, a bias-corrected estimator of the scale function is given by

$$\hat{a}_{ub}(n/k) := \hat{a}(n/k) \exp \left( -A(n/k) \frac{(1 - \hat{\gamma}_{ub}(k))(2 - \hat{\gamma}_{ub}(k)) - \hat{\rho}(k_p)(3 - 2\hat{\gamma}_{ub}(k))}{\hat{\rho}(k_p)(1 - \hat{\gamma}_{ub}(k) - \hat{\rho}(k_p))(2 - \hat{\gamma}_{ub}(k) - \hat{\rho}(k_p))} \right),$$

(3.3.12)

The second order expansion of $U$ in (3.2.2) suggests a bias-corrected high quantile estimator as

$$\hat{x}_{p,ub} := \hat{U}(n/k) + \hat{a}_{ub}(n/k) \frac{d_{\hat{n}} - \hat{\gamma}_{ub}(k)}{\hat{\gamma}_{ub}(k)} + \hat{A}(n/k)\hat{a}_{ub}(n/k) \frac{\hat{\rho}(k_p) + \hat{\gamma}_{ub}(k)}{\hat{\rho}(k_p)(\hat{\gamma}_{ub}(k) + \hat{\rho}(k_p))} - 1.$$

(3.3.13)

Here we choose to use $\hat{\gamma}_{ub}$, $\hat{\rho}$ and $\hat{A}(n/k)$ as in Section 3.3.1.

The asymptotic property of the bias-corrected quantile estimator is given in the following theorem.

**Theorem 3.3.2.** Suppose the third order condition (3.2.3) holds with $\gamma < 1/2$ and $\rho, \rho' < 0$. Assume the conditions (3.2.4) and (3.2.5) hold and the probability sequence $p(n)$ satisfies $np = o(k)$ and $\log(np) = o(\sqrt{k})$ as $n \to \infty$. Then, as $n \to \infty$, with $\gamma_- := \min(0, \gamma)$ and $q_{\gamma}(t) := \int_t^\infty s^{-\gamma} \log s ds$ for $t > 1$,

$$\sqrt{k} \frac{\hat{x}_{p,ub} - x_p}{a(n/k) q_{\gamma}(d_n)} \xrightarrow{d} \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4.$$

$\Gamma_1, \Gamma_2, \Gamma_3,$ and $\Gamma_4$ are centered normally distributed random variables stemming from the asymptotic expansions of $\hat{U}(n/k)$, $\hat{\gamma}_{ub}$, $\hat{a}_{ub}(n/k)$ and the bias correction procedure,
respectively. More specifically, with \((W(1), J_1, J_2, J_3) \overset{d}{=} (W_n(1), L_1, L_2, L_3)\),

\[
\begin{align*}
\Gamma_1 &= (\gamma_-)^2 W(1), \\
\Gamma_2 &= \frac{(1 - \gamma)(2 - \gamma)(3 - \gamma)}{2\rho} ((\gamma + \rho - 1)J_1 - 2(\gamma + \rho - 2)J_2 + (\gamma + \rho - 3)J_3), \\
\Gamma_3 &= -\frac{\gamma_-(1 - \gamma)(2 - \gamma)(3 - \gamma)}{2\rho} \left( (\gamma + \rho - 1) \left( \frac{1}{\rho} - \frac{1}{2 - \gamma} - \frac{1}{3 - \gamma} \right) J_1 \\
&\quad - 2(\gamma + \rho - 2) \left( \frac{1}{\rho} - \frac{1}{1 - \gamma} - \frac{1}{3 - \gamma} \right) J_2 \\
&\quad + (\gamma + \rho - 3) \left( \frac{1}{\rho} - \frac{1}{1 - \gamma} - \frac{1}{2 - \gamma} \right) J_3 \right), \\
\Gamma_4 &= -\frac{(\gamma_-)^2(1 - \gamma - \rho)(2 - \gamma - \rho)(3 - \gamma - \rho)}{2\rho^2(\gamma_- + \rho)} ((1 - \gamma)J_1 - 2(2 - \gamma)J_2 + (3 - \gamma)J_3).
\end{align*}
\]

Here \(W\) is the standard Brownian motion in Lemma 3.3.1. The covariance matrix of \((\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4)\) can be obtained from (3.3.2) and (3.3.3).

We remark that the asymptotic relation remains valid when replacing \(a(n/k), q_\gamma(n/k)\) by their consistent estimators. This is useful in constructing confidence interval of the high quantile estimates.

Notice that in the bias-corrected quantile estimator, we do not assume the sign of the extreme value index \(\gamma\) ex-ante. Nevertheless, once the sign is known, the asymptotic property can be significantly simplified as in the following corollary.

**Corollary 3.3.1.** If \(\gamma \geq 0\), as \(n \to \infty\),

\[
\sqrt{k} \frac{\hat{x}_{p,\text{ub}} - x_p}{a(n/k)q_\gamma(d_n)} \overset{d}{\to} N(0, \sigma_{\text{ub}}^2(\gamma, \rho)),
\]

where \(\sigma_{\text{ub}}^2(\gamma, \rho)\) is defined in Theorem 3.3.1. If \(\gamma < 0\), as \(n \to \infty\),

\[
\sqrt{k} \frac{\hat{x}_{p,\text{ub}} - x_p}{a(n/k)q_\gamma(d_n)} \overset{d}{\to} N(0, \sigma_{\text{neg}}^2),
\]

where \(\sigma_{\text{neg}}^2 = 9((1 - 2\gamma)^{-1}(1 - \gamma)^2(1 - a_2)^2 - 2(1 - 2a_2)(4\gamma a_2 - 2\gamma - 11a_2 + 5) + 2(3 - 2\gamma)^{-1}(1 - \gamma)^2(11a_2^2 - 12a_2 + 3) - 3a_2^2 + 4a_2 - 1) + (5 - 2\gamma)^{-1}(3 - \gamma)^2(1 - 3a_2)^2\) with \(a_2 = (\gamma + \rho)^{-1}\).
3.3.3 Bias correction for the endpoint estimator

The endpoint of a distribution function $F$ is defined as $x^* = \sup \{ x : F(x) < 1 \}$. If $F$ belongs to the domain of attraction with $\gamma < 0$, the endpoint of $F$ is finite. The estimation of the endpoint has been studied in literature; see Hall (1982) and Dekkers et al. (1989). The statistical procedure in such situations helps to answer many interesting questions. How long can we live (Aarssen and de Haan (1994))? What is the ultimate world record in a specific athletic event (Einmahl and Magnus (2008))? Similar to other estimation procedure in extreme value statistics, the estimation of the endpoint also bears a potential bias.

A bias-corrected endpoint estimator can be derived from the bias-corrected quantile estimator defined in (3.3.13). By taking $p = 0$ in (3.3.13), we get

\[
\hat{x}_{ub}^* := X_{n,n-k} - \frac{\hat{\tau}_{ub}(n/k)}{\hat{\gamma}_{ub}(k)} - \frac{\hat{A}(n/k)\hat{\tau}_{ub}(n/k)}{\hat{\rho}(k_p)(\hat{\gamma}_{ub}(k) + \hat{\rho}(k_p))}.
\]  

(3.3.14)

The asymptotic property of the bias-corrected endpoint estimator is analogous to that of the bias-corrected high quantile estimator as shown in the following theorem.

**Theorem 3.3.3.** Under the conditions (3.2.4), (3.2.5) and (3.2.3) with $\gamma < 0$ and $\rho, \rho' < 0$, as $n \to \infty$,

\[
\frac{\sqrt{k}}{\alpha(n/k)} (\hat{x}_{ub}^* - x^*) \overset{d}{\to} N\left(0, \frac{\sigma^2_{neg}}{\gamma^2}\right),
\]  

(3.3.15)

where $\sigma^2_{neg}$ is defined in Corollary 3.3.1.

**Remark 3.3.3.** Notice that the rate of convergence in (3.3.15) is approximately $\frac{k^{1/2+\gamma}}{n}$. When $\gamma < -1/2$ it is a decreasing function of $k$. Thus choosing a high $k$ will not reduce the estimation error. In this case, the sample maximum $X_{n,n}$ converges to $x^*$ faster than $\hat{x}_{ub}^*$ (see de Haan and Ferreira (2006, Remark 4.5.5)). When $\gamma > -1/2$, theoretically, the bias-corrected estimator can benefit from choosing a higher level of $k$. Nevertheless, the high asymptotic variance imposed by the bias correction procedure dilutes the benefit in finite sample exercise. Such an effect is particularly severe when $\gamma$ is close to zero. See the simulation results below.
3.4 Simulations

In this section, we perform finite sample simulations on the proposed bias-corrected estimators. Data are simulated from four distributions: the standard Student-t distribution with degree of freedom 3, the Gumbel distribution \( F(x) = e^{-e^{-x}}, x \in \mathbb{R} \), the reversed Burr distribution \( F(x) = 1 - (1 + (4 - x)^{-4})^{-5/4}, x \leq 4 \) and the Weibull distribution \( F(x) = 1 - e^{-x^2}, x \in \mathbb{R} \). The first, second and third order indices of the four distributions are listed in Table 3.2. Notice that the Weibull distribution is not under our framework as we assume \( \rho \) negative. We draw 100 samples with sample size \( n = 5000, 2000, 1000 \).

Table 3.2: Distributions for simulation

<table>
<thead>
<tr>
<th>parameters</th>
<th>Student-t(3)</th>
<th>Gumbel</th>
<th>reversed Burr</th>
<th>Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>( 1/3 )</td>
<td>0</td>
<td>-0.2</td>
<td>0</td>
</tr>
<tr>
<td>( \rho )</td>
<td>-2/3</td>
<td>-1</td>
<td>-0.8</td>
<td>0</td>
</tr>
<tr>
<td>( \rho' )</td>
<td>-4/3</td>
<td>-1</td>
<td>-0.8</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The Student-t(3) distribution is the standard Student-t distribution with degree of freedom 3. The other three distribution functions are: \( F(x) = e^{-e^{-x}}, x \in \mathbb{R} \) (Gumbel); \( F(x) = 1 - (1 + (4 - x)^{-4})^{-5/4}, x \leq 4 \) (reversed Burr); and \( F(x) = 1 - e^{-x^2}, x \in \mathbb{R} \) (Weibull).

As discussed in section 3.2.2, a higher level \( k_\rho \) is chosen for estimating the second order index \( \rho \). With some pre-simulation, we choose a fixed \( k_\rho \) at \( k_\rho = \lfloor n^{0.98} \rfloor \).

3.4.1 Simulations on \( \gamma \) estimation

Recall that the bias-corrected PWM estimator of \( \gamma \) is defined in (3.3.10). We compare our estimator \( \hat{\gamma}_{ub} \) with the moment estimator \( \hat{\gamma}_M \) in Dekkers et al. (1989) and the original PWM estimator \( \hat{\gamma}_{pwm} = \hat{\gamma}_{2,1} \). To eliminate the impact of \( \rho \) estimation, we construct a pseudo bias-corrected estimator with the real value of \( \rho \), defined as

\[
\hat{\gamma}_\rho := \hat{\gamma}_{2,1}(k) - \hat{A}(n/k) \frac{(1 - \hat{\gamma}_{2,1}(k))(2 - \hat{\gamma}_{2,1}(k))}{(1 - \hat{\gamma}_{2,1}(k) - \rho)(2 - \hat{\gamma}_{2,1}(k) - \rho)}.
\]  

(3.4.1)

We only apply \( \hat{\gamma}_\rho \) to data simulated from distributions with \( \rho < 0 \). In addition, for simulation for the Student-t(3) distribution, a bias-corrected Hill estimator \( \hat{\gamma}_{uH} \) given
by (1.15) in Caeiro et al. (2005) is also included. Since both $\hat{\gamma}_M$ and $\hat{\gamma}_{uH}$ are applicable to positive observations only, the estimates are set to be zero if the $(n - k)$-th order statistic is negative.

![Graphs showing average of $\gamma$ estimates for 100 samples](image)

**Figure 3.1:** Average of $\gamma$ estimates for 100 samples.

Note: Each sample consists of 5000 observations. $\hat{\gamma}_{ub}$ is the bias-corrected PWM estimator in (3.3.10); $\hat{\gamma}_{pwm}$ is the PWM estimator in (3.1.1); $\hat{\gamma}_M$ is the moment estimator in Dekkers et al. (1989); $\hat{\gamma}_{\rho}$ is a pseudo bias-corrected PWM estimator in (3.4.1); $\hat{\gamma}_{uH}$ is a bias-corrected Hill estimator in Caeiro et al. (2005). Horizontal lines indicate the real values of $\gamma$. 
Figure 3.2: Root mean square error of $\gamma$ estimates for 100 samples.

Note: Each sample consists of 5000 observations. $\hat{\gamma}_{ub}$ is the bias-corrected PWM estimator in (3.3.10); $\hat{\gamma}_{pwm}$ is the PWM estimator in (3.1.1); $\hat{\gamma}_M$ is the moment estimator in Dekkers et al. (1989); $\hat{\gamma}_p$ is a pseudo bias-corrected PWM estimator in (3.4.1); $\hat{\gamma}_{uH}$ is a bias-corrected Hill estimator in Caeiro et al. (2005).

In Figure 3.1, we plot the average of $\gamma$ estimates against $k$ from 100 samples with sample size 5000 for each distribution. The plots show that the bias-corrected estimates remain stable for a large range of $k$, which is close to the real value. Other estimates bear dramatically increasing biases as the choice of $k$ increases. It is remarkable that
Chapter 3. *Bias Correction of PWM Estimators*

![Graphs](image)

**Figure 3.3:** Average and RMSE of $\gamma$ estimates for 100 samples.

Note: Data are simulated from the Student-$t(3)$ distribution. The sample sizes are 1000 for plots in the upper panel and 2000 for plots in the lower panel. $\hat{\gamma}_{ub}$ is the bias-corrected PWM estimator in (3.3.10); $\hat{\gamma}_{pwm}$ is the PWM estimator in (3.1.1); $\hat{\gamma}_{M}$ is the moment estimator in Dekkers et al. (1989); $\hat{\gamma}_{\rho}$ is a pseudo bias-corrected PWM estimator in (3.4.1); $\hat{\gamma}_{uH}$ is a bias-corrected Hill estimator in Caeiro et al. (2005).

$\hat{\gamma}_{\rho}$ does not perform better than $\hat{\gamma}_{ub}$. Hence the inaccuracy in estimating $\rho$ does not impose a significant error when estimating $\gamma$. Although for the Weibull distribution, the bias-corrected estimator also has a bias, it is still the most stable estimator with
respect to the variation of \( k \). Moreover, it is closest to the real value.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Average of the high quantile estimates for 100 samples.}
\end{figure}

Note: Each sample consists of 5000 observations. \( \hat{x}_{p,ub} \) is the bias-corrected high quantile estimator in (3.3.13); \( \hat{x}_{p,pwm} \) is the PWM quantile estimator in de Haan and Ferreira (2006); \( \hat{x}_{p,M} \) is the moment quantile estimator in Dekkers et al. (1989). Horizontal lines indicate the value of \( x_p \) with \( p = 1/2000 \).

In our procedure, bias is corrected at the cost of imposing extra asymptotic variance. With the choice of a higher level \( k \), the asymptotic root mean square error (RMSE)
of the bias-corrected estimator is theoretically of a lower order compared to that of the original estimator. In Figure 3.2, we plot the RMSE against \( k \) and confirm this theoretical property. In general, the bias-corrected estimators are much less sensitive with respect to the choice of \( k \).

In order to get more insight on finite sample behavior of the estimators, simulations for smaller samples are also implemented. The plots in the upper panel of Figure 3.3 present the results for \( n = 1000 \), which suggest that although the bias is corrected by the procedure it is not sufficient to compensate the increase in variance. When we raise the sample size to 2000, the advantage of the bias-corrected PWM estimator becomes apparent: the RMSE remains at a low level for a wider choice of \( k \) compared to other estimators.

### 3.4.2 Simulations on the high quantile and endpoint estimation

The bias-corrected quantile estimator \( \hat{x}_{p,ub} \) is given in (3.3.13). The competitors are the moment quantile estimator \( \hat{x}_{p,M} \) introduced in Dekkers et al. (1989) and the PWM quantile estimator without bias correction \( \hat{x}_{p,pwm} \); see Exercise 4.7 in de Haan and Ferreira (2006). We study the estimators of the \( 1 - p \) quantile with \( p = 1/2000 \).

Figures 3.4 and 3.5 demonstrate the performance of the high quantile estimators in terms of the average of the estimates and the RMSE. The bias-corrected quantile estimator \( \hat{x}_{p,ub} \) performs better than the other two estimators.

We further apply the bias-corrected endpoint estimator, \( \hat{x}^*_{ub} \) defined in (3.3.14) to the random samples generated from the reversed Burr distribution. We again employ the moment and the original PWM methods to produce two competitors. The results are shown in Figure 3.6. As explained in Remark 3.3.3, the \( \hat{x}^*_{ub} \) does not have an obviously better performance in terms of RMSE. With an alternative evaluation criteria, the median absolute deviation (MAD) (see Hampel (1974) and Rousseeuw and Croux (1993)), we focus on comparing the estimators in terms of the estimation bias. The MAD is defined as

\[
MAD := \text{median}(\vert \hat{x}^*(k) - \hat{x}^* \vert).
\]
Chapter 3. Bias Correction of PWM Estimators

Figure 3.5: Root mean square error of the high quantile estimates for 100 samples.

Note: Each sample consists of 5000 observations. \( \hat{x}_{p,ub} \) is the bias-corrected high quantile estimator in (3.3.13); \( \hat{x}_{p,pwm} \) is the PWM quantile estimator in de Haan and Ferreira (2006); \( \hat{x}_{p,M} \) is the moment quantile estimator in Dekkers et al. (1989).

As shown in the right bottom plot in Figure 3.6, the MAD of \( \hat{x}_{ub}^* \) remains low and stable for a larger range of \( k \) compared to those of \( \hat{x}_{p,pwm}^* \) and \( \hat{x}_{p,M}^* \).
Chapter 3. Bias Correction of PWM Estimators

Figure 3.6: The endpoint estimates for 100 samples.

Note: Each sample consists of 5000 observations generated from reversed Burr distribution with distribution function $F(x) = 1 - (1 + (4 - x)^{-4})^{-5/4}$. The endpoint of the distribution is $x_0 = 4$. $\hat{x}_{ub}$ is the bias-corrected endpoint estimator in (3.3.14); $\hat{x}_{pwm}$ is the PWM endpoint estimator; $\hat{x}_M$ is the moment endpoint estimator in Dekkers et al. (1989).

3.5 Application to the still water level

As part of the defence against extreme flood in the Netherlands, it is crucial to have accurate estimate of the extreme still water level on various stations along the Dutch
coast. The question is to estimate the “once per 10,000 years” level of the still water. For that reason, Center for Water Management (in Dutch: “Waterdienst”) monitored the still water levels in the storm seasons.

We employ the data collected at the station Hoek van Holland. In 122 years, \( n = 1965 \) severe wind storms have been identified. We treat the still water levels during those storms as i.i.d. observations from a distribution function \( F \). Thus, the “once per 10,000 years” still water level corresponds to a high quantile of \( F \) with tail probability \( p = \frac{122}{1965} \times 10^{-4} \approx 6.2 \times 10^{-6} \).

Assuming that \( F \) satisfies the third order condition, we start with estimating the extreme value index, \( \gamma \). We plot the point estimates of \( \gamma \) using our bias-corrected PWM estimator\(^5\) as well as the moment estimator and the original PWM estimator in the upper panel of Figure 3.7. Similar to the simulation exercises, the bias-corrected PWM method yields a relatively wider range of \( k \), for which the corresponding \( \gamma \) estimates remain at a stable level (\( k \) varies from 150 to 280).

Notice that the asymptotic variance of \( \hat{\gamma}_{ub} \) can be consistently estimated by \( \frac{1}{K} \sigma^2_{ub}(\hat{\gamma}_{ub}(k), \hat{\rho}(k_p)) \). Therefore, we construct an approximate \( 1 - \alpha \) confidence interval of \( \gamma \) as following,

\[
\left[ \hat{\gamma}_{ub}(k) - z_{\alpha/2} \frac{\sigma_{ub}(\hat{\gamma}_{ub}(k), \hat{\rho}(k_p))}{\sqrt{K}}, \hat{\gamma}_{ub}(k) + z_{\alpha/2} \frac{\sigma_{ub}(\hat{\gamma}_{ub}(k), \hat{\rho}(k_p))}{\sqrt{K}} \right],
\]

where \( z_{\alpha/2} \) is the \( 1 - \alpha/2 \) quantile of the standard normal distribution. We plot the lower and upper bounds of the 95% asymptotic confidence intervals in the bottom panel of Figure 3.7. Since the value zero belongs to most of the confidence intervals with respect to different \( k \), we can not rule out the possibility that \( \gamma = 0 \). Hence it is necessary to consider estimation methods valid for a general \( \gamma \) without the prior knowledge on the sign. That is why we use the PWM method. We give the point estimates and the confidence intervals of \( \gamma \) for several choices of \( k \) in the second and third columns of Table 3.3.

\(^5\) Throughout the application, we choose \( k_p = \lfloor n^{0.98} \rfloor = 1688 \) for the estimation of the second order index, \( \rho \).
Chapter 3. Bias Correction of PWM Estimators

Still water level application: point estimates of $\gamma$

Figure 3.7: Still water level data: estimation of $\gamma$.

Note: The number of observations is 1965. The upper panel shows the point estimates of $\gamma$ using three estimators: the bias-corrected PWM estimator $\hat{\gamma}_{ub}$ in (3.3.10), the PWM estimator $\hat{\gamma}_{pwm}$ in (3.1.1), and the moment estimator $\hat{\gamma}_M$ in Dekkers et al. (1989). The lower panel gives the 95% confidence interval from the bias-corrected estimator.

We apply the bias-corrected high quantile estimator to estimate $x_p$ with $p = \frac{122}{1965} \approx 6.2 \times 10^{-6}$. The approximate $1 - \alpha$ confidence interval of $x_p$ is given by

$$\left[ \hat{x}_p - \frac{z_{\alpha/2} \sigma_2(\hat{\gamma}_{ub}(k),\hat{\rho}(k_p))\hat{a}_{ub}(n/k)q_{\hat{\gamma}_{ub}(k)}(d_n)}{\sqrt{k}}, \hat{x}_p + \frac{z_{\alpha/2} \sigma_2(\hat{\gamma}_{ub}(k),\hat{\rho}(k_p))\hat{a}_{ub}(n/k)q_{\hat{\gamma}_{ub}(k)}(d_n)}{\sqrt{k}} \right],$$

where $\sigma_2(\gamma,\rho)$ denotes the standard deviation of the limit distribution in Theorem 3.3.2. The high quantile estimators by the PWM and moment methods are also employed. The results are shown in Figure 3.8 and the fourth and fifth columns of Table 3.3. With a choice of $k = 250$, we conclude that the “once per 10000 years” still water
Still water level application: estimation of the “once per 10,000 years” level

Figure 3.8: Still water level data: estimation of the “once per 10,000 years” quantile.

Note: The number of observations is 1965. The “once per 10,000” level corresponds to a high quantile with tail probability \( p = \frac{122}{1965} \times 10^{-4} \approx 6.2 \times 10^{-6} \). The upper panel shows the point estimates of the “once per 10,000” level using three estimators: the bias-corrected high quantile estimator \( \hat{x}_{p, ub} \) in (3.3.13), the PWM quantile estimator \( \hat{x}_{p,pwm} \) in de Haan and Ferreira (2006), and the moment quantile estimator \( \hat{x}_{p,M} \) in Dekkers et al. (1989). The lower panel gives the 95% confidence intervals of the “once per 10,000” level. The scale of the vertical axis is per centimeter.

level is estimated at 611 cm.
### Table 3.3: Application to the still water level data

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\hat{\gamma}_{ub}$</th>
<th>95% confidence interval of $\gamma$</th>
<th>$\hat{x}_{p,ub}$</th>
<th>95% confidence interval of $x_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.03</td>
<td>$[-0.23, 0.30]$</td>
<td>526</td>
<td>[170, 881]</td>
</tr>
<tr>
<td>150</td>
<td>0.07</td>
<td>$[-0.14, 0.29]$</td>
<td>559</td>
<td>[183, 935]</td>
</tr>
<tr>
<td>250</td>
<td>0.13</td>
<td>$[-0.05, 0.30]$</td>
<td>611</td>
<td>[206, 1016]</td>
</tr>
<tr>
<td>280</td>
<td>0.10</td>
<td>$[-0.06, 0.27]$</td>
<td>577</td>
<td>[235, 918]</td>
</tr>
<tr>
<td>300</td>
<td>0.04</td>
<td>$[-0.11, 0.19]$</td>
<td>510</td>
<td>[269, 752]</td>
</tr>
<tr>
<td>400</td>
<td>−0.03</td>
<td>$[-0.16, 0.10]$</td>
<td>456</td>
<td>[301, 610]</td>
</tr>
</tbody>
</table>

Note: There are 1965 observations in the dataset. We choose $k_\rho = 1688$ in the estimation of the second order parameter $\rho$. $\hat{\gamma}_{ub}$ is defined in (3.3.10) and $\hat{x}_{p,ub}$ in (3.3.13) with $p = \frac{122}{1965} \times 10^{-4} \approx 6.2 \times 10^{-6}$.

### 3.6 Appendix

**Proof of Lemma 3.3.1** Under the third order condition, applying the expansion of the excesses above a high threshold as in the proof of Drees (1998, Theorem 2.1), one can prove that for each $\epsilon > 0$,

$$
\sup_{0 < s \leq 1} \min \left(1, s^{\gamma + 1/2 + \epsilon}\right) \left| \sqrt{k} \left( \frac{X_{n,n-[k\epsilon]} - X_{n,n-k}}{a(n/k)} - \frac{s^{-\gamma} - 1}{\gamma} \right) - s^{-\gamma-1}W_n(s) + W_n(1) 
- \sqrt{k}A(n/k)s^{-(\gamma+\rho)} - 1 \right| \xrightarrow{p} 0.
$$

Hence it follows that as $n \to \infty$,

$$
\frac{I_q}{a(n/k)} = \frac{1}{k} \sum_{i=0}^{k-1} \left( \frac{i}{k} \right)^{q-1} \frac{X_{n,n-i} - X_{n,n-k}}{a(n/k)} 
= \int_0^1 \left( \frac{[ks]}{k} \right)^{q-1} \frac{X_{n,n-[ks]} - X_{n,n-k}}{a(n/k)} ds 
= \int_0^1 s^{q-1} \left( \frac{s^{-\gamma} - 1}{\gamma} + \frac{1}{\sqrt{k}} \left( s^{-\gamma-1}W_n(s) - W_n(1) \right) + A(n/k)s^{-(\gamma+\rho)} - 1 \right) \rho(\gamma + \rho) 
+ A(n/k)B(n/k)O_p(1) + O_p \left( \frac{1}{k} \right) 
= \frac{1}{q(q-\gamma)} + \frac{L_q}{\sqrt{k}} \frac{B_q}{q(q-\gamma)} + A(n/k)B_q + \varepsilon_{n,k}.
$$
Here we apply $\left| \frac{[ks]}{k} - s \right| \leq \frac{1}{k}$ for the third equality.

Next we verify the covariance structure of $L_q$. Note that $L_q = q(q-\gamma)\int_0^1 W_n(s)s^{q-\gamma-2}ds - (q-\gamma)W_n(1)$ and $E((q-\gamma)W_n(1)) = 1$. Hence $E(L_qW_n(1)) = \gamma$ follows immediately. Furthermore, we have that

$$E(L_qL_r) = E\left(L_q \left( r(r-\gamma) \int_0^1 W_n(s)s^{r-\gamma-2}ds - (r-\gamma)W_n(1) \right) \right)$$

$$= E\left(r(r-\gamma)L_q \int_0^1 W_n(s)s^{r-\gamma-2}ds \right) - (r-\gamma)E(L_qW_n(1))$$

$$= qr(q-\gamma)(r-\gamma)E\left(\int_0^1 W_n(s)s^{q-\gamma-2}ds \int_0^1 W_n(t)t^{r-\gamma-2}dt \right) - r(q-\gamma) - \gamma(r-\gamma)$$

$$= qr\frac{q + r - 2\gamma}{q + r - 2\gamma - 1} - qr + \gamma^2 = \frac{qr}{q + r - 2\gamma - 1} + \gamma^2.$$

For the proof of the Theorems 3.3.1, 3.3.2 and 3.3.3, we need several lemmas. Firstly, we give a lemma on the uniform convergence of the third order condition. The proof is analogue to that in Fraga Alves et al. (2003), thus we omit it here.

**Lemma 3.6.1.** Suppose $F$ satisfies the third order condition (3.2.3) with $\rho, \rho' < 0$. There exists a particular choice of the first, second and third order scale functions $\tilde{a}(t)$, $\tilde{A}(t)$ and $\tilde{B}(t)$ such that for any $\delta > 0$, there exists a threshold $T(\delta)$, the inequality

$$\left| \frac{U(t_x) - U(t)}{A(t)} - \frac{x^{\gamma-1}}{\rho(\gamma+\rho')} - \frac{x^{\gamma+\rho+\rho'} - 1}{\gamma + \rho + \rho'} \right| < \delta x^{\gamma+\rho+\rho' + \delta}$$

holds for all $t, x$ such that $t, tx > T(\delta)$.

Without loss of generality, we still use the notations $a, A$ and $B$ for $\tilde{a}, \tilde{A}$ and $\tilde{B}$.

Lemma 3.6.2 and 3.6.3 give the asymptotic properties of the estimators of $\rho$ and $A(n/k)$.

**Lemma 3.6.2.** With an intermediate sequence $k_\rho$ satisfying the condition (3.2.5), we have that as $n \to \infty$,

$$\sqrt{k_\rho}A(n/k_\rho)(\hat{\rho}(k_\rho) - \rho) = O_p(1).$$

**Lemma 3.6.3.** With an intermediate sequence $k$ satisfying the condition (3.2.4), we have that as $n \to \infty$

$$\hat{A}(n/k) - A(n/k) = \frac{1}{\sqrt{k}} \left( \frac{1 - \gamma - \rho}{2\rho} \right) \frac{(1 - \gamma - \rho)(2 - \gamma - \rho)(3 - \gamma - \rho)}{(1 - \gamma)L_1}. $$
\[ -2(2-\gamma)L_2 + (3-\gamma)L_3 + A(n/k)A(n/k_p)O_p(1). \]  

(3.6.2)

**Corollary 3.6.1.** Under the condition in Lemma 3.6.3, \( \hat{\gamma}(n/k) \xrightarrow{p} 1 \) as \( n \to \infty \).

The following lemma gives the asymptotic property of an intermediate order statistics, which can be regarded as an estimator of \( U(n/k) \), see de Haan and Ferreira (2006).

**Lemma 3.6.4.** Suppose the third order condition (3.2.3) holds with \( \rho, \rho' < 0 \). Let \( k = k(n) \) satisfy \( \sqrt{k} |A(n/k)B(n/k)| = O(1) \), as \( n \to \infty \) Then

\[
\frac{X_{n,n-k} - U(n/k)}{a(n/k)} = \frac{1}{\sqrt{k}} W(1) + A(n/k)B(n/k)O_p(1),
\]

where \( W \) is the Brownian motion defined in Lemma 3.3.1.

Lemma 3.6.5 states the asymptotic property of the bias-corrected estimator of the scale function \( a \).

**Lemma 3.6.5.** Suppose the third order condition (3.2.3) holds with \( \gamma < 1/2, \rho, \rho' < 0 \) and the intermediate sequence \( k \) satisfies condition (3.2.4). For \( \hat{a}_{ub}(n/k) \) defined in (3.3.12), we have that, as \( n \to \infty \),

\[
\frac{\hat{a}_{ub}(n/k)}{a(n/k)} - 1 = \frac{1}{\sqrt{k}} \left( \frac{1 - \gamma}{2} \right) \left( \frac{1}{\rho} - \frac{1}{2 - \gamma} - \frac{1}{3 - \gamma} \right) L_1 \\
- 2(\gamma + \rho - 2) \left( \frac{1}{\rho} - \frac{1}{1 - \gamma} - \frac{1}{3 - \gamma} \right) L_2 \\
+ (\gamma + \rho - 3) \left( \frac{1}{\rho} - \frac{1}{1 - \gamma} - \frac{1}{2 - \gamma} \right) L_3 \right) + A(n/k)A(n/k_p)O_p(1).
\]

(3.6.3)

**Proof of Lemma 3.6.2** Since \( \rho = 1 - \gamma - \frac{1}{2 - \gamma - \rho} \), it is sufficient to notice that (3.3.5) implies—using condition (3.2.5)—that

\[
\sqrt{k} \rho A(n/k_p) \left( \frac{\hat{\gamma}_{3.1}(k_p) - \hat{\gamma}_{4.1}(k_p)}{\hat{\gamma}_{3.2}(k_p) - \hat{\gamma}_{4.2}(k_p)} - \frac{1 - \gamma}{2 - \gamma} \cdot \frac{2 - \gamma - \rho}{1 - \gamma - \rho} \right) = O_p(1)
\]

and that (3.3.4) implies (with condition (3.2.5)) that

\[
\sqrt{k} \rho A(n/k_p)(\hat{\gamma}_{2.1}(k_p) - \gamma) = O_p(1).
\]
Proof of Lemma 3.6.3 With $g(\gamma, \rho) := \frac{(1-\gamma-\rho)(2-\gamma-\rho)(3-\gamma-\rho)}{\rho(1-\gamma)}$, 
\[
\frac{A(n/k)}{\hat{A}(n/k)} - 1 = \left(\frac{g(\hat{\gamma}_{2,1}(k) \cdot \hat{\gamma}_{3,1}(k))}{g(\gamma, \rho)} - 1\right).
\]

By (3.3.8), the right hand side is approximately 
\[
\frac{g(\gamma, \rho) N_{3,2,1}}{\sqrt{\hat{k} A(n/k)}} + \left(\frac{g(\hat{\gamma}_{2,1}(k), \hat{\rho}(k_p))}{g(\gamma, \rho)} - 1\right).
\]

Thus it is sufficient to show that 
\[
g(\hat{\gamma}_{2,1}(k), \hat{\rho}(k_p)) = g(\gamma, \rho) + A(n/k)O_p(1),
\]
which follows from $\hat{\gamma}_{2,1}(k) - \gamma = A(n/k)O_p(1)$ (from (3.3.4)) and $\hat{\rho} - \rho = \frac{1}{\sqrt{\hat{k} A(n/k)}} O_p(1) = A(n/k)O_p(1)$ (from Lemma 3.6.2).

Proof of Lemma 3.6.4 Write $X_{n,n-k} = U(Y_{n,n-k})$ with $Y_{n,n-k}$ the corresponding intermediate order statistics from the distribution function $1 - 1/x$, $x > 1$. By taking $t = n/k$ and $x = \frac{k}{n} Y_{n,n-k}$ in (3.6.1), we get 
\[
\sqrt{k} \frac{X_{n,n-k} - U(n/k)}{a(n/k)} = \sqrt{k} \frac{\left(\frac{k}{n} Y_{n,n-k}\right)^\gamma - 1}{\gamma} + A(n/k) \sqrt{k} \frac{\left(\frac{k}{n} Y_{n,n-k}\right)^{\gamma + \rho} - 1}{\rho(\gamma + \rho)} + \sqrt{k} A(n/k) B(n/k) \left(\frac{\left(\frac{k}{n} Y_{n,n-k}\right)^{\gamma + \rho + \rho'} - 1}{\gamma + \rho + \rho'} + \left(\frac{k}{n} Y_{n,n-k}\right)^{\gamma + \rho + \rho'} o(1)\right).
\]

Since $\sqrt{k} \left(\frac{k}{n} Y_{n,n-k} - 1\right) \xrightarrow{d} W(1)$, as $n \to \infty$, we get that 
\[
\sqrt{k} \frac{X_{n,n-k} - U(n/k)}{a(n/k)} = W(1) + A(n/k) W(1) + o_p(1) + A(n/k) B(n/k) O_p(1) + \sqrt{k} A(n/k) B(n/k) o_p(1).
\]

Hence the result follows from $\sqrt{k} \left| A(n/k) B(n/k) \right| = O(1)$.

Proof of Lemma 3.6.5 Write $h(\gamma, \rho) := \frac{(1-\gamma)(2-\gamma-\rho)(3-\gamma-\rho)}{\rho(1-\gamma-\rho)(2-\gamma-\rho)}$. Then $\hat{a}_{ab}(n/k) = \hat{a}(n/k) \exp(-\hat{A}(n/k) h(\hat{\gamma}, \hat{\rho}))$. From Lemma 3.3.1, one can obtain that 
\[
\frac{\hat{a}(n/k)}{a(n/k)} = \frac{L_1}{L_2} \frac{\hat{a}(n/k)}{a(n/k)} = 1 + \frac{1}{\sqrt{k}} \left(\frac{\gamma - 1}{L_1} - \frac{\gamma - 2}{L_2}\right) + A(n/k) h(\gamma, \rho) + \epsilon_n.$
\]
(3.6.5)
Similar to the proof of Lemma 3.6.3, we get

\[ h(\hat{\gamma}, \hat{\rho}) = h(\gamma, \rho) + A(n/k_p)O_p(1) \quad (3.6.6) \]

Moreover, from Lemma 3.6.3, we get

\[ \bar{A}(n/k) = A(n/k) + \frac{1}{\sqrt{k}} N_A + A(n/k)A(n/k_p)O_p(1) \quad (3.6.7) \]

with

\[ N_A = \frac{(1-\gamma-\rho)(2-\gamma-\rho)(3-\gamma-\rho)}{2\rho} \left( (1-\gamma)L_1 - 2(2-\gamma)L_2 + (3-\gamma)L_3 \right). \]

By combining the expansions (3.6.5), (3.6.6) and (3.6.7), we have that

\[ \bar{a}_{ub}(n/k) = \frac{1}{\sqrt{k}} \left( (\gamma - 1)L_1 - (\gamma - 2)L_2 - h(\gamma, \rho)N_A \right) + A(n/k)A(n/k_p)O_p(1), \]

which can be verified as equivalent to (3.6.3).

**Proof of Theorem 3.3.1** Write \( b(\gamma, \rho) := \frac{(1-\gamma)(2-\gamma)}{(1-\gamma-\rho)(1-\gamma-\rho)} \). Following the same arguments as for (3.6.4), we have

\[ b(\gamma_{2,1}(k), \hat{\rho}(k)) = b(\gamma, \rho) + A(n/k_p)O_p(1). \]

Combining this with (3.3.4) and (3.6.7), we have

\[ \sqrt{k}(\gamma_{2,1} - \gamma) = \sqrt{k} \left( \gamma_{2,1}(k) - \gamma - \bar{A}(n/k)b(\gamma_{2,1}(k), \hat{\rho}(k_p)) \right) \]

\[ = - (2 - \gamma)(1 - \gamma)(L_2 - L_1) + \sqrt{k}A(n/k)b(\gamma, \rho) + \sqrt{k}\varepsilon_{n,k} \]

\[ - \left( \sqrt{k}A(n/k) + N_A + \sqrt{k}A(n/k)A(n/k_p)O_p(1) \right) \left( b(\gamma, \rho) + A(n/k_p)O_p(1) \right) \]

\[ = - \left( (2 - \gamma)(1 - \gamma)(L_2 - L_1) + b(\gamma, \rho)N_A \right) + \sqrt{k}\varepsilon_{n,k} + \sqrt{k}A(n/k)A(n/k_p)O_p(1) \]

\[ \Rightarrow \frac{1 - \gamma)(2 - \gamma)(3 - \gamma)}{2\rho} \left( (\gamma + \rho - 1)J_1 - 2(\gamma + \rho - 2)J_2 + (\gamma + \rho - 3)J_3 \right), \]

where the last convergence follows from \( \sqrt{k}\varepsilon_{n,k} \to 0 \) and \( \sqrt{k}A(n/k)A(n/k_p) \to 0 \) by condition (3.2.5). The limit distribution follows from the fact that \( (J_1, J_2, J_3) \) is normal distributed with mean vector zero and covariance matrix

\[
\begin{pmatrix}
(1 - 2\gamma)^{-1} + \gamma^2 & (1 - \gamma)^{-1} + \gamma^2 & 3(3 - 2\gamma)^{-1} + \gamma^2 \\
(1 - \gamma)^{-1} + \gamma^2 & 4(3 - 2\gamma)^{-1} + \gamma^2 & 3(2 - \gamma)^{-1} + \gamma^2 \\
3(3 - 2\gamma)^{-1} + \gamma^2 & 3(2 - \gamma)^{-1} + \gamma^2 & 9(5 - 2\gamma)^{-1} + \gamma^2
\end{pmatrix}.
\]
Proof of Theorem 3.3.2 \( \) Notice that \( x_p = U(1/p) \). Thus, we can write

\[
\frac{\sqrt{k}}{a(n/k)q_\gamma(a_n)}(\hat{x}_p - U(1/p)) = \frac{\sqrt{k}}{q_\gamma(a_n)} X_{n,n-k} - U(n/k)
\]

\[
+ \frac{\hat{a}_{ub}(n/k)}{a(n/k)} \frac{d^{\hat{n}}}{\gamma q_\gamma(a_n)} \left( \frac{d^{\hat{n}}}{\gamma} - 1 \right)
\]

\[
+ \frac{d^{\gamma_1}}{\gamma q_\gamma(a_n)} \sqrt{k} \left( \frac{\hat{a}_{ub}(n/k)}{a(n/k)} - 1 \right)
\]

\[
+ \frac{\sqrt{k}A(n/k)}{q_\gamma(a_n)} \left( \frac{A(n/k)\hat{a}(n/k) d^{\hat{n}}}{A(n/k) a(n/k)} \frac{d^{\hat{n}}}{\rho(\gamma + \hat{\rho})} - \frac{d^{\gamma_1}}{\rho(\gamma + \hat{\rho})} \right)
\]

\[
- \frac{\sqrt{k}A(n/k)}{q_\gamma(a_n)} \left( \frac{U(1/p) - U(n/k)}{a(n/k)} - \frac{d^{\gamma_1}}{\rho(\gamma + \hat{\rho})} \right)
\]

\[
= I_1 + I_2 + I_3 + I_4 + I_5.
\]

By Lemma 3.6.4, \( I_1 \xrightarrow{d} \Gamma_1 \). As in the proof of Theorem 4.3.1 in de Haan and Ferreira (2006, page 136-137), \( I_2 \xrightarrow{d} \Gamma_2 \). By Lemma 3.6.5, \( I_3 \xrightarrow{d} \Gamma_3 \).

For the term \( I_4 \), similar to the term \( I_2 \), we have that

\[
\frac{d^{\hat{n}}}{\hat{\gamma} + \hat{\rho}} - 1 - \frac{d^{\gamma_1}}{\gamma + \rho} = q_{\gamma + \rho}(d_n)(\hat{\gamma} + \hat{\rho} - \gamma - \rho)(1 + \alpha_p(1)) = A(n/k) q_{\gamma + \rho}(d_n) O_p(1).
\]

Since \( \frac{1}{\rho} = \frac{1}{\hat{\rho}}(1 + A(n/k) O_p(1)) \), we get

\[
\frac{\hat{a}_{ub}(n/k)}{a(n/k)} = 1 + \frac{1}{\sqrt{k}} O_p(1)
\]

(3.6.9)

and by Lemma 3.6.3

\[
\frac{\hat{A}(n/k)}{A(n/k)} = 1 + \frac{N_A}{\sqrt{k}A(n/k)}(1 + \alpha_p(1)).
\]

(3.6.10)

Combing (3.6.8), (3.6.9) and (3.6.10) we get the expansion of \( I_4 \) as

\[
I_4 = \frac{\sqrt{k}A(n/k)}{q_\gamma(a_n)} \frac{d^{\hat{n}}}{\hat{\rho}(\gamma + \hat{\rho})} \left( 1 + \frac{N_A}{\sqrt{k}A(n/k)}(1 + \alpha_p(1)) \right) \left( 1 + \frac{1}{\sqrt{k}} O_p(1) \right)
\]
\[
\left(1 + \frac{(\gamma + \rho)A(n/k^\rho)q_{\gamma+\rho}(d_n)}{d_n^{\gamma+\rho} - 1}O_p(1) \right) - 1
\]
\[
= \frac{d_n^{\gamma+\rho} - 1}{\rho(\gamma+\rho)q_{\gamma}(d_n)} N_A + \frac{d_n^{\gamma+\rho} - 1}{\rho(\gamma+\rho)q_{\gamma}(d_n)}o_p(1) + \frac{\sqrt{kA(n/k)A(n/k^\rho)q_{\gamma+\rho}(d_n)}}{q_{\gamma}(d_n)}O_p(1).
\]

Hence, together with the relation \(\lim_{n \to \infty} \frac{d_n^{\gamma+\rho} - 1}{\rho(\gamma+\rho)q_{\gamma}(d_n)} = -\frac{(\gamma - \rho^2)}{(\rho(\gamma + \rho))},\) we get \(I_4 \overset{d}{\to} \Gamma_4.\)

For part \(I_5,\) we use the inequality (3.6.1) and obtain that
\[
I_5 = -\frac{\sqrt{kA(n/k)}}{q_{\gamma}(d_n)} B(n/k) \left( \frac{d_n^{\gamma+\rho+\rho^\prime} - 1}{\gamma + \rho + \rho^\prime} + d_n^{\gamma+\rho+\rho^\prime}o(1) \right) = \sqrt{kA(n/k)B(n/k)}o(1) \to 0.
\]

**Proof of Theorem 3.3.3** The proof is similar to that of Theorem 3.3.2.
Chapter 4

Estimation of marginal expected shortfall: the mean when a related variable is extreme

[Based on joint work with John H.J. Einmahl and Laurens de Haan and Chen Zhou]

Abstract. Denote the loss return on the equity of a financial institution as $X$ and that of the entire market as $Y$. For a given very small value of $p > 0$, the marginal expected shortfall (MES) is defined as $E(X \mid Y > Q_Y(1-p))$, where $Q_Y(1-p)$ is the $(1-p)$-th quantile of the distribution of $Y$. The MES is an important factor when measuring the systemic risk of financial institutions. For a wide nonparametric class of bivariate distributions, we construct an estimator of the MES and establish the asymptotic normality of the estimator when $p \downarrow 0$, as the sample size $n \to \infty$. Since we are in particular interested in the case $p = O(1/n)$, we use extreme value techniques for deriving the estimator and its asymptotic behavior. The finite sample performance of the estimator and the adequacy of the limit theorem are shown in a detailed simulation study. We also apply our method to estimate the MES of three large U.S. investment banks.
Chapter 4. *Estimation of marginal expected shortfall* 68

4.1 Introduction

The financial crisis has led market participants, academics and regulators to further investigate the systemic risk in the financial industry. One key issue is to have a measure of systemic risk that is both economically sound and statistically assessable. Among others, Acharya et al. (2010) proposes a systemic risk measure under a general equilibrium model. An important factor in constructing this systemic risk measure is the contribution of a financial institution to a systemic crisis measured by the Marginal Expected Shortfall (MES). The MES of a financial institution is defined as the expected loss on its equity return conditional on the occurrence of an extreme loss in the aggregated return of the financial sector. In this paper, we provide an empirical methodology based on multivariate extreme value theory (EVT) to estimate the MES.

Denote the loss of the equity return of a financial firm and that of the entire market as $X$ and $Y$, respectively. Then the MES is defined as $E(X \mid Y > t)$, where $t$ is a high threshold such that $p = P(Y > t)$ is extremely small. In other words, the threshold $t$ is the $(1 - p)$-th quantile of the distribution of $Y$ defined by $P(Y > Q_Y(1 - p)) = p$. Thus, the MES at probability level $p$ is defined as

$$MES(p) = E(X \mid Y > Q_Y(1 - p)).$$

Notice that in applications the probability $p$ is at an extremely low level that can be even lower than $1/n$, where $n$ is the sample size of historical data that are used for estimating the MES. Therefore, this is a typical extreme value problem.

We tackle this problem by a two-stage approach. Firstly, we consider the estimation of the MES at an intermediate probability level. More specifically, we consider an intermediate sequence $k = k(n)$ such that $k/n \to 0$, $k \to \infty$, as $n \to \infty$, and estimate $MES(k/n)$. At such an intermediate level, there are many observations $(X, Y)$ such that $Y > Q_Y(1 - k/n)$. Thus the $MES(k/n)$ can be estimated non-parametrically by taking the average over the $X$ components of those selected observations. Secondly, we follow the usual extrapolation method in extreme value statistics to obtain an estimator of $MES(p)$ for the intended low probability level $p$. Assuming that $X$ follows a heavy-tailed distribution with finite mean, and the dependence structure of $(X, Y)$ follows the bivariate EVT framework, we show that the MES can be extrapolated in a similar way.

---

1In Acharya et al. (2010), the probability of such an extreme tail event is specified as “that happen once or twice a decade (or less)”, whereas the estimation is based on daily data from one year.
as that for high quantiles of the distribution of $X$. More precisely, we estimate $MES(p)$ by multiplying the estimator of $MES(k/n)$ with an extrapolation factor, which is the same as that used in extrapolating the $(1 - k/n)$-quantile to the $(1 - p)$-th quantile of the distribution of $X$. Our main result establishes the asymptotic normality of the thus obtained estimator of $MES(p)$. This result makes statistical inference for the MES feasible for a large class of bivariate distributions. In addition, we show through a simulation study that the estimator performs well and that this limit theorem provides an adequate approximation for finite sample sizes.

The MES has been studied under the name “Conditional Tail Expectation” (CTE, or TCE) in statistics and actuarial science. The definition of CTE in a univariate context is the same as that of the tail value at risk. Mathematically, it is given by $E(X \mid X > Q_X(1-p))$ where $Q_X(1-p)$ is the $(1-p)$-th quantile of the distribution of $X$. In case $X$ has a continuous distribution, this is also called the expected shortfall. Compared to the MES, it can be viewed as the special case that $Y = X$. The concept of CTE has been defined more generally in a multivariate setup. It is possible to have the conditioning event defined by another, related random variable $Y$ exceeding its high quantile. In that case, the CTE coincides with the MES.

A few studies show how to calculate the CTE when the joint distribution of $(X, Y)$ follows specific parametric models. For example, Landsman and Valdez (2003) and Kostadinov (2006) deal with elliptical distributions with heavy-tailed marginals. Cai and Li (2005) studies the CTE for multivariate phase-type distributions. Vernic (2006) considers skewed-normal distributions. Compared to these studies, our approach does not impose any parametric structure on $(X, Y)$. A comparable result in the literature is the approach in Joe and Li (2011), where under multivariate regular variation, a formula for calculating the CTE is provided. The multivariate regularly varying distributions form a subclass of our model. Note that we do not make any assumption on the marginal distribution of $Y$. Furthermore, in contrast to these papers where only probabilistic calculations of the MES are provided, we focus on the statistical problem of estimating it and studying the performance of the estimator.

In Acharya et al. (2010), an estimation procedure for the MES is provided which is similar to our approach. Nevertheless, in order to obtain such an estimator, the probabilistic model is restricted to a specific linear relationship between $X$ and $Y$. Such an assumption falls into the scope of our general framework. Thus their approach can be seen as a special case of our much broader setup. A similar parametric approach
has been adopted in Brownlees and Engle (2011), where, as one step forward, a non-parametric kernel estimator of the MES is proposed. Such a kernel estimation method, however, is valid only if the threshold for defining a systemic crisis is not too high, that is, there are still several observations corresponding to the event $\{Y > Q_Y(1-p)\}$. In other words, the tail probability level $p$ should be substantially larger than $1/n$. Such a method cannot handle extreme events, that is $p < 1/n$, which is particularly required for systemic risk measures.

To summarize, compared to existing studies on MES or CTE, four advantages arise for our method. Firstly, we allow a wide, nonparametric class of dependence structures of $(X,Y)$, which covers in particular most of the parametric models used in literature. Secondly, we do not impose any assumption on the marginal distribution of $Y$. This allows other potential applications of our estimation method on the MES measure. For example, in the context of measuring systemic risk, the condition $\{Y > Q_Y(1-p)\}$ is considered as the definition of a “systemic crisis”. Our distribution-free feature on $Y$ allows other potential indicators in defining “systemic crisis”. Thirdly, we handle an extreme value statistics problem: we provide a novel estimator of $\text{MES}(p)$, corresponding to an extremely low probability level $p$, that can be lower than $1/n$. Lastly, the main contribution of this paper is the unraveling of the asymptotic behavior of this estimator.

The paper is organized as follows. Section 2 provides the main result: asymptotic normality of the estimator. In Section 3, a simulation study shows the good performance of the estimator. An application on estimating the MES for U.S. financial institutions is given in Section 4. The proofs are deferred to Section 5.

### 4.2 Main Results

Let $(X,Y)$ be a random vector with a continuous distribution function $F$. Denote the marginal distribution functions as $F_1(x) = F(x, \infty)$ and $F_2(y) = F(\infty, y)$ with corresponding tail quantile functions given by $U_j = \left(\frac{1}{1-F_j} \right)^{\leftarrow}$, $j = 1, 2$, where $\leftarrow$ denotes the left-continuous inverse. Then the MES at a probability level $p$ can be written as

$$\theta_p := E(X \mid Y > U_2(1/p)).$$

The goal is to estimate $\theta_p$ based on independent and identically distributed (i.i.d.) observations, $(X_1, Y_1), \ldots, (X_n, Y_n)$ from $F$, where $p = p(n) \to 0$ as $n \to \infty$. 

We adopt the bivariate EVT framework for modeling the tail dependence structure of \((X,Y)\). Suppose for all \((x,y) \in [0, \infty]^2 \setminus \{(+\infty, +\infty)\}\), the following limit exists:

\[
\lim_{t \to \infty} t P(1 - F_1(X) \leq x/t, 1 - F_2(Y) \leq y/t) =: R(x,y).
\] (4.2.1)

The function \(R\) completely determines the so-called stable tail dependence function \(l\), as for all \(x, y \geq 0\),

\[l(x,y) = x + y - R(x,y);\]

see Drees and Huang (1998); Beirlant et al. (2004, Chapter 8.2).

For the marginal distributions, we assume that only \(X\) follows a distribution with a heavy right tail: there exists \(\gamma_1 > 0\) such that for \(x > 0\),

\[
\lim_{t \to \infty} \frac{U_1(tx)}{U_1(t)} = x^{\gamma_1}.
\] (4.2.2)

Then it follows that \(1 - F_1\) is regularly varying with index \(-\frac{1}{\gamma_1}\) and \(\gamma_1\) is the extreme value index.

We first focus on \(X\) being positive, then we consider \(X \in \mathbb{R}\). Throughout, there is no assumption, apart from continuity, on the marginal distribution of \(Y\).

### 4.2.1 \(X\) Positive

Assume \(X\) takes values in \((0, \infty)\). The following limit result gives an approximation for \(\theta_p\).

**Proposition 4.2.1.** Suppose that (4.2.1) and (4.2.2) hold with \(0 < \gamma_1 < 1\). Then,

\[
\lim_{\nu \downarrow 0} \frac{\theta_p}{U_1(1/p)} = \int_0^{\infty} R\left(x^{-1/\gamma_1}, 1\right) dx.
\]

In Joe and Li (2011, Theorem 2.4), this result is derived under the stronger assumption of multivariate regular variation.

Next, we construct an estimator of \(\theta_p\) based on the limit given in Proposition 4.2.1. Let \(k\) be an intermediate sequence of integers, that is, \(k \to \infty, k/n \to 0, \text{ as } n \to \infty\).

By Proposition 4.2.1 and a strengthening of (4.2.2) (see condition (b) below), we have
that as \(n \to \infty\),
\[
\theta_p \sim \frac{U_1(1/p)}{U_1(n/k)} \frac{k}{np} \frac{\gamma_1}{\pi} \theta_{\frac{k}{\pi}}. 
\]
For estimating \(\theta_p\), it thus suffices to estimate \(\gamma_1\) and \(\theta_{\frac{k}{\pi}}\).

We estimate \(\gamma_1\) with the Hill (1975) estimator:
\[
\hat{\gamma}_1 = \frac{1}{k_1} \sum_{i=1}^{k_1} \log X_{n-i+1,n} - \log X_{n-k_1,n}, 
\]
where \(k_1\) is another intermediate sequence of integers and \(X_{i,n}, i = 1, \ldots, n\) is the \(i\)-th order statistic of \(X_1, \ldots, X_n\).

By regarding the \((n-k)\)-th order statistic \(Y_{n-k,n}\) of \(Y_1, \ldots, Y_n\) as an estimator of \(U_2(n/k)\), we construct a nonparametric estimator of \(\theta_{\frac{k}{\pi}}\) which is the average of the selected \(X_i\) corresponding to the highest \(k\) values of \(Y\):
\[
\hat{\theta}_{\frac{k}{\pi}} = \frac{1}{k} \sum_{i=1}^{n} X_i I(Y_i > Y_{n-k,n}). 
\]

Combining (4.2.3), (4.2.4) and (4.2.5), we estimate \(\theta_p\) by
\[
\hat{\theta}_p = \left(\frac{k}{np}\right) \frac{\gamma_1}{\pi} \hat{\theta}_{\frac{k}{\pi}}. 
\]

We prove the asymptotic normality of \(\hat{\theta}_p\) under the following conditions.

(a) There exist \(\beta > \gamma_1\) and \(\tau < 0\) such that, as \(t \to \infty\),
\[
\sup_{0 < x < \infty} \frac{\left| tP(1 - F_1(X) < x/t, 1 - F_2(Y) < y/t) - R(x, y) \right|}{x^\beta \wedge 1} = O(t^\tau).
\]

(b) There exist \(\rho_1 < 0\) and an eventually positive or negative function \(A_1\) with \(\lim_{t \to \infty} A_1(t) = 0\) such that
\[
\lim_{t \to \infty} \frac{U_1(tx)/U_1(t) - x^{\gamma_1}}{A_1(t)} = x^{\gamma_1} x^{\rho_1} - 1. 
\]

As a consequence, \(|A_1|\) is regularly varying with index \(\rho_1\). Conditions (a) and (b) are natural second-order strengthenings of (4.2.1) and (4.2.2), respectively. We further require the following conditions on the intermediate sequences \(k_1\) and \(k\).
(c) As \( n \to \infty \), \( \sqrt{k_1}A_1(n/k_1) \to 0 \).

(d) As \( n \to \infty \), \( k = O(n^\alpha) \) for some \( \alpha < \min\left(\frac{2\gamma_1}{2\gamma_1 + 1}, \frac{2\gamma_1}{2\gamma_1 + \rho_1 - 1}\right) \).

To characterize the limit distribution of \( \hat{\theta}_p \), we define a mean zero Gaussian process \( W_R \) on \([0, \infty) \times (\infty, \infty)\) with covariance structure

\[
E(W_R(x_1, y_1)W_R(x_2, y_2)) = R(x_1 \wedge x_2, y_1 \wedge y_2),
\]

i.e., \( W_R \) is a Wiener process. Set

\[
\Theta = (\gamma_1 - 1)W_R(\infty, 1) + \left( \int_0^\infty R(s, 1)ds^{-\gamma_1} \right)^{-1} \int_0^\infty W_R(s, 1)ds^{-\gamma_1},
\]

and

\[
\Gamma = \gamma_1 \left( -W_R(1, \infty) + \int_0^1 s^{-1}W_R(s, \infty)ds \right).
\]

It will be shown (see Proposition 4.5.2 and (4.5.15)) that \( \hat{\theta}_p \) and \( \hat{\gamma}_1 \) are asymptotically normal with \( \Theta \) and \( \Gamma \) as limit, respectively. The following theorem gives the asymptotic normality of \( \hat{\theta}_p \).

**Theorem 4.2.1.** Suppose conditions (a)–(d) hold and \( \gamma_1 \in (0, 1/2) \). Assume \( d_n := \frac{k}{n^p} \geq 1 \) and \( r := \lim_{n \to \infty} \frac{\sqrt{\log d_n}}{\sqrt{|x|_1}} \in [0, \infty] \). If \( \lim_{n \to \infty} \frac{\log d_n}{\sqrt{|x|_1}} = 0 \), then, as \( n \to \infty \),

\[
\min\left( \sqrt{k}, \frac{\sqrt{k_1}}{\log d_n} \right) \frac{\hat{\theta}_P - 1}{\theta_P} \xrightarrow{d} \begin{cases} 
\Theta + r\Gamma, & \text{if } r \leq 1, \\
\frac{1}{2}\Theta + \Gamma, & \text{if } r > 1,
\end{cases}
\]

where \( \text{Var}(\Theta) = (\gamma_1^2 - 1) - 2 \int_0^\infty R(s, 1)ds^{-2\gamma_1} \), \( \text{Var}(\Gamma) = \gamma_1^2 \) and

\[
\text{Cov}(\Gamma, \Theta) = \gamma_1(1-\gamma_1+b)R(1,1)-\gamma_1 \int_0^1 ((1-\gamma_1)+bs^{-\gamma_1}(1-\gamma_1-\gamma_1\ln s)) R(s,1)s^{-1}ds
\]

with \( b = \left( \int_0^\infty R(s,1)ds^{-\gamma_1} \right)^{-1} \).

### 4.2.2 X Real

In this section, \( X \) takes values in \( \mathbb{R} \), that is, we do not restrict \( X \) to be positive. Define \( X^+ = \max(X, 0) \) and \( X^- = X - X^+ \). Besides the conditions of Theorem 4.2.1, we require two more conditions:

(e) \( E|X^-|^{1/\gamma_1} < \infty \);
(f) As \( n \to \infty \), \( k = o \left( p^{2\tau(1-\gamma_1)} \right) \).

It can be shown that condition (e), together with (a), ensure that \( \theta_p \sim E(X^+ \mid Y > U_2(1/p)) \), as \( p \downarrow 0 \). Therefore, we estimate \( \theta_p \) with

\[
\hat{\theta}_p = \left( \frac{k}{np} \right)^{\hat{\gamma}_1} \frac{1}{k} \sum_{i=1}^{n} X_i I(X_i > 0, Y_i > Y_{n-k,n}),
\]

(4.2.7)

with \( \hat{\gamma}_1 \) as in Section 4.2.1. Observe that when \( X \) is positive, this definition coincides with that in (4.2.6). As stated in the following theorem, the asymptotic behavior of the estimator remains the same as that for positive \( X \).

**Theorem 4.2.2.** Under the conditions of Theorem 4.2.1 and conditions (e) and (f), as \( n \to \infty \),

\[
\min \left( \sqrt{k}, \frac{\sqrt{k}}{\log d_n} \right) \left( \frac{\hat{\theta}_p}{\theta_p} - 1 \right) d \to \begin{cases} 
\Theta + r\Gamma, & \text{if } r \leq 1; \\
\frac{1}{\theta_p} \Theta + \Gamma, & \text{if } r > 1,
\end{cases}
\]

where \( r, \Theta \) and \( \Gamma \) are defined as in Theorem 4.2.1.

### 4.3 Simulation Study

In this section, a simulation and comparison study is implemented to investigate the finite sample performance of our estimator. We generate data from three bivariate distributions.

- **A transformed Cauchy distribution on \((0, \infty)^2\)** defined as

\[
(X,Y) = \left( |Z_1|^{2/5}, |Z_2| \right),
\]

where \((Z_1, Z_2)\) is a standard Cauchy distribution on \( \mathbb{R}^2 \) with density \( \frac{1}{2\pi} (1 + x^2 + y^2)^{-3/2} \). It follows that \( \gamma_1 = 2/5 \) and \( R(x,y) = x + y - \sqrt{x^2 + y^2}, x, y \geq 0 \). It can be shown that this distribution satisfies conditions (a) and (b) with \( \tau = -1 \), \( \beta = 2 \), and \( \rho_1 = -2 \). We shall refer to this distribution as “transformed Cauchy distribution (1)”.  

- **Student-\(t_3\) distribution on \((0, \infty)^2\)** with density

\[
f(x,y) = \frac{2}{\pi} \left( 1 + \frac{x^2 + y^2}{3} \right)^{-5/2}, \quad x, y > 0.
\]
We have $\gamma_1 = 1/3$, $R(x, y) = x + y - \frac{x^{4/3} + \frac{1}{3}x^{2/3}y^{2/3} + y^{4/3}}{\sqrt{x^{2/3} + y^{2/3}}}$, $\tau = -1/3$, $\beta = 4/3$ and $\rho_1 = -2/3$.

- A transformed Cauchy distribution on the whole $\mathbb{R}^2$ defined as 
  
  $$(X, Y) = \left( Z_1^{2/5} I (Z_1 \geq 0) + Z_1^{1/5} I (Z_1 < 0), Z_2 I (Z_1 \geq 0) + Z_2^{1/3} I (Z_1 < 0) \right).$$

  We have $\gamma_1 = 2/5$, $R(x, y) = x/2 + y - \sqrt{x^2/4 + y}$, $\tau = -1$, $\beta = 2$, and $\rho_1 = -2$.
  We shall refer to this distribution as “transformed Cauchy distribution (2)”. 

We draw 500 samples from each distribution with sample sizes $n = 2,000$ and $n = 5,000$. Based on each sample, we estimate $\theta_p$ for $p = 1/500$, $1/5,000$ or $1/10,000$.

Besides the estimator given by (4.2.7), we construct two other estimators. Firstly, for $np \geq 1$, an empirical counterpart of $\theta_p$, given by

$$\hat{\theta}_{\text{emp}} = \frac{1}{[np]} \sum_{i=1}^{n} X_i I(Y_i > Y_{n-[np], n}), \quad (4.3.1)$$

is studied, where $[\cdot]$ denotes the integer part. Secondly, exploiting the relation in Proposition 4.2.1 and using the empirical estimator of $R$ given by $\hat{R}(x, y) = \frac{1}{k} \sum_{i=1}^{n} I(X_i > X_{n-[kx], n}, Y_i > Y_{n-[ky], n})$ and the Weissman (1978) estimator of $U_1(1/p)$ given by $\hat{U}_1(1/p) = d_1^{\gamma_1} X_{n-k,n}$, we define an alternative EVT estimator as

$$\tilde{\theta}_p = - \hat{U}_1(1/p) \int_0^\infty \hat{R}(x, 1) dx^{-\gamma_1}$$

$$= d_1^{\gamma_1} X_{n-k,n} \frac{1}{k} \sum_{i=1}^{n} I(Y_i > Y_{n-k,n}) \left( \frac{n - \text{rank}(X_i) + 1}{k} \right)^{-\gamma_1}. \quad (4.3.2)$$

The comparison of the three estimators is shown in Figure 4.1, where we present boxplots of the ratios of the estimates and the true values. For all three distributions, the empirical estimator underestimates the MES and is consistently outperformed by the EVT estimators. Additionally, it is not applicable for $p < 1/n$. The two EVT estimators, $\tilde{\theta}_p$ and $\hat{\theta}_p$, both perform well. Their behavior is similar and remains stable when $p$ changes from $1/500$ to $1/10,000$. The results for the transformed Cauchy distribution (1) are the best among the three distributions, as the medians of the ratios are closest to one and the variations are smallest.
Chapter 4. Estimation of marginal expected shortfall

Figure 4.1: Boxplots on ratios of estimates and true values. Each plot is based on 500 samples with sample size \(n = 2000\) or 5000 from the transformed Cauchy distributions (1), (2) or Student-\(t_3\) distribution. The estimators are \(\hat{\theta}_p\) of (4.2.7), \(\bar{\theta}_p\) of (4.3.2) and \(\hat{\theta}_{emp}\) of (4.3.1); \(p = 1/500\) (p1), 1/5000 (p2) and 1/10000 (p3).

Next, we investigate the normality of the estimator, \(\hat{\theta}_p\), with \(p = 1/n\). For \(r < \infty\), the asymptotic normality of \(\hat{\theta}_p\) in Theorem 4.2.1 can be expressed as

\[\sqrt{k} \left( \hat{\theta}_p - 1 \right) \xrightarrow{d} \Theta + r\Gamma,\]

or equivalently,

\[\sqrt{k} \log \frac{\hat{\theta}_p}{\theta_p} \xrightarrow{d} \Theta + r\Gamma.\]
Notice that the limit distribution is a centered normal distribution. Write $\sigma_p^2 = \frac{1}{k} \text{Var}(\Theta + r\Gamma)$ with $r = \frac{\sqrt{k} \log \frac{\hat{\sigma}}{\sigma_p}}{\sqrt{k_1}}$. We compare the distribution of $\log \frac{\hat{\sigma}}{\sigma_p}$ with the limit distribution $N(0, \sigma_p^2)$. Table 4.1 reports the standardized mean of $\log \frac{\hat{\sigma}}{\sigma_p}$, i.e., the

<table>
<thead>
<tr>
<th></th>
<th>$n = 2,000$</th>
<th>$n = 5,000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1/2,000$</td>
<td>0.152 (1.027)</td>
<td>0.107 (1.054)</td>
</tr>
<tr>
<td>$p = 1/5,000$</td>
<td>0.232 (0.929)</td>
<td>0.148 (0.964)</td>
</tr>
<tr>
<td>Transformed Cauchy distribution (1)</td>
<td>-0.147 (1.002)</td>
<td>-0.070 (1.002)</td>
</tr>
<tr>
<td>Student-$t_3$ distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transformed Cauchy distribution (2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbers are the standardized mean of $\log \frac{\hat{\sigma}}{\sigma_p}$ and between brackets, the ratio of the standard deviation and $\sigma_p$, based on 500 estimates with $n = 2,000$ or $5,000$ and $p = 1/n$. The average value divided by $\sigma_p$, and between brackets, the ratio of the “sample” standard deviation and $\sigma_p$. As indicated by the numbers, the mean and standard deviation of $\log \frac{\hat{\sigma}}{\sigma_p}$ are both close to that of the limit distribution.

After the numerical assessment on the parameters, we illustrate the normality of $\log \frac{\hat{\sigma}}{\sigma_p}$. Figure 4.2 shows the densities of the $N(0, \sigma_p^2)$-distribution and the histograms of $\log \frac{\hat{\sigma}}{\sigma_p}$, based on 500 estimates. The normality of the estimates is supported by the large overlap between the histograms and the areas under the density curves. Hence we conclude that the limit theorem provides an adequate approximation for finite sample sizes.

### 4.4 Application

In this section, we apply our estimation method to estimate the MES for some financial institutions. We consider three large investment banks in the U.S., namely, Goldman Sachs (GS), Morgan Stanley (MS) and T. Rowe Price (TROW), all of which have a market capitalization greater than 5 billion USD as of the end of June 2007. The dataset consists of the loss returns (i.e., minus log returns) on their equity prices at a daily frequency from July 3, 2000 to June 30, 2010.\(^2\) Moreover, for the same time

\(^2\)The choice of the banks, data frequency and time horizon follows the same setup as in Brownlees and Engle (2011).
Chapter 4. *Estimation of marginal expected shortfall*

Figure 4.2: Histograms of $\hat{\theta}_p$ for $p = 1/n$, based on 500 samples with sample size $n = 2000$ or $5000$ from the transformed Cauchy distributions (1), (2) or Student-$t_3$ distribution. The choices of $k$ and $k_1$ are the same as in Figure 4.1. The curves are the densities of the $N(0, \sigma_p^2)$-distribution.

We extract daily loss returns of a value weighted market index aggregating three markets: NYSE, AMES and Nasdaq. We use our method to estimate the MES, $E(X \mid Y > U_2(1/p))$, where $X$ and $Y$ refer to the daily loss returns of a bank equity and the market index, respectively and $p = 1/n = 1/2513$, that corresponds to a once per decade systemic event.
Chapter 4. Estimation of marginal expected shortfall

Figure 4.3: The Hill estimates of the extreme value indices of the daily loss returns on the three equities.

Since $X$ may take negative values (i.e. positive returns of the equities of the banks), it is necessary to apply the estimator for the general case as defined in (4.2.7). For that purpose, we first verify two of the conditions required for the procedure. First of all, the assumption that $\gamma_1 < 1/2$ is confirmed by the plot of the Hill estimates in Figure 4.3. Secondly, since the estimation relies on the approximation of $\theta_p \sim E(X^+ | Y > U_2(1/p))$, it is important to check that high values of $Y$ do not coincide with negative values of $X$, generally. Intuitive empirical evidence for this is presented in Figure 4.4. It plots the loss returns of the equity prices against the market index. The horizontal lines indicate the 50-th largest loss of the index. As one can see, from the upper parts of the plots, the largest 50 losses of the index are mostly associated with losses ($X > 0$).

Hence, we can apply our method to obtain the estimates of $\gamma_1$ and $MES(p) = \theta_p$ for the three banks, see Table 4.2. It follows that in case of a once per decade market crisis, we estimate that on average the equity prices of Goldman Sachs and T. Rowe Price drop about 25% and Morgan Stanley falls even about 45% on that day.
Chapter 4. *Estimation of marginal expected shortfall* 80

**Figure 4.4:** The points are the daily loss returns of the three equity prices and the market index. The horizontal lines indicate the 50-th largest loss of the market index. The vertical lines, at 0, distinguish the occurrence of losses and profits.

**Table 4.2:** MES of the three investment banks

<table>
<thead>
<tr>
<th>Bank</th>
<th>$\hat{\gamma}_1$</th>
<th>$\hat{\theta}_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldman Sachs (GS)</td>
<td>0.386</td>
<td>0.301</td>
</tr>
<tr>
<td>Morgan Stanley (MS)</td>
<td>0.473</td>
<td>0.593</td>
</tr>
<tr>
<td>T. Rowe Price (TROW)</td>
<td>0.379</td>
<td>0.312</td>
</tr>
</tbody>
</table>

Here $\hat{\gamma}_1$ is computed by taking the average of the Hill estimates for $k_1 \in [70, 90]$. $\hat{\theta}_p$ is given in (4.2.7) with $n = 2513$, $k = 50$, and $p = 1/n = 1/2513$.

### 4.4.1 Discussion on Serial Dependence

The sound property of our estimator is proved under the assumption of independence. Failure of the independence assumption would lead to a consistent however less efficient estimator, see Drees (2003) and Poon et al. (2004). In order to assess the standard errors more precisely, one might apply bootstrap, see Ferro and Segers (2003). Rigorous study on this issue would make another interesting research project and hence out of the scope the paper.

We apply our method to the tri-daily loss returns. Denote $(X, Y)$ as the tri-daily loss returns. The sample size is now $n = 837$. In order to consider still a once per decade
event, we choose \( p = 1/n = 1/837 \). The tail dependence between \((X, Y)\) is supported by the data plot in Figure 4.5. The estimates are reported in Table 4.3.

![Figure 4.5](image)

**Figure 4.5:** The points are the tri-daily loss returns of the three equity prices and the market index from July 3, 2000 to June 30, 2010. The horizontal lines indicate the 30-th largest loss of the market index. The vertical lines, at 0, distinguish the occurrence of losses and profits.

**Table 4.3:** MES of the three investment banks

<table>
<thead>
<tr>
<th>Bank</th>
<th>( \hat{\gamma}_1 )</th>
<th>( \hat{\theta}_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldman Sachs (GS)</td>
<td>0.399</td>
<td>0.323</td>
</tr>
<tr>
<td>Morgan Stanley (MS)</td>
<td>0.410</td>
<td>0.496</td>
</tr>
<tr>
<td>T. Rowe Price (TROW)</td>
<td>0.285</td>
<td>0.223</td>
</tr>
</tbody>
</table>

Here \( \hat{\gamma}_1 \) is computed by taking the average of the Hill estimates for \( k_1 \in [20, 40] \). \( \hat{\theta}_p \) is given in (4.2.7) with \( n = 837 \), \( k = 30 \), and \( p = 1/n = 1/837 \).

## 4.5 Proofs

**Proof of Proposition 4.2.1** Recall that for a non-negative random variable \( Z \),

\[
E(Z) = \int_0^\infty P(Z > x)dx.
\]
Hence,
\[
\frac{\theta_p}{U_1(1/p)} = \int_0^\infty \frac{1}{p} P(X > x, Y > U_2(1/p)) \frac{dx}{U_1(1/p)} = \int_0^\infty \frac{1}{p} P(X > U_1(1/p)x, Y > U_2(1/p))dx.
\]
(4.5.1)

The limit relations (4.2.1) and (4.2.2) implies that
\[
\lim_{p \downarrow 0} \frac{1}{p} P(X > U_1(1/p)x, Y > U_2(1/p)) = R(x^{-1/\gamma_1}, 1).
\]

Hence, we only have to prove that the integral in (4.5.1) and the limit procedure \( p \to 0 \) can be interchanged. This is ensured by the dominated convergence theorem as follows. Notice that for \( x \geq 0 \),
\[
\frac{1}{p} P(X > U_1(1/p)x, Y > U_2(1/p)) \leq \min\left(1, \frac{1}{p} (1 - F_1(U_1(1/p)x))\right).
\]
For \( 0 < \varepsilon < 1/\gamma_1 - 1 \), there exists \( p(\varepsilon) \) (see Proposition B.1.9.5 in de Haan and Ferreira (2006)) such that for all \( p < p(\varepsilon) \) and \( x > 1 \),
\[
\frac{1}{p} (1 - F_1(U_1(1/p)x)) \leq 2x^{-1/\gamma_1 + \varepsilon}.
\]
Write
\[
h(x) = \begin{cases} 1, & 0 \leq x \leq 1; \\ 2x^{-1/\gamma_1 + \varepsilon}, & x > 1. \end{cases}
\]
Then \( h \) is integrable and \( \frac{1}{p} P(X > U_1(1/p)x, Y > U_2(1/p)) \leq h(x) \) on \([0, \infty)\) for \( p < p(\varepsilon) \). Hence we can apply the dominated convergence theorem to complete the proof of the proposition. \( \square \)

Next, we prove Theorem 4.2.1. The general idea of the proof is described as follows. It is clear that the asymptotic behavior of \( \hat{\theta}_p \) results from that of \( \hat{\gamma}_1 \) and \( \hat{\theta}_k_n \). The asymptotic normality of \( \hat{\gamma}_1 \) is well-known, see, e.g., de Haan and Ferreira (2006). To prove the asymptotic normality of \( \hat{\theta}_k_n \), write
\[
\hat{\theta}_k_n = \frac{1}{k} \sum_{i=1}^n X_i I(Y_i > U_2(n/(k\epsilon_n))),
\]
where $e_n = \frac{n}{k} (1 - F_2(Y_{n-k}, n)) \xrightarrow{P} 1$, as $n \to \infty$. Hence, with denoting $\hat{\theta}_{ky} := \frac{1}{ky} \sum_{i=1}^n X_i I(Y_i > U_2(n/(ky)))$, we first investigate the asymptotic behavior of $\hat{\theta}_{ky}$ for $y \in [1/2, 2]$. Then, by applying the result for $y = e_n$ and considering the asymptotic behavior of $e_n$, we obtain the asymptotic normality of $\hat{\theta}_k$. Lastly, together with the asymptotic normality of $\hat{\gamma}_1$, we prove that of $\hat{\theta}_p$.

To obtain the asymptotic behavior of $\hat{\theta}_{ky}$, we introduce some new notation and auxiliary lemmas. Write $R_n(x, y) := \frac{n}{k} \mathbb{P}(1 - F_1(X) < kx/n, 1 - F_2(Y) < ky/n)$. A pseudo non-parametric estimator of $R_n$ is given as $T_n(x, y) := \frac{1}{k} \sum_{i=1}^n I(1 - F_1(X_i) < kx/n, 1 - F_2(Y_i) < ky/n)$. The following lemma shows the asymptotic behavior of the pseudo estimator. The limit process is characterized by the aforementioned $W_R$ process. For convenient presentation, all the limit processes involved in the lemma are defined on the same probability space, via the Skorohod construction. However, they are only equal in distribution to the original ones. The proof of the lemma is analogous to that of Proposition 3.1 in Einmahl et al. (2006) and is thus omitted.

**Lemma 4.5.1.** Suppose (4.2.1) holds. For any $\eta \in [0, 1/2)$ and $T$ positive, with probability 1,

\[
\sup_{x, y \in (0, T]} \left| \frac{\sqrt{k}(T_n(x, y) - R_n(x, y)) - W_R(x, y)}{x^\eta} \right| \xrightarrow{\mathbb{P}} 0,
\]

\[
\sup_{x \in (0, T]} \left| \frac{\sqrt{k}(T_n(x, \infty) - x) - W_R(x, \infty)}{x^\eta} \right| \xrightarrow{\mathbb{P}} 0,
\]

\[
\sup_{y \in (0, T]} \left| \frac{\sqrt{k}(T_n(\infty, y) - y) - W_R(\infty, y)}{y^\eta} \right| \xrightarrow{\mathbb{P}} 0.
\]

The following lemma shows the boundedness of the $W_R$ process with proper weighing function. It follows from, for instance, a modification of Example 1.8 in Alexander (1986) or that of Lemma 3.2 in Einmahl et al. (2006).

**Lemma 4.5.2.** For any $T > 0$ and $\eta \in [0, 1/2)$, with probability 1,

\[
\sup_{0 < x \leq T, 0 < y < \infty} \frac{|W_R(x, y)|}{x^\eta} < \infty \quad \text{and} \quad \sup_{0 < x < \infty, 0 < y < T} \frac{|W_R(x, y)|}{y^\eta} < \infty.
\]

Next, denote $s_n(x) = \frac{n}{k} (1 - F_1(U_1(n/k)x^{-\gamma_1}))$ for $x > 0$. From the regular variation condition (4.2.2), we get that $s_n(x) \to x$ as $n \to \infty$. The following lemma shows that when handling proper integrals, $s_n(x)$ can be substituted by $x$ in the limit.

---

3It is called “pseudo” estimator because the marginal distribution functions are unknown.
Lemma 4.5.3. Suppose (4.2.2) holds. Denote $g$ as a bounded and continuous function on $[0, S_0] \times [a, b]$ with $0 < S_0 \leq \infty$ and $0 \leq a < b < \infty$. Moreover, suppose there exist $\eta_1 > \gamma_1$ and $m > 0$ such that

$$\sup_{0 < x \leq S_0, \ a \leq y \leq b} \frac{|g(x, y)|}{x^{\eta_1}} \leq m.$$ 

If $S_0 < +\infty$, we further require that $0 < S < S_0$. Then,

$$\lim_{n \to \infty} \sup_{a \leq y \leq b} \left| \int_0^S g(s_n(x), y) - g(x, y) dx^{-\gamma_1} \right| = 0. \quad (4.5.2)$$

Furthermore, suppose $|g(x_1, y) - g(x_2, y)| \leq |x_1 - x_2|$ holds for all $0 \leq x_1, x_2 < S_0$ and $a \leq y \leq b$. Under conditions (b) and (d), we have that

$$\lim_{n \to \infty} \sup_{a \leq y \leq b} \sqrt{\frac{\epsilon}{k}} \left| \int_0^S g(s_n(x), y) - g(x, y) dx^{-\gamma_1} \right| = 0. \quad (4.5.3)$$

**Proof of Lemma 4.5.3** We prove (4.5.2) and (4.5.3) for $S = S_0 = \infty$. The proof for $0 < S < S_0 < +\infty$ is similar but simpler. For any $0 < \epsilon < 1$, denote $T(\epsilon) = \epsilon^{-1/\gamma_1}$. It follows from (4.2.2) and Proposition B.1.10 of de Haan and Ferreira (2006) that

$$\lim_{n \to \infty} \sup_{0 < x \leq 1} \frac{s_n(x)}{x^{\eta_1}} = 1,$$

and

$$\lim_{n \to \infty} \sup_{0 < x \leq T(\epsilon)} |s_n(x) - x| = 0.$$

With $\delta(\epsilon) = \epsilon^{1/(\eta_1 - \gamma_1)}$, we have that

$$\sup_{a \leq y \leq b} \left| \int_0^\infty (g(s_n(x), y) - g(x, y)) dx^{-\gamma_1} \right| \leq \sup_{a \leq y \leq b} \left( \left| \int_0^\delta (g(s_n(x), y) - g(x, y)) dx^{-\gamma_1} \right| + \left| \int_\delta^T (g(s_n(x), y) - g(x, y)) dx^{-\gamma_1} \right| + \left| \int_T^\infty (g(s_n(x), y) - g(x, y)) dx^{-\gamma_1} \right| \right) \leq m \int_0^\delta \left( x^{\frac{\eta_1 + \gamma_1}{2}} + x^{\eta_1} \right) dx^{-\gamma_1} + \delta^{-\gamma_1} \sup_{\delta \leq x \leq T} |g(s_n(x), y) - g(x, y)| + 2 \epsilon \sup_{0 < x < \infty} |g(x, y)|$$
\[ \leq c_1 \varepsilon^{1/2} + \delta^{-\gamma_1} \sup_{\frac{\delta}{\varepsilon} \leq x \leq T} \sup_{0 \leq y \leq b} |g(s_n(x), y) - g(x, y)| + 2\varepsilon \sup_{0 \leq x \leq \infty} |g(x, y)|, \]

where \( c_1 \) is a finite constant. Hence, (4.5.2) follows from the uniform continuity of \( g \) on \([\delta, T] \times [a, b]\) and the boundedness of \( g \) on \([0, \infty) \times [a, b]\).

Next we prove (4.5.3). Denote \( \tilde{T}_n = |A_1(n/k)|^{\frac{1}{\gamma_1 - 7}} \). By the Lipschitz property of \( g \),

\[
\sup_{a \leq y \leq b} \left| \int_0^\infty (g(s_n(x), y) - g(x, y)) \, dx^{\gamma_1} \right| \leq \int_0^{\tilde{T}_n} |s_n(x) - x| \, dx^{\gamma_1} + 2 \sup_{0 \leq x < \infty} |g(x, y)| \tilde{T}_n^{-\gamma_1}. \tag{4.5.4}
\]

It is thus necessary to prove that both terms in the right hand side of (4.5.4) are \( o(1/\sqrt{k}) \). For the second term, condition (d) implies that \( \alpha \frac{\gamma_1}{\gamma_1 - 7} < \frac{\gamma_1}{\rho_1 - 1} \). Thus for any \( \varepsilon_0 \in \left(0, \frac{\gamma_1}{\rho_1 - 1} - \frac{\alpha}{2(1 - \alpha)}\right) \), as \( n \to \infty \), we have that

\[
\sqrt{k} \left( n \frac{\gamma_1}{\rho_1} + \varepsilon_0 \right) = O\left(n \frac{\gamma_1}{\rho_1} + \varepsilon_0\left( \frac{\gamma_1}{\rho_1} + \frac{\varepsilon_0}{2}\right)\right) \to 0,
\]

which leads to

\[
\sqrt{k} \tilde{T}_n^{-\gamma_1} = \sqrt{k} |A_1(n/k)|^{\frac{1}{\gamma_1 - 7}} \to 0. \tag{4.5.5}
\]

For the first term, notice that for \( x \in (0, \tilde{T}_n) \) and \( 0 < \varepsilon_1 < \frac{\gamma_1}{\rho_1 - 1} \), when \( n \) is large enough,

\[
U_1(n/k)x^{\gamma_1} \geq U_1(n/k)\tilde{T}_n^{\gamma_1} = U_1(n/k)|A_1(n/k)|^{\frac{\gamma_1}{\rho_1}} \geq \left( n \frac{\gamma_1}{\rho_1} \right)^{-\varepsilon_1},
\]

which implies that \( U_1(n/k)x^{\gamma_1} \to +\infty \) as \( n \to \infty \). Hence we can apply Theorems 2.3.9 and B.3.10 in de Haan and Ferreira (2006) to condition (b) and obtain that for sufficiently large \( n \),

\[
\left| \frac{s_n(x) - x}{A_1(n/k)} - x^{\frac{1}{\gamma_1 \rho_1}} \right| \leq x^{1 - \rho_1} \max(x^{\varepsilon_0}, x^{-\varepsilon_0}).
\]

Thus, we get that

\[
\sqrt{k} \int_0^{\tilde{T}_n} |s_n(x) - x| \, dx \to 0
\]

\[
\leq \sqrt{k} |A_1(n/k)| \int_0^{\tilde{T}_n} \left( x^{\frac{1}{\gamma_1 \rho_1}} - 1 \right) + x^{1 - \rho_1} \max(x^{\varepsilon_0}, x^{-\varepsilon_0}) \, dx^{\gamma_1}
\]

Chapter 4. Estimation of marginal expected shortfall

85
with \( c_2 \) and \( c_3 \) some positive constants. Again, by condition (d), as \( n \to \infty \), \( c_3 \sqrt{k} (n/k)^{\frac{\gamma_1 - \gamma_2}{1 - \gamma_2}} \to 0 \). Hence, (4.5.3) is proved by combining (4.5.4), (4.5.5) and (4.5.6).

With those auxiliary lemmas, we obtain the asymptotic behavior of \( \tilde{\theta}_{k\gamma_k} \) as follows.

**Proposition 4.5.1.** Suppose (4.2.1) and (4.2.2) hold with \( 0 < \gamma_1 < 1/2 \). Then,

\[
\sup_{1/2 \leq y \leq 2} \left| \frac{\sqrt{k}}{U_1(n/k)} \left( \tilde{\theta}_{k\gamma_k} - \theta_{k\gamma_k} \right) + \frac{1}{y} \int_0^\infty W_R(s, y)ds^{-\gamma_1} \right| \to 0.
\]

**Proof of Proposition 4.5.1** Recall \( s_n(x) = \frac{n}{k} (1 - F_1(U_1(n/k)x^{-\gamma_1})) \), \( x > 0 \). Similar to (4.5.1),

\[
y\theta_{k\gamma_k} = \int_0^\infty \frac{n}{k} P(X > s, Y > U_2(n/(ky)))ds
= \int_0^\infty \frac{n}{k} P(1 - F_1(X) < 1 - F_1(s), 1 - F_2(Y) < ky/n)ds
= \int_0^\infty R_n \left( \frac{n}{k} (1 - F_1(s)), y \right) ds
= - U_1(n/k) \int_0^\infty R_n(s_n(x), y)dx^{-\gamma_1}.
\] (4.5.7)

Similarly, \( y\tilde{\theta}_{k\gamma_k} = -U_1(n/k) \int_0^\infty T_n(s_n(x), y)dx^{-\gamma_1} \). For any \( T > 0 \), we have

\[
\sup_{1/2 \leq y \leq 2} \left| \frac{\sqrt{k}}{U_1(n/k)} (y\tilde{\theta}_{k\gamma_k} - y\theta_{k\gamma_k}) + \int_0^\infty \frac{1}{y} W_R(x, y)dx^{-\gamma_1} \right|
= \sup_{1/2 \leq y \leq 2} \left| \int_0^\infty W_R(x, y)dx^{-\gamma_1} - \int_0^\infty \sqrt{k} (T_n(s_n(x), y) - R_n(s_n(x), y))dx^{-\gamma_1} \right|
\leq \sup_{1/2 \leq y \leq 2} \left| \int_0^T W_R(x, y)dx^{-\gamma_1} \right| + \sup_{1/2 \leq y \leq 2} \left| \int_T^\infty \sqrt{k} (T_n(s_n(x), y) - R_n(s_n(x), y))dx^{-\gamma_1} \right|
\leq I_1(T) + I_{2,n}(T) + I_{3,n}(T).
\]
It suffices to prove that for any $\varepsilon > 0$, there exists $T_0 = T_0(\varepsilon)$ such that

$$P(I_1(T_0) > \varepsilon) < \varepsilon,$$  \hspace{1cm} (4.5.8)

and $n_0 = n_0(T_0)$ such that for any $n > n_0$,

$$P(I_{2,n}(T_0) > \varepsilon) < \varepsilon; \hspace{1cm} (4.5.9)$$

$$P(I_{3,n}(T_0) > \varepsilon) < \varepsilon. \hspace{1cm} (4.5.10)$$

Firstly, for the term $I_1(T)$, by Lemma 4.5.2 with $\eta = 0$, there exists $T_1 = T_1(\varepsilon)$ such that

$$P\left(\sup_{0 < x < \infty, 0 \leq y \leq 2} |W_R(x, y)| > T_1^{\gamma_1}\varepsilon\right) < \varepsilon.$$  \hspace{1cm} (4.5.11)

Then for any $T > T_1$,

$$P(I_1(T) > \varepsilon) \leq P\left(\sup_{x > T, 1/2 \leq y \leq 2} \sqrt{k}(T_n(s_n(x), y) - R_n(s_n(x), y)) dx \gamma_1 > \varepsilon T_1^{\gamma_1}\right).$$

Thus (4.5.8) holds provided that $T_0 > T_1$.

Next we deal with the term $I_{2,n}(T)$. Let $\tilde{P}$ be the probability measure defined by $(1 - F_1(X), 1 - F_2(Y))$ and $\tilde{P}_n$ the empirical probability measure defined by $(1 - F_1(X_i), 1 - F_2(Y_i))_{1 \leq i \leq n}$. We have

$$P(I_{2,n}(T) > \varepsilon) = P\left(\sup_{1/2 \leq y \leq 2} \left| \int_T^{\infty} \sqrt{k}(T_n(s_n(x), y) - R_n(s_n(x), y)) dx \gamma_1 \right| > \varepsilon T_1^{\gamma_1}\right)$$

$$\leq P\left(\sup_{x > T, 1/2 \leq y \leq 2} \sqrt{n}(\tilde{P}_n - \tilde{P}) \left\{ \left(0, \frac{k s_n(x)}{n}\right), \left(0, \frac{k y}{n}\right) \right\} > \varepsilon T_1^{\gamma_1} \sqrt{k/n}\right)$$

$$=: p_2.$$  \hspace{1cm} (4.5.12)

Define $S_n = \{[0, 1] \times (0, 2k/n)\}$, then $\tilde{P}(S_n) = 2k/n$. Now by Inequality 2.5 in Einmahl (1987), there exists a constant $c$ and a function $\psi$ with $\lim_{t \to 0} \psi(t) = 1$, such that

$$p_2 \leq c \exp\left(-\frac{(\varepsilon T_1^{\gamma_1} \sqrt{k/n})^2}{4P(S_n)} \psi\left(\frac{\varepsilon T_1^{\gamma_1} \sqrt{k/n}}{\sqrt{n}P(S_n)}\right)\right).$$  \hspace{1cm} (4.5.13)
\[ c \exp \left( -\frac{\varepsilon^2 T^{\gamma_1}}{8} \psi \left( \frac{\varepsilon T^{\gamma_1/2}}{2\sqrt{k}} \right) \right) \]

Choose \( T_2(\varepsilon) \) such that \( c \exp \left( -\frac{\varepsilon^2 T^{\gamma_1}}{16} \right) \leq \varepsilon. \) Then, for any \( T > T_2, c \exp \left( -\frac{\varepsilon^2 T^{\gamma_1}}{16} \right) \leq \varepsilon. \)

Moreover, we can choose \( n_1 = n_1(T) \) such that for \( n > n_1, \psi \left( \frac{\varepsilon T^{\gamma_1/2}}{2\sqrt{k}} \right) > 1/2. \)

Therefore, for \( T > T_2(\varepsilon) \) and \( n > n_1(T), \) we have \( p_2 < \varepsilon, \) which leads to (4.5.9) provided that \( T_0 > T_2 \) and \( n_0 > n_1. \)

Lastly, we deal with \( I_{3,n}(T). \) We have that

\[
P(I_{3,n}(T) > \varepsilon) = P \left( \sup_{1/2 \leq y \leq 2} \left| \int_0^T \sqrt{k} \left( T_n(s_n(x), y) - R_n(s_n(x), y) \right) - W_R(s_n(x), y) dx^{\gamma_1} \right| > \varepsilon/2 \right)
+ P \left( \sup_{1/2 \leq y \leq 2} \left| \int_0^T W_R(s_n(x), y) - W_R(x, y) dx^{\gamma_1} \right| > \varepsilon/2 \right)
= p_{31} + p_{32}.
\]

We first consider \( p_{31}. \) Notice that for any \( T, \) there exists \( n_2 = n_2(T) \) such that for all \( n > n_2, s_n(T) < T + 1. \) Hence, for \( n > n_2 \) and any \( \eta_0 \in (\gamma_1, 1/2), \)

\[
p_{31} \leq P \left( \sup_{0 < s < T + 1} \left| \frac{\sqrt{k}(T_n(s, y) - R_n(s, y)) - W_R(s, y)}{s^{\eta_0}} \right| \left| \int_0^T (s_n(x))^{\eta_0} dx^{\gamma_1} \right| > \varepsilon/2 \right)
\]

Notice that by (4.5.2), as \( n \to \infty, \left| \int_0^T (s_n(x))^{\eta_0} dx^{\gamma_1} \right| \to \frac{\gamma_1}{\eta_0 - \gamma_1} T^{\eta_0 - \gamma_1}. \) Together with Lemma 4.5.1, there exists \( n_3(T) > n_2(T) \) such that for \( n > n_3(T), \) \( p_{31} < \varepsilon/2. \)

Then, we consider \( p_{32}. \) Applying Lemma 4.5.2, with the aforementioned \( \eta_0 \in (\gamma_1, 1/2), \) there exists \( \lambda_0 = \lambda(\eta_0, \varepsilon) \) such that

\[
P \left( \sup_{0 < x < \infty, 1/2 \leq y \leq 2} \frac{|W_R(x, y)|}{x^{\eta_0}} \geq \lambda_0 \right) \leq \varepsilon/3. \quad (4.5.11)
\]

Moreover, \( W_R(x, y) \) is continuous on \((0, \infty) \times [1/2, 2], \) see Corollary 1.11 in Adler (1990). Hence applying (4.5.11) and (4.5.2) with \( g = W_R, S = T \) and \( S_0 = T + 1, \) we have that there exists a \( n_4 = n_4(T) \) such that for \( n > n_4, p_{32} < \varepsilon/2. \) Thus, (4.5.10) holds for any \( T_0 \) and \( n_0 > \max(n_3(T_0), n_4(T_0)). \)
To summarize, choose $T_0 = T_0(\varepsilon) > \max(T_1, T_2)$, and define $n_0(T_0) = \max_{1 \leq j \leq 4} n_j(T_0)$. We get that for the chosen $T_0$ and any $n > n_0$, the three inequalities (4.5.8)–(4.5.10) hold, which completes the proof of the proposition. □

Next, we proceed with the second step: establishing the asymptotic normality of $\hat{\theta}_k$. 

**Proposition 4.5.2.** Under the condition of Theorem 4.2.1, we have

$$\sqrt{k} \left( \frac{\hat{\theta}_k}{\theta_k} - 1 \right) \overset{d}{\rightarrow} \Theta.$$

**Proof of Proposition 4.5.2** Observe that $\lim_{n \to \infty} \frac{\theta_k}{U_1(n/k)} \overset{p}{\rightarrow} \Theta \int_0^\infty R(s^{-1/\gamma_1}, 1) ds$. Therefore it is sufficient to show that

$$\frac{\sqrt{k}}{U_1(n/k)} \left( \frac{\hat{\theta}_k}{\theta_k} - 1 \right) \overset{d}{\rightarrow} \Theta \int_0^\infty R(s^{-1/\gamma_1}, 1) ds.$$

Recall $e_n = \frac{n}{k} (1 - F_2(Y_{n-k,n}))$. Hence, with probability 1, $\frac{\hat{\theta}_k}{\theta_k} = e_n \tilde{\theta}_{k,n}$, we thus have that

$$\sqrt{k} \left( \frac{e_n \tilde{\theta}_{k,n} - \theta_k}{\theta_k} \right) = \left( \frac{\sqrt{k}}{U_1(n/k)} (e_n \tilde{\theta}_{k,n} - e_n \tilde{\theta}_{k,n}) + \int_0^\infty W_R(s, 1) ds^{-\gamma_1} \right) + \left( \frac{\sqrt{k}}{U_1(n/k)} (e_n \tilde{\theta}_{k,n} - \theta_k) - W_R(\infty, 1)(\gamma_1 - 1) \int_0^\infty R(s^{-1/\gamma_1}, 1) ds \right) =: J_1 + J_2.$$

We prove that both $J_1$ and $J_2$ converge to zero in probability as $n \to \infty$.

Firstly, we deal with $J_1$. By Lemma 4.5.1 and $T_n(\infty, e_n) = 1$, we get that

$$\sqrt{k} (e_n - 1) \overset{p}{\rightarrow} -W_R(\infty, 1), \quad (4.5.12)$$

which implies that

$$\lim_{n \to \infty} P(|e_n - 1| > k^{-1/4}) = 0.$$
Chapter 4. Estimation of marginal expected shortfall

Hence, with probability tending to 1,

\[ |J_1| \leq \sup_{|y-1|<k^{-1/4}} \left| \frac{\sqrt{k}}{U_1(n/k)} \left( y \hat{\theta}_1 - y \hat{\theta}_k \right) + \int_0^\infty W_R(s,y)ds^{-\gamma_1} \right| + \sup_{|y-1|<k^{-1/4}} \left| \int_0^\infty W_R(s,y) - W_R(s,1)ds^{-\gamma_1} \right|. \]

The first part converges to zero in probability by Proposition 4.5.1. For the second part, notice that for any \( \varepsilon > 0, 0 < \delta < 1 \) and \( \eta \in (\gamma_1, 1/2) \),

\[
P \left( \sup_{|y-1|<k^{-1/4}} \left| \int_0^\infty W_R(s,y) - W_R(s,1)ds^{-\gamma_1} \right| > \varepsilon \right) \leq P \left( \sup_{|y-1|<k^{-1/4}} \left| \int_0^\delta W_R(s,y) - W_R(s,1)ds^{-\gamma_1} \right| > \varepsilon/2 \right) + P \left( \sup_{|y-1|<k^{-1/4}} \left| \int_\delta^\infty W_R(s,y) - W_R(s,1)ds^{-\gamma_1} \right| > \varepsilon/2 \right)
\]

\[
\leq P \left( \sup_{0<s\leq1.1/2\leq y \leq 2} \frac{|W_R(s,y)|}{s^\eta} > \frac{\varepsilon(\eta - \gamma_1)}{4\gamma_1} \eta^{-1-\eta} \right) + P \left( \sup_{s>0,|y-1|<k^{-1/4}} |W_R(s,y) - W_R(s,1)| \delta^{-\gamma_1} > \varepsilon/2 \right) \]

\[=: p_{11} + p_{12}. \]

For any fixed \( \varepsilon \), Lemma 4.5.2 ensures that there exists a positive \( \delta(\varepsilon) \) such that for all \( \delta < \delta(\varepsilon) \), we have that \( p_{11} < \varepsilon \). Then, for any fixed \( \delta \), there must exist an positive integer \( n(\delta) \) such that for \( n > n(\delta) \) we can achieve that \( p_{12} < \varepsilon \), because we have that as \( n \to \infty \),

\[ \sup_{s>0,|y-1|<k^{-1/4}} |W_R(s,y) - W_R(s,1)| \xrightarrow{a.s.} 0, \]

see Corollary 1.11 in Adler (1990). Hence we proved that \( J_1 \xrightarrow{P} 0 \) as \( n \to \infty \).

Next we deal with \( J_2 \). We first prove a non-stochastic limit relation: as \( n \to \infty \),

\[
\sup_{1/2 \leq y \leq 2} \sqrt{k} \left| \int_0^\infty R_n(s_n(x),y) - R(x,y)dx^{-\gamma_1} \right| \to 0. \quad (4.5.13)
\]
Chapter 4. *Estimation of marginal expected shortfall* 91

Condition (a) implies that as \( n \to \infty \),

\[
\sup_{0 < x < \infty} \sup_{1/2 \leq y \leq 2} \frac{|R_n(x, y) - R(x, y)|}{x^\beta \wedge 1} = O \left( \left( \frac{n}{k} \right)^\tau \right).
\]

Hence, as \( n \to \infty \),

\[
\sup_{1/2 \leq y \leq 2} \sqrt{k} \left| \int_0^\infty R_n(s_n(x), y) - R(s_n(x), y)dx^{-\gamma_1} \right|
\leq \sqrt{k} \sup_{0 < x < \infty} \frac{|R_n(x, y) - R(x, y)|}{x^\beta \wedge 1} \left| \int_0^\infty (s_n(x))^{\beta} \wedge 1 dx^{-\gamma_1} \right|
= O \left( \sqrt{k} \left( \frac{n}{k} \right)^\tau \right) \left( - \int_0^{1/2} (s_n(x))^{\beta} dx^{-\gamma_1} - \int_0^\infty 1 dx^{-\gamma_1} \right) \to 0.
\]

The last step follows from the following two facts. Firstly, condition (d) ensures that \( k = O(n^\alpha) \) with \( \alpha < \frac{2\tau}{2\tau - 1} \). Secondly, we have that

\[
\lim_{n \to \infty} - \int_0^{1/2} (s_n(x))^{\beta} dx^{-\gamma_1} = - \int_0^{1/2} x^{\beta} dx^{-\gamma_1} < \infty,
\]

which is a consequence of (4.5.2).

To complete the proof of relation (4.5.13), it is still necessary to show that as \( n \to \infty \),

\[
\sup_{1/2 \leq y \leq 2} \sqrt{k} \left| \int_0^\infty R(s_n(x), y) - R(x, y)dx^{-\gamma_1} \right| \to 0.
\]

This is achieved by applying (4.5.3) to the \( R \) function which satisfies the Lipschitz condition: \( |R(x_1, y) - R(x_2, y)| \leq |x_1 - x_2| \), for \( x_1, x_2, y \geq 0 \). Hence, we proved the relation (4.5.13).

Combining (4.5.7) and (4.5.13), we obtain that

\[
\frac{\theta_k}{U_1(n/k)} = - \int_0^\infty R(s_n(x), 1)dx^{-\gamma_1} = - \int_0^\infty R(x, 1)dx^{-\gamma_1} + o \left( \frac{1}{\sqrt{k}} \right), \tag{4.5.14}
\]

and

\[
\frac{e_n \theta_k}{U_1(n/k)} = - \int_0^\infty R_n(s_n(x), e_n)dx^{-\gamma_1} = - \int_0^\infty R(x, e_n)dx^{-\gamma_1} + o_P \left( \frac{1}{\sqrt{k}} \right).
\]
Chapter 4. *Estimation of marginal expected shortfall*

From the homogeneity of the $R$ function, for $y > 0$, we have that

$$
\int_{0}^{\infty} R(x, y) dx^{-\gamma_1} = y^{1-\gamma_1} \int_{0}^{\infty} R(x, 1) dx^{-\gamma_1}.
$$

Hence, we get that

$$
e_n \theta_{\frac{\gamma_1}{n}} = e_n^{1-\gamma_1} \theta_{\frac{1}{n}} + o_p \left( \frac{U_1(n/k)}{\sqrt{k}} \right).
$$

By applying (4.5.12), Proposition 4.2.1 and the Cramér’s delta method, we get that, as $n \to \infty$, 

$$
\frac{\sqrt{k}}{U_1(n/k)} \left( e_n \theta_{\frac{\gamma_1}{n}} - \theta_{\frac{1}{n}} \right) \overset{P}{\to} \left( e_n^{1-\gamma_1} - 1 \right) \frac{\theta_{\frac{1}{n}}}{U_1(n/k)} + o_p(1) 
$$

which implies that to $J_2 \overset{P}{\to} 0$. The proposition is thus proved. \(\Box\)

Finally, we can combine the asymptotic relations on $\hat{\theta}_n$ and $\hat{\gamma}_1$ to obtain the proof of Theorem 4.2.1.

**Proof of Theorem 4.2.1** Write

$$
\frac{\hat{\theta}_p}{\theta_p} = \frac{d_{\hat{\gamma}_1}}{d_{\gamma_1}} \times \frac{\hat{\theta}_n}{\theta_n} \times \frac{d_{\hat{\gamma}_1 \theta_n}}{\theta_p} =: L_1 \times L_2 \times L_3.
$$

We deal with the three factors separately.

Firstly, handling $L_1$ uses the asymptotic normality of the Hill estimator. Under conditions (b) and (c), we have that, as $n \to \infty$,

$$
\sqrt{k_1} (\hat{\gamma}_1 - \gamma_1) \overset{P}{\to} \Gamma; \quad (4.5.15)
$$

see, e.g., Example 5.1.5 in de Haan and Ferreira (2006). As in the proof of Theorem 4.3.8 of de Haan and Ferreira (2006), this leads to

$$
\frac{\sqrt{k_1}}{\log d_n} (L_1 - 1) - \Gamma \overset{P}{\to} 0. \quad (4.5.16)
$$

Secondly, the asymptotic behavior of the factor $L_2$ is given by Proposition 4.5.2.
Lastly, for $L_3$, by condition (b) and Theorem 2.3.9 in de Haan and Ferreira (2006), we have that

$$\frac{U_1(1/p)}{U_1(n/k)d_{n1}} - \frac{1}{A_1(n/k)} \rightarrow -\frac{1}{\rho_1}.$$  

Together with the fact that as $n \rightarrow \infty$, $\sqrt{k}|A_1(n/k)| \rightarrow 0$ (implied by condition (d)), we get that

$$\frac{U_1(1/p)}{U_1(n/k)d_{n1}} - \frac{1}{A_1(n/k)} = o \left( \frac{1}{\sqrt{k}} \right)$$  

(4.5.17)

Following the same reasoning of (4.5.14) for $p \leq k/n$, we have

$$\frac{\theta_p}{\theta_p} = \frac{1}{\sqrt{k}n} \int_0^\infty R(s^{-1/\gamma}, 1) ds = o \left( \frac{1}{\sqrt{k}} \right).$$

Combining this with (4.5.17), we have

$$L_3 = \frac{\theta_p}{\theta_p} \frac{U_1(n/k)}{U_1(1/p)} \times \frac{U_1(n/k)d_{n1}}{U_1(1/p)} = 1 + o \left( \frac{1}{\sqrt{k}} \right).$$  

(4.5.18)

Combining the asymptotic relations (4.5.16), (4.5.18) and Proposition 4.5.2, we get that

$$\hat{\theta}_p / \theta_p - 1 = L_1 \times L_2 \times L_3 - 1$$

$$= \left( 1 + \frac{\log d_n}{\sqrt{k_1}} + o_P \left( \frac{\log d_n}{\sqrt{k_1}} \right) \right) \left( 1 + \frac{\Theta}{\sqrt{k}} + o_P \left( \frac{1}{\sqrt{k}} \right) \right) \left( 1 + o \left( \frac{1}{\sqrt{k}} \right) \right) - 1$$

$$= \frac{\log d_n}{\sqrt{k_1}} + \frac{\Theta}{\sqrt{k}} + o_P \left( \frac{\log d_n}{\sqrt{k_1}} \right) + o_P \left( \frac{1}{\sqrt{k}} \right).$$

The covariance matrix of $(\Theta, \Gamma)$ follows from the straightforward calculation. □

**Proof of Theorem 4.2.2** Write $\theta_p^+ := E(X^+ | Y > U_2(1/p))$. Then,

$$\frac{\hat{\theta}_p}{\theta_p} = \frac{\hat{\theta}_p}{\theta_p} \times \frac{\theta_p^+}{\theta_p}.$$

Hence, it suffices to prove that $\frac{\hat{\theta}_p}{\theta_p}$ follows the asymptotic normality stated in Theorem 4.2.1 and $\frac{\theta_p^+}{\theta_p} = 1 + o \left( \frac{1}{\sqrt{k}} \right)$.

We first show that $(X^+, Y)$ satisfies conditions (a) and (b) of Section 4.2.1. Let $\tilde{F}_1$ be the distribution function of $X^+$ and $\bar{U}_1 = \left( \frac{1}{1 - \tilde{F}_1} \right)^{-1}$. It is obvious that $U_1(t) = \bar{U}_1(t)$, for $t > \frac{k}{1 - \tilde{F}_1(t)}$. Hence $X^+$ satisfies condition (b).
Before checking condition (a) for \((X^+, Y)\), we prove that, as \(t \to \infty\),

\[
\mathbb{P}(X < 0, 1 - F_2(Y) < 1/t) = O(t^\tau). \tag{4.5.19}
\]

Observe that condition (a) implies that

\[
\sup_{1/2 \leq y \leq 2} \left| y - R(t, y) \right| = O(t^\tau). \tag{4.5.20}
\]

Because of the homogeneity of \(R\), we have \(1 - R(ct, 1) = O(t^\tau)\) for any \(c \in (0, \infty)\). Hence, (4.5.19) is proved by

\[
\begin{align*}
\mathbb{P}(X < 0, 1 - F_2(Y) < 1/t) &= 1 - \mathbb{P}(X > 0, 1 - F_2(Y) < 1/t) \\
&= 1 - \mathbb{P}(1 - F_1(X) < 1 - F_1(0), 1 - F_2(Y) < 1/t) \\
&\leq 1 - R(t(1 - F_1(0)), 1) + |\mathbb{P}(1 - F_1(X) < 1 - F_1(0), 1 - F_2(Y) < 1/t) - R(t(1 - F_1(0)), 1)| \\
&= O(t^\tau).
\end{align*}
\]

Now we show that \((X^+, Y)\) satisfies condition (a), that is, as \(t \to \infty\),

\[
\sup_{0 < x \leq \infty, 1/2 \leq y \leq 2} \left| \frac{\mathbb{P}(1 - F_1(X) < x/t, 1 - F_2(Y) < y/t) - R(x, y)}{x^{\beta} \wedge 1} \right| = O(t^\tau). \tag{4.5.21}
\]

Firstly, observe that for \(0 < x \leq t(1 - F_1(0))\),

\[
\{1 - \tilde{F}_1(X^+) < x/t\} = \{1 - F_1(X^+) < x/t\} = \{1 - F_1(X) < x/t\}.
\]

Hence, the uniform convergence (in (4.5.21) ) on \((0, t(1 - F_1(0))) \times [1/2, 2]\) follows from the fact that \((X, Y)\) satisfies condition (a). Secondly, for \(x > t(1 - F_1(0))\), we have \(1 - \tilde{F}_1(X^+) < x/t\). Therefore,

\[
\begin{align*}
\sup_{t(1 - F_1(0)) < x < \infty, 1/2 \leq y \leq 2} \left| \mathbb{P}(1 - \tilde{F}_1(X^+) < x/t, 1 - F_2(Y) < y/t) - R(x, y) \right| \\
&= \sup_{t(1 - F_1(0)) < x < \infty} (y - R(x, y)) \\
&\leq \sup_{1/2 \leq y \leq 2} (y - R(t(1 - F_1(0)), y)) = O(t^\tau),
\end{align*}
\]

Hence, (4.5.21) is proved.
where the last relation follows from (4.5.20). This completes the verification of (4.5.21).

As a result, Theorem 4.2.1 applies to \( \hat{\theta}_p + \theta_p \).

Next we show that \( \theta_p = 1 + O \left( \frac{1}{\sqrt{k}} \right) \). By Proposition 4.2.1, \( \frac{\theta_p}{U_1(1/p)} \to \int_0^\infty R \left( x^{-1/\gamma_1}, 1 \right) dx \).

By Hölder’s inequality, condition (e) and (4.5.19),

\[
- E(X^- | Y > U_2(1/p)) = - \frac{1}{p} E(X^- I(X < 0, Y > U_2(1/p))) \\
\leq \frac{1}{p} \left( E |X^-|^{1/\gamma_1} \right)^{\gamma_1} \left( P(X < 0, Y > U_2(1/p)) \right)^{1-\gamma_1} \\
= O(p^{-1+(1-\tau)(1-\gamma_1)}).
\]

Condition (b) can be written as:

\[
\lim_{t \to \infty} \frac{U_1(tx)(tx)^{-\gamma_1} - U_1(t)t^{-\gamma_1}}{A_1(t)U_1(t)t^{-\gamma_1}} = \frac{x^{\rho_1} - 1}{\rho_1}.
\]

It follows from Theorem B.2.2 in de Haan and Ferreira (2006) that \( \frac{1}{U_1(1/p)} = O(p^{\gamma_1}) \), as \( p \downarrow 0 \). Hence by condition (f),

\[
\frac{\theta_p}{\theta_p^+} = 1 + \frac{E(X^- | Y > U_2(1/p))}{\theta_p^+} = 1 + O \left( \frac{p^{-1+(1-\tau)(1-\gamma_1)}}{U_1(1/p)} \right) \\
= 1 + O \left( p^{-\tau(1-\gamma_1)} \right) = 1 + o \left( \frac{1}{\sqrt{k}} \right).
\]

\( \square \)
Bibliography


