Statistics of Extremes: Estimation and Optimality

(Extremenstatistiek: Schatten en Optimaliteit)

PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Tilburg,

op gezag van de rector magnificus,

prof. dr. F.A. van der Duyn Schouten,

in het openbaar te verdedigen ten overstaan van

een door het college voor promoties aangewezen commissie

in de aula van de Universiteit op

maandag 25 november 2002, om 16.15 uur

door

Ana Maria Santos Ferreira Gorjão Henriques,

geboren op 21 september 1969 te Lissabon, Portugal.

Promotores: Prof. dr L. de Haan
Prof. dr J.H.J. Einmahl

**European Institute for
Statistics, Probability, Stochastic Operations Research and their
Applications**
Seat: Eindhoven University of Technology
EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

The research for this thesis was done while the author was working at EURANDOM.
The author was supported by the Thomas Stieltjes Institute for Mathematics.

THOMAS STIELTJES INSTITUTE FOR MATHEMATICS

Acknowledgements

First of all I would like to express my gratitude to my supervisor Laurens de Haan, for his encouragement, guidance and how he believed in me, even when I came to him with the most fundamental questions. I find an immense privilege to have had the opportunity of sharing his knowledge and wisdom.

To everybody at EURANDOM (including those who left by now) I would like to mention the great environment I experienced. The stimulating atmosphere with seminars, workshops and all that research activity, and so many interesting people visiting EURANDOM. The great support that everybody from the office management staff gives us, especially foreigners and non-Dutch speakers. Moreover our extra-work activities, like korfbal tournaments. I thank you all.

I would like to thank my co-authors Casper, Gerrit, Holger and Peng. Sometimes it was just fun, when work went well, but there were also difficult periods. In the end I find great that we have shared our knowledge.

I would like to thank Ivette Gomes, John Einmahl and Claudia Klüppelberg for their support.

I would like to thank my colleagues at ISA, who have agreed with my leave-period, even if it was asked sooner than expected.

To family and all friends who supported and helped me, when I was in Eindhoven asking things from Lisbon, or when I was in Munich asking help from one of these other places... Many thanks.

Finally my deepest gratitude goes to Bernardo who, most of the time from distance but always present, supported and encouraged me through all this period.

Contents

Chapter 1

Introduction

This thesis is devoted to statistics in extreme value theory, where the estimation of quantities related to extreme events is of particular interest. For example in the design of dikes, a typical requirement is that the sea wall must be high enough so that the chance of a flood is no more than once in ten thousand years. We consider such a flood an extreme event. Or, in insurance mathematics it is of great interest to have statistical insight into the occurrence of large claims, due to natural catastrophes such as floods, hurricanes, high temperatures that give potential risk of fires of enormous proportions, etc. A common feature of this kind of events is that not many (if any at all) events of similar size have been observed in the past. Hence when making inferences related to extreme events, in particular one faces the problem of estimation where information from previous 'experiments' is scarce.

Let us illustrate our problem with an example. In Figure 1.1.a, in the real line are indicated 1877 observations from the sea level (in cm) at Delfzijl, which is located in the north coast of The Netherlands, measured during winterstorms in the years 1882-1991. The storm season lasts from October 1 until March 15. For more details on the data set see Dillingh et al. (1993). Now suppose that we are interested in estimating those sea levels (during winterstorms) that have probability .05 or .0001 of being exceeded. To start organising the information contained in the sample, we construct the empirical distribution function F_n, i.e. we put mass $1/n$ at every one of the observations (where n represents the sample size throughout). This is shown in Figure 1.1.b. Then, from this distribution the desired levels could be obtained by making the correspondence between the given probability and the level, as shown in Figure 1.1.b. But, it becomes clear that extra information is needed as the sea level to be estimated becomes larger, with the most extreme cases being when the given probability is smaller than $1/n$. Figures 1.2.a display the empirical distribution on a log-scale, i.e. the step function $-\log(1-F_n)$, which is often an appropriate scale when one is mainly interested in the larger values of a sample.

Under rather general conditions, extreme value theory provides a class of functions to fit to the distribution of the largest observations. Figure 1.2.b shows some

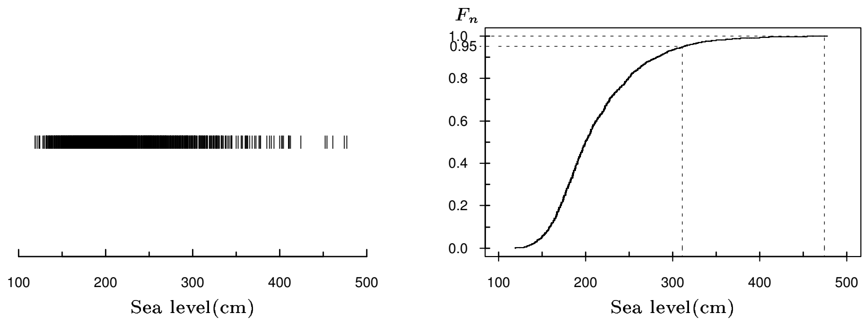


Figure 1.1: a) Left: sea level (cm) at Delfzijl (the data was trend corrected). b) Right: empirical distribution of the sea level sample.

of these functions. A real parameter, γ say, comes into play which determines their shape. To fit the appropriate function to the tail of the distribution, one has then to decide on the shape, and moreover on the appropriate shift and scaling constants; for instance in case of deciding for a straight line ($\gamma = 0$) then one has to decide on the appropriate slope and origin of the line to fit to the tail.

Estimators of the shape parameter, and the normalising constants are known from e.g. Hill (1975), Dekkers et al. (1989), and many others. Their accuracy depends strongly on the size of the sample fraction actually used in the estimation. On the one hand one should only take the larger observations of the sample. Since in general the extreme value model is valid only in the limit, the consequent approximation from the extreme value model should only be applied in the range of the largest obervations. On the other hand one wants to extract the most information from the sample, and so to have as many observations as possible. In Figures 1.2.a are examples of various models that can be fitted to the tail of the distribution. They only differ in the number of largest observations that were used in the estimation. Then, from each model the sea level estimate is obtained by making the correspondence between the level and the given probability, as shown in these figures for the case .0001.

Therefore a key question in the estimation procedure is: *How to choose the sample fraction to use in the estimation?* Recently there have been proposals to obtain the optimal sample fraction in the estimation of γ. In part of this thesis we address this question but from a quantile perspective, where throughout a quantile is an unkown quantity (like the sea level we wanted to estimate above) which has some given probability of being exceeded.

The reverse problem of estimating tail (or exceedance) probabilities is similar. Then one can ask, given some high value: *What is the optimal sample fraction when estimating the probability of observing a value larger than the given value?*

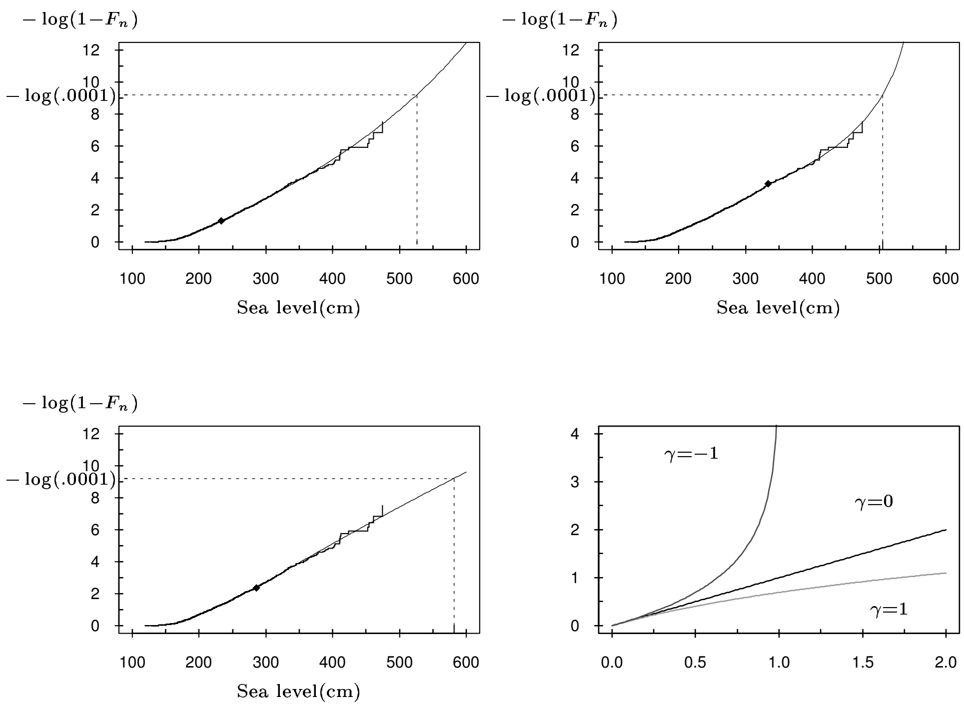


Figure 1.2: a) Top and bottom left: step function $-\log(1-F_n)$ of the sea level sample with estimated models. b) Bottom right: theoretical models.

Consider now the bivariate setting. For example, in the design of the dikes it is natural to consider not only the sea level but also the wave height, and of major importance might be the interaction between these two variables for large values of both. As before we are interested in inferences concerning large values of both variables. Then questions of the following type are of interest: *What is the probability that some given high values of the sea level and wave height are both exceeded?*

Statistical models in multivariate extreme value theory usually consider separately the marginal structure and the dependence structure. For the first one then one is back to the univariate setting. For the dependence structure problems arise with the multivariate theory (e.g. from Resnick, 1987), when extremes of the marginal variables are independent. For instance this holds for all bivariate normal variables with correlation less than 1. We shall study a model for the dependence part which comprises a parameter governing the asymptotic dependence of the variables.

In the bivariate setting not so much is known about estimation in general. Hence in this thesis, while in the univariate setting we focus on estimation and optimality issues, in the bivariate setting we shall address the estimation problem and leave out the optimality questions.

1.1 Extreme value theory

In the following two subsections we formalise a little some of the ideas discussed above.

1.1.1 Univariate setting

Suppose that one has a sample $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables from some unknown distribution function F. The basic condition for F in extreme value theory is, that F belongs to the domain of attraction of the generalised extreme value distribution, for some extreme value index $\gamma \in \mathbb{R}$ (Fisher and Tippett, 1928; Gnedenko, 1943). In other words, suppose there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$, such that the normalised sample maximum converges in distribution

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} P\left(\frac{\max\{X_1, X_2, \ldots, X_n\} - b_n}{a_n} \le x\right) = G(x), \quad (1.1.1)$$

for all $x \in \mathbb{R}$ with G a non-degenerate function. Then G is known to be an extreme value distribution, of the type

$$G_\gamma(x) = \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}, \quad 1+\gamma x > 0,\ \gamma \in \mathbb{R}$$

(read e^{-x} for $(1+\gamma x)^{-1/\gamma}$ in case $\gamma = 0$).

Condition (1.1.1) in terms of tail probabilities gives that,

$$\lim_{t\to\infty} t\left\{1 - F\left(b(t) + xa(t)\right)\right\} = 1 - H_\gamma(x) \quad (1.1.2)$$

for all x for which $0 < H_\gamma(x) < 1$, $a(t) > 0$ and $b(t) \in \mathbb{R}$ are suitable normalising functions and $H_\gamma(x)$ is the generalised Pareto distribution:

$$H_\gamma(x) = 1 - (1+\gamma x)^{-1/\gamma}, \quad 1+\gamma x > 0, \gamma \in \mathbb{R}.$$

Formula (1.1.2) suggests, for large x,

$$1 - F(x) \approx \frac{1}{t}\left\{1 + \gamma\frac{x - b(t)}{a(t)}\right\}^{-1/\gamma}, \quad \text{as} \quad t \to \infty.$$

This motivates the following estimator for the exceedance probability (e.g. Dekkers et al., 1989),

$$\hat{p}_n(k) = \frac{k}{n} \max\left\{0, \left(1 + \hat{\gamma}_n(k) \frac{x_n - \hat{b}(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\right)\right\}^{-1/\hat{\gamma}_n(k)}, \tag{1.1.3}$$

where n is the sample size, k is an intermediate sequence ($k = k(n) \to \infty$ and $k(n)/n \to 0$, as $n \to \infty$), and $\hat{a}$, $\hat{b}$ and $\hat{\gamma}_n$ are estimators of a, b and γ, respectively. In the sequel we use moment type estimators (Hill, 1975; Dekkers et al., 1989) to estimate γ. The following quantile estimator can be similarly motivated,

$$\hat{x}_n(k) = \hat{b}(\frac{n}{k}) + \hat{a}(\frac{n}{k}) \frac{\left(\frac{k}{np_n}\right)^{\hat{\gamma}_n(k)} - 1}{\hat{\gamma}_n(k)}. \tag{1.1.4}$$

These estimators can be somewhat simplified, if the underlying distribution F is restricted to the family of distributions which are in the domain of attraction of G_γ with $\gamma > 0$.

Let x^* be the upper endpoint of F, i.e. $x^* = \sup\{x : F(x) < 1\}$, which can be a finite number or infinity. It is known that if F verifies (1.1.1) with $\gamma < 0$ then x^* is finite, and in this case one can obtain the following estimator,

$$\hat{x}^*_n(k) = \hat{b}(\frac{n}{k}) - \frac{\hat{a}(\frac{n}{k})}{\hat{\gamma}_n(k)}. \tag{1.1.5}$$

In endpoint estimation, in addition to using moment type estimators to estimate γ, we shall also consider a shift and scale invariant estimator of γ.

In this thesis under quite general conditions we obtain limiting distributions for all these estimators. This is a major step in the optimal sample fraction analysis. A common feature in all these estimators is: their variance is large if one takes in the estimation a small number of upper order statistics, a bias component is introduced when the number of upper order statistics included is large. A criterion to obtain the optimal number of upper order statistics to use in the estimation, $k_0(n)$ say, is to minimise in a special way, over k, the mean square error of the limiting random variable. Then typically one finds

$$k_0(n) \sim \text{ const } n^{-2\rho/(1-2\rho)}, \quad n \to \infty \tag{1.1.6}$$

where $\rho < 0$ is a second order regular variation parameter, depending on the underlying distribution function F. The constant in (1.1.6) depends on the variance and bias of the limiting random variable. An interesting feature is that this asymptotic optimal rate does not depend on the given high value when estimating a tail probability, or on the given probability when estimating a quantile.

1.1.2 Bivariate setting

Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a sequence of independent and identically distributed random vectors with distribution function $\mathbf{F}$, with marginal distributions

F_1, F_2. We are interested in probabilities of the type

$$P(X_i > u \text{ and } Y_i > v),$$

where u and v are large threshold values. Since only large values of X_i and Y_i are involved, one would expect multivariate extreme value theory to provide the appropriate framework for systematic estimation of the above probability. Namely, the assumption that there exist normalising constants $a_n, c_n > 0$ and $b_n, d_n \in \mathbb{R}$ such that

$$\begin{aligned}
&\lim_{n\to\infty} \mathbf{F}^n(a_n x + b_n, c_n y + d_n) \\
&= \lim_{n\to\infty} P\left(\frac{\max\{X_1,\ldots,X_n\} - b_n}{a_n} \le x, \frac{\max\{Y_1,\ldots,Y_n\} - d_n}{c_n} \le y\right) \\
&= G(x,y) \qquad (1.1.7)
\end{aligned}$$

for all but denumerable many vectors (x, y) . Here G is a distribution function with non-degenerate marginals (Resnick, 1987).

We say that the maxima of the X_i and those of the Y_i are asymptotically independent if the marginals of the limiting distribution are independent. Unfortunately, in this case the limit assumption (1.1.7) is of little help to estimate the above probability. Indeed, this is a rather common situation. For instance, it holds for nondegenerate bivariate normal distributions.

In order to overcome this problem, Ledford and Tawn (1996) introduced a submodel, where the penultimate tail dependence is characterised by a coefficient $\eta \in (0, 1]$. More precisely, they assumed that the function $t \mapsto P(1-F_1(X) < t \text{ and } 1-F_2(Y) < t)$ is regularly varying at 0 with index $1/\eta$. Then $\eta = 1$ in case of asymptotic dependence, whereas $\eta < 1$ implies asymptotic independence.

As an extension to Ledford and Tawn's sub-model, we introduce the following assumption:

$$\lim_{t\downarrow 0} \frac{\frac{P\{1-F_1(X)<tx \text{ and } 1-F_2(Y)<ty\}}{q(t)} - c(x,y)}{q_1(t)} = c_1(x,y) \qquad (1.1.8)$$

exists, for $x, y \ge 0$ (but $x + y > 0$), with q positive, $q_1 \to 0$ as $t \downarrow 0$ and c_1 nonconstant and not a multiple of c. Moreover, we assume that the convergence is uniform on $\{(x,y) \in [0,\infty)^2 \mid x^2 + y^2 = 1\}$. Then, in particular it follows that the function q is regularly varying at zero with index $1/\eta$, $\eta \in (0, 1]$, and without loss of generality we may assume that $q(t) = \Pr\{1 - F_1(X) < t \text{ and } 1 - F_2(Y) < t\}$. Our assumptions imply that (1.1.8) holds locally uniformly on $(0,\infty)^2$. The bivariate normal distribution satisfies these conditions.

Under (1.1.8) we propose a new estimator for the dependence parameter η and prove its asymptotic normality. Moreover we propose a procedure to estimate the probability of an extreme set (like the one mentioned above) that works under asymptotic dependence as well as under asymptotic independence.

1.2 Outline of the thesis

This thesis is a collection of five articles. Each of them has its own introduction, where the relation of the subjects with the literature is discussed. Thus in the following we just focus on the key ideas of the contents of each chapter.

In Chapter 2 (paper Ferreira et al., 1999) we consider high quantiles and endpoint estimation. We establish distributional limiting results for these estimators. Then we obtain an (asymptotically) optimal number, $k_0(n)$ say, of upper order statistics to use in the estimation (as briefly explained in (1.1.6)).

In this chapter we also deal with the estimation of $k_0(n)$. We propose an adaptive bootstrap procedure, from which we obtain a consistent estimator of $k_0(n)$ in the sense that, $\hat{k}_0(n)/k_0(n) \to 1$ $(n \to \infty)$, in probability. In order to study the behaviour of this estimator for finite samples, we also present simulation results.

Also in Chapter 2 we apply our results, to the estimation of the endpoint of the distribution of the total life span of a certain group of individuals.

Chapter 3 (paper Ferreira, 2002) deals with tail (or exceedance) probability estimation. Following similar ideas as in the previous chapter, we give the limiting distribution of the tail probability estimator and the optimal sample fraction. A bootstrap procedure and simulation results are also considered.

The methods in Chapters 2 and 3 are rather general, in the sense that they can be applied to most consistent, asymptotically normal sets of estimators of (γ, a, b) involved in (1.1.3)-(1.1.5).

In Chapter 4 we use the results obtained in the previous chapters to optimise the construction of confidence intervals. We shall focus on the shape parameter and high quantiles, for $\gamma > 0$. When obtaining confidence intervals for these quantities, the common approach that we find in the literature is to use the normal distribution approximation with a non-optimal rate. We propose to use the optimal rate, but then additional problems arise, since a bias term with unkown sign has to be estimated. We provide an estimator for this sign and the full programme to obtain these optimal confidence intervals. We demonstrate the gain in coverage. Moreover we show the relevance of these confidence intervals by calculating the reduction in capital requirements in a Value at Risk exercise. Simulation results are also presented.

A further step in the optimisation of confidence intervals for the tail index γ, would have been to use the first term in Edgeworth expansion of the distribution of Hill's estimator. In Chapter 4 this expansion is obtained for the optimal rate. Although we consider this theoretical result interesting by itself, it turns out that to be used in the optimisation of confidence intervals brings extra difficulties. Namely the need to estimate new parameters.

In Chapter 5 we prove asymptotic normality of the so-called maximum likelihood estimator of the extreme value index. We start from the same equations as considered in Smith (1987). We use recently obtained limiting results on the empirical tail quantile function (Drees, 1998a).

In Chapter 6 (paper Draisma et al., 2001) we address questions related to bivariate extreme value theory. As mentioned before we extend Ledford and Tawn's model. Then one is able to prove asymptotic normality of the estimators they proposed for the dependence parameter η. Moreover we propose a new estimator and prove its asymptotic normality. The models can be used to device a test for asymptotic independence.

Another main question addressed in Chapter 6 is the estimation of the probability of a failure set. That is, the estimation of the probability that given high values of both variables, these values are both exceeded. We propose a procedure that works under asymptotic dependence as well as under asymptotic independence, and we prove consistency of the resulting estimator.

Chapter 2

On optimising the estimation of high quantiles of a probability distribution

Co-authors: Laurens de Haan and Liang Peng

To appear in Statistics

Abstract. One of the major aims of one-dimensional extreme value theory is to estimate quantiles outside the sample or at the boundary of the sample. The underlying idea of any method to do this is to estimate a quantile well inside the sample but near the boundary and then to shift it somehow to the right place. The choice of this "anchor quantile" plays a major role in the accuracy of the method. We present a bootstrap method to achieve the optimal choice of sample fraction in the estimation of either high quantile or endpoint estimation which extends earlier results by Hall and Weissman (1997) in the case of high quantile estimation. We give detailed results for the estimators used in Dekkers et al. (1989). An alternative way of attacking problems like this one is given in a paper by Drees and Kaufmann (1998).

2.1 Introduction

In problems of coastal safety, one wants to estimate the 10,000 years return level based on one hundred years of observations (de Haan, 1990). In finance one seeks a Value-at-Risk which is basically also a quantile at the boudary of the range of available observations (Jansen and de Vries, 1991; Danielsson and de Vries, 1997).

The situation is the following: we have a sample $X_1, X_2, \ldots, X_n$ from some unknown distribution function (d.f.) F and want to estimate the quantile corresponding to a probability close to 1 i.e. we want x_p with $1-F(x_p)=p$ and $p \le c/n$. This inequality means that, if we want to apply asymptotic theory and if in the limiting process we want to maintain this essential feature, we are forced to assume that in fact p depends on $n, p = p_n$ and $\lim_{n\to\infty} p_n = 0$. Then there are still several possibilities: $np_n \to c \in (0,\infty)$ or $np_n \to 0$ $(n\to\infty)$. In both cases purely non-parametric methods do not work. Only if $np_n \to \infty$ non-parametric methods can be successful (Einmahl, 1990). The use of models for the tail suggested by extreme value theory gives us a sensible way of extrapolating from an intermediate quantile to one outside the sample unless one uses one of the generalised Pareto distributions (GPd)

$$H_\gamma(x) := 1-(1+\gamma x)^{-1/\gamma} \quad \text{for those } x \text{ for which } 1+\gamma x > 0, \tag{2.1.1}$$

$(\gamma \in \mathbb{R})$ for modelling the tail of F. The tail condition for F is:

$$\lim_{t\to\infty} t\left\{1-F\left((1-F)^{\leftarrow}(\frac{1}{t})+xa(t)\right)\right\} = 1-H_\gamma(x) \tag{2.1.2}$$

for all x for which $0 < H_\gamma(x) < 1$ where $a(t)$ is a suitable positive function. This means for the quantile function that for $x > 0$

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

For our problem this means

$$(1-F)^{\leftarrow}(p_n) \approx (1-F)^{\leftarrow}(\frac{k}{n}) + a(\frac{n}{k})\frac{(\frac{k}{np_n})^\gamma-1}{\gamma},$$

i.e. an extreme quantile is linked to an intermediate quantile (which can be estimated via the empirical d.f.) by using the GPd approximation. The extreme quantile estimator based on this relation is

$$\hat{x}_{n,1}(k) := X_{n-k,n} + \hat{a}_1(\frac{n}{k})\frac{(\frac{k}{np_n})^{\hat{\gamma}_{n,1}(k)}-1}{\hat{\gamma}_{n,1}(k)} \tag{2.1.3}$$

where $X_{1,n} \le X_{2,n} \le \dots \le X_{n,n}$ are the order statistics and $\hat{a}_1(n/k)$ and $\hat{\gamma}_{n,1}(k)$ suitable estimators for $a(n/k)$ and γ (Weissman, 1978; Smith, 1984; Boos, 1984; Joe, 1987 and many others). A boundary case is $\gamma < 0$ and $p = 0$. Then the same expression (with $p_n \to 0$) can be used as an estimator of the right endpoint of the probability distribution, in the same GPd set-up.

The choice of k (or rather $n-k$, the index of the order statistics from where on the GPd approximation is believed to be valid) is crucial for the accuracy of the procedure. The optimal value depends on the underlying distribution and is a result of balancing variance and bias components. In this paper we present a bootstrap procedure to obtain this optimal value adaptively. The method is an extension of what we used for obtaining the optimal number of order statistics in estimating γ (Danielsson et al., 2001 and Draisma et al., 1999). The paper Hall and Weissman (1997) presents a (similar but different) bootstrap method for solving the same optimality problem, not for the quantile but for the exceeding probability of a high level which is similar. Unlike that paper, we do not assume any of the parameters known. Also our conditions allow for much smaller values of p_n. The quantile problem is more common in applications than the inverse problem of exceedance probabilities of a high level.

We restrict ourselves to the range $\gamma > -\frac{1}{2}$. This range is most important in applications and in this range it is most efficient to choose a sequence $k = k(n)$ in (2.1.3) that goes to infinity with n. Also, since we consider tail properties, we have to limit ourselves to sequence $k(n) = o(n), n \to \infty$. Hence we are dealing with intermediate sequences $k(n)$ (i.e., the corresponding order statistics $X_{n-k,n}$ are intermediate):

$$k(n) \to \infty, \ k(n)/n \to 0 \ (n \to \infty). \tag{2.1.4}$$

The main idea is the following. We seek

$$k_0(n) := \arg\inf_k \text{ as. } E(\hat{x}_{n,1}(k) - x_n)^2 \tag{2.1.5}$$

where as. E means the asymptotic expectation (according to the limit distribution, cf. Theorem 2.2.1 and the discussion thereafter) and k ranges from, say, $\log n$ to $n/(\log n)$ (this expresses the restriction to intermediate sequences and includes the optimal one). Since we are looking for an adaptive method for optimisation and since x_n and the averaging probability measure in (2.1.5) are unknown, we replace them with sample analogues. So we consider

$$E_n \left(\hat{x}_{n,1}(k) - \hat{x}_{n,2}(k)\right)^2 \mathbf{1}_{(|\hat{x}_{n,1}(k) - \hat{x}_{n,2}(k)| \leq k^\delta)}, \quad \delta > -1/2 \tag{2.1.6}$$

where $\hat{x}_{n,1}(k)$ is as before, E_n denotes averaging with respect to the empirical d.f. and

$$\hat{x}_{n,2}(k) := X_{n-k,n} + \hat{a}_2(\frac{n}{k}) \frac{(\frac{k}{np_n})^{\hat{\gamma}_{n,2}(k)} - 1}{\hat{\gamma}_{n,2}(k)} \tag{2.1.7}$$

with $\hat{a}_2(n/k)$ and $\hat{\gamma}_{n,2}(k)$ alternative estimators.

The reason why we put the indicator function $\mathbf{1}_{()}$ in (2.1.6) is to ensure the convergence of the mean square error (mse, say). For details see Draisma et al. (1999). Since $\delta > -1/2$, by the asymptotic normality the condition expressed by the indicator function will be satisfied most of the time.

The quantity (2.1.6) depends on the sample only and can be approximated using a bootstrap procedure where the bootstrap sample size has to be chosen of lower order than n in order to avoid unwanted extra randomness. Solving the optimisation problem for (2.1.6) makes sense since the value $k_0^*(n_1)$ minimizing (2.1.6) is asymptotically related to the value k_0 from (2.1.5) and in fact with the help of a second bootstrap we can get k_0 from $k_0^*(n_1)$.

For technical reasons (cf. Lemma 2.4.1) we have to exclude the case $\gamma = 0$ and also the cases where the convergence in (2.1.2) is very slow ($\rho = 0$, cf. Lemma 2.4.1).

The procedure for quantile and endpoint estimation is explained in Section 2.2 which also contains the main results. The most general setting is accounted in Section 2.2.1. We also consider two special cases separately. In quantile estimation, if one restricts to the case γ positive, the asymptotic results may be simplified. This is analysed in Section 2.2.2. All these results use the moment estimator (Dekkers et al., 1989) or simplified versions of it to estimate γ. In Section 2.2.3 we use instead a location-scale invariant estimator of γ in endpoint estimation. In Section 2.3 we present some simulation results and an application. Finally in Section 2.4 are the proofs of the results of Section 2.2.

Our methods could be applied to most consistent, asymptotically normal, sets of estimators. However we work out the details only for the estimators used in Dekkers et al. (1989).

2.2 Main results

2.2.1 Results for high quantile and endpoint estimation

We start by explaining the method in detail. Then we shall state the precise conditions and present the formal results.

We shall use explicit estimators for $a(\frac{n}{k})$ and γ which are as follows. Define for $j = 1, 2, 3$

$$M_n^{(j)} := \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n-i,n} - \log X_{n-k,n})^j, \tag{2.2.1}$$

$$\hat{\gamma}_{n,1}(k) := M_n^{(1)} + 1 - \frac{1}{2}(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}})^{-1}, \tag{2.2.2}$$

$$\hat{\gamma}_{n,2}(k) := \sqrt{M_n^{(2)}/2} + 1 - \frac{2}{3}(1 - \frac{M_n^{(1)}M_n^{(2)}}{M_n^{(3)}})^{-1}, \tag{2.2.3}$$

$$\hat{a}_1(\frac{n}{k}) := X_{n-k,n}M_n^{(1)}/\rho_1(\hat{\gamma}_{n,1}(k)) \tag{2.2.4}$$

$$\hat{a}_2(\frac{n}{k}) := X_{n-k,n}M_n^{(1)}/\rho_1(\hat{\gamma}_{n,2}(k)) \tag{2.2.5}$$

where $\hat{\gamma}_{n,1}(k)$ and $\hat{a}_1(\frac{n}{k})$ are the estimators in (2.1.3) and $\hat{\gamma}_{n,2}(k)$ and $\hat{a}_2(\frac{n}{k})$ the alternative estimators in (2.1.7), and $\rho_1(\gamma) = (1-\gamma_-)^{-1}$. We denote $\min(\gamma, 0)$ by γ_- and $\max(\gamma, 0)$ by γ_+.

Step 1 Select randomly and independently n_1 times ($n_1 = O(n^{1-\varepsilon})$, $0 < \varepsilon < 1/2$) a member from the set $\{X_1, X_2, \ldots, X_n\}$. Indicate the result by $X_1^*, X_2^*, \ldots X_{n_1}^*$. Form the order statistics $X_{1,n_1}^* \leq X_{2,n_1}^* \leq \cdots \leq X_{n_1,n_1}^*$ and compute the quantities

(2.1.3) and (2.1.7) from (2.2.1-2.2.5) on the basis of these order statistics.

We denote the resulting quantities by $\hat{\gamma}^*_{n_1,1}(k)$, $\hat{\gamma}^*_{n_1,2}(k)$, $\hat{a}^*_1(n_1/k)$ and $\hat{a}^*_2(n_1/k)$, $\hat{x}^*_{n,1}(k)$, $\hat{x}^*_{n,2}(k)$ for $k = 1, 2, \ldots, n_1 - 1$. Form

$$q^*_{n_1,k} = (\hat{x}^*_{n,1}(k) - \hat{x}^*_{n,2}(k))^2 \mathbf{1}_{(|\hat{x}^*_{n,1}(k) - \hat{x}^*_{n,2}(k)| \le k^\delta)}$$

on the basis of these bootstrap estimators.

Step 2 Repeat step 1 r times independently. This results in a sequence $q^*_{n_1,k,s}$, $k = 1, 2, \ldots, n_1 - 1$ and $s = 1, 2, \ldots, r$. Calculate

$$\frac{1}{r} \sum_{s=1}^{r} q^*_{n_1,k,s}.$$

Step 3 Minimize $\frac{1}{r}\sum_{s=1}^{r} q^*_{n_1,k,s}$ with respect to k but reject values which are very small or very near to n_1 (the statement of Theorem 2.2.3 will be valid if k ranges from $\log n_1$ to $n/\log n_1$). Denote the value of k where the minimum is obtained by $k^*_0(n_1)$.

Step 4 Repeat step 1 up to 3 independently with the number n_1 replaced by $n_2 = (n_1)^2/n$. So n_2 is smaller than n_1. This results in $k^*_0(n_2)$.

Step 5 Calculate

$$\hat{k}_0(n) := \frac{(k^*_0(n_1))^2}{k^*_0(n_2)} \frac{h(\hat{\gamma}^+_n(k), \hat{\gamma}^-_n(k), \hat{\rho}'_{n_1}(k^*_0))}{\overline{h}(\hat{\gamma}^+_n(k), \hat{\gamma}^-_n(k), \hat{\rho}'_{n_1}(k^*_0))} \tag{2.2.6}$$

with $\hat{\gamma}^+_n(k)$ and $\hat{\gamma}^-_n(k)$ any consistent estimators of γ_+ and γ_-,

$$\hat{\rho}'_{n_1}(k^*_0) := \frac{\log k^*_0(n_1)}{-2\log n_1 + 2\log k^*_0(n_1)} \tag{2.2.7}$$

and the functions h and $\overline{h}$ from Theorems 2.2.2 and 2.2.3 below respectively.

This $\hat{k}_0(n)$, which is obtained adaptively, is asymptotically as good as the optimal number of order statistics in (2.1.5).

Now in order to be able to present our main results we have to state the conditions.

Suppose that the underlying d.f. F is in the domain of attraction of an extreme value distribution (or equivalently that the observations above a large threshold have an asymptotic GPd distribution). We formulate this condition analytically in terms of the quantile-type function $U := (\frac{1}{1-F})^{\leftarrow}$:

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \tag{2.2.8}$$

for all positive x, where $a(t)$ is a suitable positive function. We shall need a second order refinement of this relation which reads as follows: there is a function $A(t) \to 0$ with constant sign near infinity such that for all $x > 0$

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = \frac{1}{\rho}\left[\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma}\right] \tag{2.2.9}$$

with $\rho \leq 0$. For the final result we shall have to require $\rho < 0$, $a(t) \sim c_1 t^\gamma$ and $A(t) \sim \bar{c}_2 t^\rho (t \to \infty)$. Note that (2.2.9) with $\gamma + \rho \neq 0$ is equivalent to

$$U(t) = c_0 + c_1 \frac{t^\gamma - 1}{\gamma} + c_2 t^{\gamma+\rho} + o(t^{\gamma+\rho}) \quad \text{with} \quad c_1 > 0, \quad c_2 \neq 0 \qquad (t \to \infty). \tag{2.2.10}$$

So (2.2.9) in fact motivates condition (2.2.10), which is used in Theorem 2.2.2. In the next theorem we give asymptotic normality under the more general condition (2.2.9).

In the sequel we shall need repeatedly the function $\tilde{A}(t)$ defined by

$$\tilde{A}(t) = \begin{cases} A(t) & \text{if } \gamma < \rho \\ \gamma_+ - \frac{a(t)}{U(t)} & \text{if } \rho < \gamma \leq 0 \\ & \quad \text{or } (0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) \neq 0) \\ & \quad \text{or } \gamma = -\rho \\ \rho A(t)/(\gamma+\rho) & \text{if } \gamma > -\rho \\ & \quad \text{or } (0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) = 0). \end{cases}$$

Then $|\tilde{A}(t)| \in RV_{\rho'}$, $\rho' < 0$ (cf. Lemma 2.4.1). Let $a_n = k/(np_n)$.

Theorem 2.2.1. *Suppose U satisfies* (2.2.9) *and $U(\infty) > 0$. Let $k(n)$ be an intermediate sequence and assume $\rho < 0$, $\gamma \neq 0$, $\gamma \neq \rho$ and $np_n \to c$ (finite, ≥ 0), as $n \to \infty$.*

(1) If $\tilde{A}\left(\frac{n}{k}\right)\sqrt{k} \to \lambda \in (-\infty, \infty)$ and $\log a_n/\sqrt{k} \to 0$ then

(a). ($\gamma > 0$)

$$R(n,k)\,(\hat{x}_{n,1}(k) - x_n) := \frac{\gamma\sqrt{k}}{a\left(\frac{n}{k}\right) a_n^\gamma \log a_n}\,(\hat{x}_{n,1}(k) - x_n)$$

converges in distribution to a normal r.v., N say, with mean $\lambda\sqrt{c_4}$ and variance c_3,

(b). ($\gamma < 0$)

$$R(n,k)\,(\hat{x}_{n,1}(k) - x_n) := \frac{\sqrt{k}}{a\left(\frac{n}{k}\right)}\,(\hat{x}_{n,1}(k) - x_n)$$

converges in distribution to a normal r.v., N say, with mean $\lambda\sqrt{c_6}$ and variance c_5.

(2) If $|\tilde{A}\left(\frac{n}{k}\right)|\sqrt{k} \to \infty$ *and* $\tilde{A}\left(\frac{n}{k}\right)\log a_n \to 0$ *then*

(a). ($\gamma > 0$)

$$R(n,k)\left(\hat{x}_{n,1}(k) - x_n\right) := \frac{\gamma}{\tilde{A}\left(\frac{n}{k}\right) a\left(\frac{n}{k}\right) a_n^{\gamma} \log a_n}\left(\hat{x}_{n,1}(k) - x_n\right)$$

converges in probability to $N = \sqrt{c_4}$,

(b). ($\gamma < 0$)

$$R(n,k)\left(\hat{x}_{n,1}(k) - x_n\right) := \frac{1}{\tilde{A}\left(\frac{n}{k}\right) a\left(\frac{n}{k}\right)}\left(\hat{x}_{n,1}(k) - x_n\right)$$

converges in probability to $N = \sqrt{c_6}$.

The constants c_3, c_4, c_5 *and* c_6 *are given in Theorem 2.2.2.*

Remark 2.2.1. Theorem 2.2.1 was proved under somewhat different conditions in de Haan and Rootzén (1993).

Remark 2.2.2. (i) $\tilde{A}\left(\frac{n}{k}\right)\sqrt{k} \to \lambda \in (-\infty, \infty)$ and $\log a_n/\sqrt{k} \to 0$ imply $\tilde{A}\left(\frac{n}{k}\right)\log a_n \to 0$; (ii) $|\tilde{A}\left(\frac{n}{k}\right)|\sqrt{k} \to \infty$ and $\tilde{A}\left(\frac{n}{k}\right)\log a_n \to 0$ imply $\log a_n/\sqrt{k} \to 0$.

Remark 2.2.3. In fact the conditions on $\log a_n$ are needed only when $\gamma > 0$.

Under (2.2.10) we have that $|\tilde{A}(t)| \sim \sqrt{\tilde{c}_2}\ t^{\rho'}$, $\tilde{c}_2 > 0$, as $t \to \infty$. Now we show that it follows from Theorem 2.2.1, that the best rate of convergence of $(\hat{x}_{n,1}(k) - x_n)$ is achieved in the case (1.) with $\lambda \neq 0$. For instance let $\gamma < 0$. Then, under (1.) and if $\lambda \neq 0$, k is of order $n^{-2\rho'/(1-2\rho')}$ (say k_0) and the rate of convergence $R(n, k_0)$ is of order $k_0^{\gamma+1/2}/n^{\gamma}$. Now let k_1 be other sequence for which (1.) holds but with $\lambda = 0$. Then $k_1 = o(k_0)$, hence $R(n,k_0)/R(n,k_1)$ is asymptotic to $(k_0/k_1)^{\gamma+1/2}$, which goes to infinity if $\gamma > -1/2$. Also notice that if k_2 is such that $|\tilde{A}(n/k_2)|\sqrt{k_2} \to \infty$, then the ratio of the rates $R(n,k_0)/R(n,k_2)$ is asymptotic to $k_0^{\gamma+1/2}/(k_2^{\gamma+\rho'} n^{-\rho'}) = (k_0/k_2)^{\gamma+\rho'} k_0^{1/2-\rho'}/n^{-\rho'}$, which again converges to infinity. The case $\gamma > 0$ is similar.

Also from Theorem 2.2.1, we have that $(\hat{x}_{n,1}(k) - x_n)$ is asymptotic to $R^{-1}(n,k)N$, where N is a r.v. In the next theorem we seek the optimal sequence $k_0(n)$ such that the mse of the approximating r.v. $R^{-1}(n,k)N$ is minimal. Hence

(a) ($\gamma > 0$)

$$k_0(n) = \arg\inf_k \frac{a^2\left(\frac{n}{k}\right) a_n^{2\gamma} \log^2 a_n}{\gamma^2}\left\{\frac{c_3}{k} + c_4\tilde{c}_2\left(\frac{n}{k}\right)^{2\rho'}\right\}, \tag{2.2.11}$$

(b) ($\gamma < 0$)

$$k_0(n) = \arg\inf_k a^2\left(\frac{n}{k}\right)\left\{\frac{c_5}{k} + c_6\tilde{c}_2\left(\frac{n}{k}\right)^{2\rho'}\right\}. \tag{2.2.12}$$

Next we need to get rid of the condition $\log a_n/\sqrt{k} \to 0$, since this is a condition on k and not only on p_n. So in order to stay in case (1.) of Theorem 2.2.1, so that the explicit formulation (2.2.11) holds, we require a condition on p_n alone, which is $\log p_n = o(n^{\frac{-\rho'}{1-2\rho'}})$.

For the rest of the paper we shall restrict ourselves to intermediate sequences $k(n)$ for which $(n/k)^{\rho'}\sqrt{k}$ converges to a finite or infinite constant.

Theorem 2.2.2. *Suppose U satisfies* (2.2.10) *and* $U(\infty) > 0$. *Assume* $\rho < 0$, $\gamma > -1/2$, $\gamma \neq 0$, $\gamma \neq \rho$, $\gamma \neq -\rho$, $np_n \to c$ *(finite,* ≥ 0*) and* $\log p_n = o\left(n^{\frac{-\rho'}{1-2\rho'}}\right)$. *Then* $k_0 = k_0(n)$ *(cf.* (2.2.11)-(2.2.12)*) satisfies*

$$k_0(n) \sim \begin{cases} \left(\frac{c_3}{c_4\tilde{c}_2(-2\rho')}\right)^{\frac{1}{1-2\rho'}} n^{\frac{-2\rho'}{1-2\rho'}} & \textit{for } \gamma > 0 \\ \left(\frac{c_5}{c_6\tilde{c}_2}\ \frac{1+2\gamma}{-2\rho'-2\gamma}\right)^{\frac{1}{1-2\rho'}} n^{\frac{-2\rho'}{1-2\rho'}} & \textit{for } \gamma < 0 \end{cases}$$

$$=: h(\gamma_+, \gamma_-, \rho')\, n^{\frac{-2\rho'}{1-2\rho'}}$$

with

$$\begin{aligned} c_3 &:= \frac{\gamma^2}{c_1^2} c_3(\gamma_+) := (\gamma_+^2 + 1) \\ c_4 &:= \frac{\gamma^2}{c_1^2} c_4(\gamma_+, \rho') := \\ &\begin{cases} \frac{(\gamma_+ + \rho' - \gamma_+\rho')^2}{\rho'^2(1-\rho')^4} & if\ (\lim_{t\to\infty} U(t) - \frac{a(t)}{\gamma_+} = 0 \textit{ and } 0 < \gamma < -\rho) \\ & \quad or\ \gamma > -\rho \\ \frac{(\gamma_+ + 2\rho' - \gamma_+\rho' - \rho'^2)^2}{(1-\rho')^4} & if\ (\lim_{t\to\infty} U(t) - \frac{a(t)}{\gamma_+} \neq 0 \textit{ and } 0 < \gamma < -\rho) \end{cases} \\ c_5 &:= c_5(\gamma_-) := \frac{(1-\gamma_-)^2(1-3\gamma_- + 4\gamma_-^2)}{\gamma_-^4(1-2\gamma_-)(1-3\gamma_-)(1-4\gamma_-)} \\ c_6 &:= c_6(\gamma_-, \rho') := \begin{cases} \frac{(1-\gamma_-)^2\rho'^2}{\gamma_-^4(1-\gamma_- - \rho')^2(\gamma_- + \rho')^2(1-2\gamma_- - \rho')^2} & \textit{if } \gamma < \rho \\ \frac{(3\gamma_-^2 - \gamma_- - 2\gamma_-^3 + 2\rho' - 2\gamma_-\rho' - \gamma_-^2\rho' - \rho'^2)^2}{\gamma_-^4(1-\gamma_-)^2(1-\gamma_- - \rho')^2(1-2\gamma_- - \rho')^2} & \textit{if } \rho < \gamma < 0. \end{cases} \end{aligned}$$

Remark 2.2.4. Since ρ is not known, one could alternatively require $\log p_n = o(n^{\varepsilon})$ for all $\varepsilon > 0$.

Corollary 2.2.1. *Under the conditions of Theorem 2.2.2,*

(1) if $\gamma > 0$

$$\frac{\sqrt{k_0}\,\gamma}{a\left(\frac{n}{k_0}\right)\left(\frac{k_0}{np_n}\right)^{\gamma}\log\left(\frac{k_0}{np_n}\right)}\ (\hat{x}_{n,1}(k_0) - x_n)$$

converges in distribution to a normal r.v. with variance c_3 *and mean* $\sqrt{c_4\tilde{c}_2}\,(h(\gamma_+,\gamma_-,\rho'))^{(1-2\rho')/2}$ *and,*

(2) if $\gamma < 0$

$$\frac{\sqrt{k_0}}{a\,(n/k_0)}\ (\hat{x}_{n,1}(k_0) - x_n)$$

converges in distribution to a normal r.v. with variance c_5 *and mean* $\sqrt{c_6\tilde{c}_2}\,(h(\gamma_+,\gamma_-,\rho'))^{(1-2\rho')/2}$.

Theorem 2.2.3. *Assume the conditions of Theorem 2.2.2. Then, as* $n \to \infty$ *(and* $r = r(n) \to \infty$*, with* r *the number of bootstrap repetitions, c.f. Step 2 above), for* $k_0(n)$ *as in* (2.2.11)-(2.2.12)

$$\lim_{n\to\infty}\frac{\hat{k}_0(n)}{k_0(n)} = 1$$

in probability, where (cf. (2.2.6)*)*

$$\hat{k}_0(n) = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)}\ \frac{h(\hat{\gamma}_n^+(k),\hat{\gamma}_n^-(k),\hat{\rho}'_{n_1}(k_0^*))}{\overline{h}(\hat{\gamma}_n^+(k),\hat{\gamma}_n^-(k),\hat{\rho}'_{n_1}(k_0^*))}$$

with $\hat{\gamma}_n^+(k),\hat{\gamma}_n^-(k)$ *any consistent estimators of* γ_+ *and* γ_-, $\hat{\rho}'_{n_1}(k_0^*)$ *as in* (2.2.7) *and the function* $\overline{h}$ *given by*

$$\overline{h}(\gamma_+,\gamma_-,\rho') := \begin{cases} \left(\frac{\overline{c}_3}{\overline{c}_4\,\tilde{c}_2(-2\rho')}\right)^{\frac{1}{1-2\rho'}} & \text{for } \gamma > 0 \\ \left(\frac{1+2\gamma}{-2\rho'-2\gamma}\ \frac{\overline{c}_5}{\tilde{c}_2\,\overline{c}_6}\right)^{\frac{1}{1-2\rho'}} & \text{for } \gamma < 0 \end{cases}$$

where

$$\begin{aligned}
\overline{c}_3 &:= \overline{c}_3(\gamma_+) := \frac{1}{4}(1+\gamma_+^2)\\
\overline{c}_4 &:= \overline{c}_4(\gamma_+,\rho') := \frac{(\rho'+\gamma_+-\gamma_+\rho')^2}{4(1-\rho')^6}\\
\overline{c}_5 &:= \overline{c}_5(\gamma_-) := \frac{(1-\gamma_-)^2(1-6\gamma_-+35\gamma_-^2-78\gamma_-^3+72\gamma_-^4)}{4\gamma_-^4(1-2\gamma_-)(1-3\gamma_-)(1-4\gamma_-)(1-5\gamma_-)(1-6\gamma_-)}\\
\overline{c}_6 &:= \overline{c}_6(\gamma_-,\rho') :=
\end{aligned}$$

$$\frac{(1-\gamma_-)^2\rho'^2}{4\gamma_-^4(1-\gamma_--\rho')^2(1-2\gamma_--\rho')^2(1-3\gamma_--\rho')^2} \qquad \text{if } \gamma<\rho\,,$$

$$\Bigg[\frac{-2+12\gamma_--22\gamma_-^2+12\gamma_-^3+5\rho'-22\gamma_-\rho'+21\gamma_-^2\rho'-6\rho'^2+12\gamma_-\rho'^2+2\rho'^3}{2\gamma_-^2(1-\gamma_-)(1-\gamma_--\rho')(1-2\gamma_--\rho')(1-3\gamma_--\rho')}$$
$$+\frac{2-14\gamma_-+34\gamma_-^2-34\gamma_-^3+12\gamma_-^4-6\rho'+30\gamma_-\rho'-46\gamma_-^2\rho'+22\gamma_-^3\rho'+6\rho'^2}{2\gamma_-^2(1-\gamma_-)(1-\gamma_--\rho')(1-2\gamma_--\rho')(1-3\gamma_--\rho')\sqrt{(1-\gamma_-)(1-2\gamma_-)}}$$
$$+\frac{-18\gamma_-\rho'^2+12\gamma_-^2\rho'^2-2\rho'^3+2\gamma_-\rho'^3}{2\gamma_-^2(1-\gamma_-)(1-\gamma_--\rho')(1-2\gamma_--\rho')(1-3\gamma_--\rho')\sqrt{(1-\gamma_-)(1-2\gamma_-)}}\Bigg]^2$$
$$\text{if } \rho<\gamma<0.$$

Moreover the result of Corollary 2.2.1 holds when $k_0(n)$ is replaced by $\hat{k}_0(n)$ throughout.

Remark 2.2.5. Since the order of magnitude is the same as in the case of minimizing the mean square error of the moment estimator $\hat{\gamma}_{n,1}(k)$ (only the constant differs and this factor can be estimated consistently), we could use the bootstrap procedure for one of them in order to get the optimal value for the other.

Next we turn our attention to the estimation of the right endpoint x_0 of the probability distribution when $\gamma<0$. Define (cf. Dekkers et al., 1989)

$$\hat{x}_{0,1}(k) := X_{n-k,n} - \frac{\hat{a}_1(\frac{n}{k})}{\hat{\gamma}_{n,1}^-(k)} \tag{2.2.13}$$

where

$$\hat{\gamma}_{n,1}^-(k) := 1-\frac{1}{2}\Big(1-\frac{(M_n^{(1)})^2}{M_n^{(2)}}\Big)^{-1}. \tag{2.2.14}$$

We seek

$$k_0(n) := \arg\inf_k \text{ as. } E(\hat{x}_{0,1}(k)-x_0)^2 \tag{2.2.15}$$

where as. E means the asymptotic expectation, according to the limit distribution. Similarly as before, from the proof of Theorem 2.2.4 we have that, for $\tilde{A}(n/k)\sqrt{k} \to \lambda \in (-\infty, \infty)$,

$$\frac{\sqrt{k}}{a\left(\frac{n}{k}\right)}\left(\hat{x}_{0,1}(k) - x_0\right)$$

converges in distribution to a normal r.v. with mean $\lambda\sqrt{c_8}$ and variance c_7, and if $|\tilde{A}(n/k)|\sqrt{k} \to \infty$ it converges, in probability, to $\sqrt{c_8}$. Moreover, under the conditions of Theorem 2.2.4 the best rate of convergence is attained when $\lambda \neq 0$. Hence we shall have

$$k_0(n) = \arg\inf_k \; a^2\left(\frac{n}{k}\right)\left\{\frac{c_7}{k} + c_8\tilde{c}_2\left(\frac{n}{k}\right)^{2\rho'}\right\}. \tag{2.2.16}$$

Theorem 2.2.4. *Suppose U satisfies (2.2.10) and $x_0 = U(\infty) > 0$. If $\rho < 0$, $-1/2 < \gamma < 0$ and $\gamma \neq \rho$, the value $k_0(n)$ of k minimizing the asymptotic second moment of $\hat{x}_{0,1}(k) - x_0$ (cf. (2.2.16)) satisfies*

$$k_0(n) \sim \left(\frac{1+2\gamma_-}{-2\rho' - 2\gamma_-}\frac{c_7}{\tilde{c}_2 c_8}\right)^{\frac{1}{1-2\rho'}} n^{\frac{-2\rho'}{1-2\rho'}} =: g(\gamma_-, \rho')\, n^{\frac{-2\rho'}{1-2\rho'}}.$$

with

$$c_7 := c_7(\gamma_-) := \frac{(1-\gamma_-)^2(1-3\gamma_- + 4\gamma_-^2)}{\gamma_-^4(1-2\gamma_-)(1-3\gamma_-)(1-4\gamma_-)}$$

$$c_8 := c_8(\gamma_-, \rho') := \begin{cases} \frac{(2\gamma_- - 6\gamma_-^2 + 4\gamma_-^3 + \rho' - 5\gamma_-\rho' + 6\gamma_-^2\rho' + 2\gamma_-\rho'^2)^2}{\gamma_-^4(1-\gamma_- - \rho')^2(\gamma_- + \rho')^2(1-2\gamma_- - \rho')^2} & \text{if } \gamma < \rho \\ \frac{(1-3\gamma_- + 2\gamma_-^2 + \gamma_-\rho')^2}{\gamma_-^4(1-\gamma_- - \rho')^2(1-2\gamma_- - \rho')^2} & \text{if } \rho < \gamma < 0. \end{cases}$$

In order to construct an adaptive estimator for $k_0(n)$ we consider the following alternative estimator for x_0,

$$\hat{x}_{0,2}(k) := X_{n-k,n} - \frac{\hat{a}_2(\frac{n}{k})}{\hat{\gamma}_{n,2}^-(k)} \tag{2.2.17}$$

where

$$\hat{\gamma}_{n,2}^-(k) := 1 - \frac{2}{3}\left(1 - \frac{M_n^{(1)}M_n^{(2)}}{M_n^{(3)}}\right)^{-1}. \tag{2.2.18}$$

Now for $\hat{x}_{0,1}(k)$ we apply the same bootstrap procedure as described before for $\hat{x}_{n,1}(k)$, but with the constants $h(\hat{\gamma}_n^+(k), \hat{\gamma}_n^-(k), \hat{\rho}'_{n_1}(k_0^*))$ and $\overline{h}(\hat{\gamma}_n^+(k), \hat{\gamma}_n^-(k), \hat{\rho}'_{n_1}(k_0^*))$ replaced by $g(\hat{\gamma}_n^-(k), \hat{\rho}'_{n_1}(k_0^*))$ and $\overline{g}(\hat{\gamma}_n^-(k), \hat{\rho}'_{n_1}(k_0^*))$ respectively.

Theorem 2.2.5. *Under the conditions of Theorem 2.2.4, as $n \to \infty$ (and $r = r(n) \to \infty$), $k_0(n)$ as in (2.2.16) satisfies*

$$\lim_{n\to\infty} \frac{\hat{k}_0(n)}{k_0(n)} = 1$$

in probability, where

$$\hat{k}_0(n) = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)} \frac{g(\hat{\gamma}_n^-(k), \hat{\rho}'_{n_1}(k_0^*))}{\overline{g}(\hat{\gamma}_n^-(k), \hat{\rho}'_{n_1}(k_0^*))}$$

with $\hat{\gamma}_n^-(k)$ any consistent estimate of γ_-, $\hat{\rho}'_{n_1}(k_0^*)$ as in (2.2.7) and the function $\overline{g}$ given by

$$\overline{g}(\gamma_-, \rho') := \left(\frac{1 + 2\gamma_-}{-2\rho' - 2\gamma_-} \frac{\overline{c}_7}{\tilde{c}_2 \overline{c}_8} \right)^{\frac{1}{1-2\rho'}}$$

where

$$\begin{aligned} \overline{c}_7 &:= \overline{c}_7(\gamma_-) := \frac{(1-\gamma_-)^2(1 - 6\gamma_- + 35\gamma_-^2 - 78\gamma_-^3 + 72\gamma_-^4)}{4\gamma_-^4(1-2\gamma_-)(1-3\gamma_-)(1-4\gamma_-)(1-5\gamma_-)(1-6\gamma_-)} \\ \overline{c}_8 &:= \overline{c}_8(\gamma_-, \rho') := \frac{((\gamma_- - 1)\rho')^2}{4\gamma_-^4(1-\gamma_- - \rho')^2(1-2\gamma_- - \rho')^2(1-3\gamma_- - \rho')^2}. \end{aligned}$$

Moreover when replacing $k_0(n)$ by $\hat{k}_0(n)$ the same asymptotic normal distribution is obtained.

2.2.2 Results for quantile, positive γ

Suppose we know, or assume, $\gamma > 0$ and want to estimate a high quantile. Confined to this situation, in this section we present the required asymptotic results to apply the bootstrap procedure as described in the last section. To estimate the quantile we use

$$\hat{x}_{n,1}^+(k) := X_{n-k,n} \left(\frac{k}{np_n} \right)^{\hat{\gamma}_{n,1}^+(k)} \qquad \text{where} \qquad \hat{\gamma}_n^+(k) := M_n^{(1)} \tag{2.2.19}$$

and let

$$\hat{x}_{n,2}^+(k) := X_{n-k,n} \left(\frac{k}{np_n} \right)^{\hat{\gamma}_{n,2}^+(k)} \qquad \text{where} \qquad \hat{\gamma}_{n,2}^+(k) := \sqrt{\frac{M_n^{(2)}}{2}} \tag{2.2.20}$$

be the alternative quantile estimator.

Theorem 2.2.6. *Suppose the second order condition (2.2.10) holds for $\gamma > 0$, $\rho < 0$, $\gamma \neq -\rho$ and $U(\infty) > 0$. Assume $np_n \to c$ (finite, ≥ 0) and $\log p_n = o(n^\varepsilon)$ for $\varepsilon > 0$, as $n \to \infty$. Then*

$$k_0(n) \sim \left(\frac{-\rho'(1-\rho')^2}{2\,\tilde{c}_2} \right)^{1/(1-2\rho')} n^{\frac{-2\rho'}{1-2\rho'}} =: l(\gamma_+, \rho')\, n^{\frac{-2\rho'}{1-2\rho'}}$$

where $k_0(n) := \arg\inf_k$ *as.* $E\left(\hat{x}^+_{n,1}(k) - x_n\right)^2$.

Theorem 2.2.7. *Under the conditions of Theorem 2.2.6, as* $n \to \infty$ *(and* $r = r(n) \to \infty$*),* $k_0(n)$ *satisfies*

$$\lim_{n\to\infty} \frac{\hat{k}_0(n)}{k_0(n)} = 1$$

in probability, where

$$\hat{k}_0(n) = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)} \frac{l(\hat{\gamma}_n^+(k), \hat{\rho}'_{n_1}(k_0^*))}{\bar{l}(\hat{\gamma}_n^+(k), \hat{\rho}'_{n_1}(k_0^*))}$$

with $\hat{\gamma}_n^+(k)$ *any consistent estimate of* γ_+, $\hat{\rho}'_{n_1}(k_0^*)$ *as in* (2.2.7) *and the function* $\bar{l}$ *given by*

$$\bar{l}(\gamma_+, \rho') := \left(\frac{(1-\rho')^4}{-2\rho'\tilde{c}_2}\right)^{1/(1-2\rho')}.$$

Moreover when replacing $k_0(n)$ *by* $\hat{k}_0(n)$ *the same asymptotic normal distribution is obtained.*

2.2.3 Results for endpoint with a shift-scale invariant estimator of γ

In this section, for the endpoint estimator we still use the same structure as in (2.2.13). Let

$$\hat{x}_{0,3}(k) \quad := \quad X_{n-k,n} - \frac{\hat{a}_3(\frac{n}{k})}{\hat{\gamma}^-_{n,3}(k)}. \tag{2.2.21}$$

The main difference now, lies in the quantities $M_n^{(j)}$ (cf. (2.2.1)). We propose to use instead

$$N_n^{(j)} \quad := \quad \frac{1}{k}\sum_{i=0}^{k-1}(X_{n-i,n} - X_{n-k,n})^j, \quad j = 1,2,3. \tag{2.2.22}$$

Since γ is negative we shall use

$$\hat{\gamma}^-_{n,3}(k) \quad := \quad 1 - \frac{1}{2}(1 - \frac{(N_n^{(1)})^2}{N_n^{(2)}})^{-1} \tag{2.2.23}$$

to estimate the extreme value index. Note that (2.2.23) is shift and scale invariant whilst the extreme value index estimators used in the previous sections are just scale invariant. In what concerns the estimation of $a(\frac{n}{k})$ similarly we propose

$$\hat{a}_3(\frac{n}{k}) \quad := \quad N_n^{(1)}/\rho_1(\hat{\gamma}^-_{n,3}(k)). \tag{2.2.24}$$

In what regards the alternative estimators necessary for the bootstrap procedure just apply the same scheme as in Section 2.2.1 for the endpoint. Substitute in (2.2.18) $M_n^{(j)}$, $j=1,2,3$ by $N_n^{(j)}$, $j=1,2,3$, respectively, to get $\hat{\gamma}_{n,4}^{-}(k)$. Substitute in (2.2.24) $\hat{\gamma}_{n,3}^{-}(k)$ by $\hat{\gamma}_{n,4}^{-}(k)$ to get $\hat{a}_4(\frac{n}{k})$, to finally obtain $\hat{x}_{0,4}(k)$.

We now state the main results. Note the resemblance with Theorem 2.2.4.

Theorem 2.2.8. *Suppose U satisfies (2.2.10) and $x_0 = U(\infty) > 0$. If $\rho < 0$ and $-1/2 < \gamma < 0$, the value $k_0(n)$ of k minimizing as. $E\left(\hat{x}_{0,3}(k) - x_0\right)^2$ satisfies*

$$k_0(n) \sim \left(\frac{1+2\gamma_-}{-2\rho - 2\gamma_-}\frac{c_7}{\bar{c}_2 c_8'}\right)^{\frac{1}{1-2\rho'}} n^{\frac{-2\rho'}{1-2\rho'}}.$$

The constant c_7 is given in Theorem 2.2.4 and

$$c_8' := \frac{(2\gamma_- - 6\gamma_-^2 + 4\gamma_-^3 + \rho - 5\gamma_-\rho + 6\gamma_-^2\rho + 2\gamma_-\rho^2)^2}{\gamma_-^4(1-\gamma_- - \rho)^2(\gamma_- + \rho)^2(1 - 2\gamma_- - \rho)^2}. \tag{2.2.25}$$

Theorem 2.2.9. *Under the conditions of Theorem 2.2.8, as $n \to \infty$ (and $r = r(n) \to \infty$), $k_0(n)$ satisfies*

$$\lim_{n\to\infty} \frac{\hat{k}_0(n)}{k_0(n)} = 1$$

in probability, where

$$\hat{k}_0(n) = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)} \frac{g(\hat{\gamma}_n^-(k), \hat{\rho}_{n_1}(k_0^*))}{\overline{g}(\hat{\gamma}_n^-(k), \hat{\rho}_{n_1}(k_0^*))}$$

with $\hat{\gamma}_n^-(k)$ any consistent estimate of γ_-,

$$\hat{\rho}_{n_1}(k_0^*) = \frac{\log k_0^*(n_1)}{-2\log n_1 + 2\log k_0^*(n_1)}$$

and the functions g and $\overline{g}$ from Theorems 2.2.4 and 2.2.5 respectively, with c_8 replaced by c_8' as in (2.2.25), and $\tilde{c}_2$ replaced by $\bar{c}_2$. Moreover when replacing $k_0(n)$ by $\hat{k}_0(n)$ the same asymptotic normal distribution is obtained.

2.3 Applications to simulated and real data

2.3.1 Simulation results

The simulations are based on the following three types of d.f.'s.

Generalised extreme value distribution

Let $G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}$, $1 + \gamma x > 0$. In accordance with the conditions in the previous theorems, in the following we exclude the cases $\gamma = 0, -1$. The function $U(t) = F^{\leftarrow}(1-1/t)$ is given by $U(t) = ((-\log(1-1/t))^{-\gamma}-1)/\gamma$, $t>1$, where $\lim_{t\to\infty} U(t) = U(\infty) = -1/\gamma$ if $\gamma<0$ and $U(\infty)=\infty$ if $\gamma>0$. Expanding the function $U(t)$, if $\gamma \neq 1$,

$$U(t) = \frac{t^\gamma - 1}{\gamma} - \frac{1}{2} t^{\gamma-1} + o\left(t^{\gamma-1}\right) \quad \text{as } t\to\infty ,$$

and if $\gamma = 1$,

$$U(t) = -\frac{1}{2} + (t-1) - \frac{t^{-1}}{12} + o\left(t^{-1}\right) \quad \text{as } t\to\infty .$$

Hence (2.2.10) holds with (ρ, c_0, c_1, c_2) equal to $(-1, 0, 1, -1/2)$ if $\gamma \neq 1$ and $(-2, -1/2, 1, -1/12)$ if $\gamma = 1$.

The functions required in the first and second order conditions in terms of $U(t)$ (see (2.2.9)) may be taken as $a(t) = c_1 t^\gamma = t^\gamma$ and, $A(t) = \rho(\gamma+\rho)c_2 t^\rho / c_1 = (\gamma-1)t^{-1}/2$ if $\gamma \neq 1$ and $t^{-2}/6$ if $\gamma = 1$, as $t\to\infty$. The function required in the second order condition for $\log U(t)$ (cf. Lemma 2.4.1) may be taken as $(t\to\infty)$

$$\widetilde{A}(t) = \begin{cases} A(t) = \frac{\gamma-1}{2}\, t^{-1} & ,\gamma<-1 \\ \frac{a(t)}{U(\infty)} = -\gamma\, t^\gamma & ,-1<\gamma<0 \\ \gamma - \frac{a(t)}{U(t)} \sim \frac{c_0 - c_1/\gamma}{c_1/\gamma}\, t^{-\gamma} = t^{-\gamma} & ,0<\gamma<1 \\ \gamma - \frac{a(t)}{U(t)} \sim \frac{3}{2}\, t^{-1} & ,\gamma = 1 \\ \frac{\rho A(t)}{\gamma+\rho} = \frac{t^{-1}}{2} & ,\gamma>1. \end{cases} \tag{2.3.1}$$

Note that $\lim_{t\to\infty}(U(t) - a(t)/\gamma) = -1/\gamma$ if $\gamma<0$, $-1/\gamma$ if $0<\gamma<1$, $-3/2$ if $\gamma=1$ and $-\infty$ if $\gamma>1$. Hence

$$\tilde{c}_2 = \begin{cases} (\gamma-1)^2/4 & ,\gamma<-1 \\ \gamma^2 & ,-1<\gamma<0 \\ 1 & ,0<\gamma<1 \\ 9/4 & ,\gamma=1 \\ 1/4 & ,\gamma>1. \end{cases}$$

Reversed Burr distribution

A random variable (r.v.) Y is said to have Burr d.f. with parameters β, λ and τ if $F_Y(y) = 1 - \beta^\lambda/(\beta + y^\tau)^\lambda$, $y>0$, $\beta,\lambda,\tau>0$. Let $X = -Y^{-1}$. Then X is said to have a Reversed Burr distribution, say $RB_{\beta,\lambda,\tau}$, with d.f. given by $F_X(x) = 1 - \beta^\lambda/(\beta + (-x)^{-\tau})^\lambda$, $x<0=x_0$, $\beta,\lambda,\tau>0$. In accordance with the conditions in the previous theorems, in the following we exclude the case $\tau = 1$. In order to properly use simulated data from this model it must be shifted by a positive constant, say a, so that $x_0 = a > 0$. Therefore we consider $U(t) = a - \beta^{-1/\tau}(t^{1/\lambda}-1)^{-1/\tau}$, $t>1$, and $\lim_{t\to\infty} U(t) = U(\infty) = a$. Expanding this function we get, as $t\to\infty$

$$U(t) = a - \beta^{-1/\tau} + \frac{\beta^{-1/\tau}}{\lambda\tau}\frac{t^{-1/\lambda\tau}-1}{-1/\lambda\tau} - \frac{\beta^{-1/\tau}}{\tau} t^{-1/\lambda\tau - 1/\lambda} + o\left(t^{-1/\lambda\tau-1/\lambda}\right) .$$

Hence (2.2.10) holds with $(\gamma, \rho, c_0, c_1, c_2)$ equal to $(-1/\lambda\tau$, $-1/\lambda$, $a-\beta^{-1/\tau}$, $\beta^{-1/\tau}/\lambda\tau$, $-\beta^{-1/\tau}/\tau)$. The functions required in the first and second order conditions in terms of $U(t)$ may be taken as $a(t) = \beta^{-1/\tau}t^{-1/\lambda\tau}/\lambda\tau$ and $A(t) = (1+\tau)t^{-1/\lambda}/\lambda\tau$, as $t \to \infty$. The function required in the second order condition in terms of $\log U(t)$ may be taken as $(t \to \infty)$

$$\widetilde{A}(t) = \begin{cases} \frac{1+\tau}{\lambda\tau}\, t^{-1/\lambda} & ,\tau < 1 \\ \frac{\beta^{-1/\tau}}{a\lambda\tau}\, t^{-1/\lambda\tau} & ,\tau > 1 \ . \end{cases}$$

Hence

$$\tilde{c}_2 = \begin{cases} (1+\tau)^2/(\lambda\tau)^2 & ,\tau < 1 \\ \beta^{-2/\tau}/(a\lambda\tau)^2 & ,\tau > 1. \end{cases}$$

Cauchy distribution.

Let X with d.f. $F_X(x) = (\arctan x + \pi/2)/\pi$, $x \in \mathbb{R}$. Then $U(t) = \tan(\pi/2 - \pi/t)$, $t > 1$ and $\lim_{t\to\infty} U(t) = U(\infty) = \infty$. Expanding this function we get

$$U(t) = 1/\pi + \frac{1}{\pi}(t-1) - \frac{\pi}{3}t^{-1} + o\left(t^{-1}\right) \quad \text{as} \quad t \to \infty \ ,$$

and so (2.2.10) holds with $(\gamma, \rho, c_0, c_1, c_2)$ equal to $(1, -2, 1/\pi, 1/\pi, -\pi/3)$. The functions required in the first and second order conditions in terms of $U(t)$ may be taken as $a(t) = t/\pi$, $A(t) = 2\pi^2 t^{-2}/3$, as $t \to \infty$, and the function required in the second order condition in terms of $\log U(t)$ may be taken as $\widetilde{A}(t) = 4\pi^2 t^{-2}/3$, as $t \to \infty$. Note that $\lim_{t\to\infty}(U(t) - a(t)/\gamma) = 0$. Hence $\tilde{c}_2 = 16\pi^4/9$.

Simulation results

We present results for the following distributions: $G_{-.25}$, $G_{.5}$, $G_{1.5}$, Cauchy and $RB_{4,4,2}$. For quantile estimation, for each df we estimate the quantile corresponding to a tail probability of $p_n = 1/(n \log n)$.

In Tables 2.1 and 2.2 are bootstrap results on endpoint and quantile estimation (based on the results of Section 2.2.1), based on 200 independent samples of size n = 2000 from $RB_{4,4,2}$ and $G_{-.25}$, respectively. We show the bootstrap estimates of k_0 and x_0 (Table 2.1) and, k_0 and x_n (Table 2.2), for several choices of n_1, namely $n_1 = 500(250)1750$. A general feature, which can be observed in these tables, is that the mean of the bootstrap estimates and respective mse (in the tables we show the square root of the mse), are quite stable along n_1.

We compare the performance of the bootstrap estimates with the true values, calculating the correspondent ratios. In Table 2.1, the 'true' value of k_0 was obtained from minimizing, over k, the sample mse of $(\hat{x}_n(k) - x_0)$, based on the same 200 independent samples used to obtain the estimates in the table, and where x_0 is the true value of the endpoint. The 'min rootmse' is the average over these samples of $(\hat{x}_n(k_0) - x_0)^2$. Similarly for the other tables.

n_1 (Interval to look for $k_0^*(n_1)$)	n_2 (Interval to look for $k_0^*(n_2)$)	$\hat{k}_0$ mean	mean/ true value	st. dev.	$\hat{x}_0$ $(x_0 = 0)$ mean	rootmse	rootmse/ min rootmse	% success
500 (10,400)	125 (10,100)	110.9	2.3	85.7	.00	.04	1.33	90
750 (10,600)	281 (10,224)	121.8	2.5	77.5	.00	.04	1.33	74
1000 (10,800)	500 (10,400)	121.8	2.6	84.3	.00	.04	1.33	74
1250 (10,1 000)	781 (10,624)	119.9	2.6	87.4	.00	.04	1.33	70
1500 (10,1 200)	1125 (10,900)	122.8	2.6	144.0	.00	.05	1.33	70
1750 (10,1 400)	1531 (10,1 224)	103.8	2.2	132.5	.00	.05	1.33	64

Table 2.1: Simulation results, bootstrap endpoint estimation with 200 independent samples of size 2 000 from $RB_{4,4,2}$ and $r = 200$ (see text for details).

As mentioned in Section 2.2.1, we reject values of $k_0^*(n_1)$ (and of $k_0^*(n_2)$) which are very small or very near to n_1 (repectively n_2). In all simulations (except when mentioned otherwise) we used as a lower bound the value 10. For the upper bound in all cases, except $RB_{4,4,2}$, it was determined according to the data: for each d.f. it is such that as many positive values as possible are used in the bootstrap samples. In fact all samples have approximately half of positive values, so this upper bound does not restrict our estimation. Indeed $k_0^*(n_1)$ and $k_0^*(n_2)$ are in general much lower than half of the respective bootstrap sample size. For the particular case $RB_{4,4,2}$ notice that the original samples are constituted of negative values. In this case we shifted the data by 16, which is the smallest integer such that all values are positive. Then we just took for the upper bound $.8\,n_i$, that is, a large enough value.

In what concerns the number of bootstrap resamples, 200 replications (denoted by $r = 200$) seem fairly enough in all cases.

As shown in the tables, not all simulations worked well. Mainly the abortions were because $k_0^*(n_2) > k_0^*(n_1)$, a situation we classified as inconsistent. For a few times we also observed other reasons like $\hat{k}_0(n) = 0$, $\hat{k}_0(n)$ exceeds the number of positive observations or $\hat{\gamma} > 0$ on endpoint estimation.

In Table 2.3 we consider samples of size 10 000. We give results for: (1) quantile and endpoint estimation based on the bootstrap algorithm described in Section 2.2.1; (2) endpoint estimation based on the bootstrap algorithm described in Section 2.2.3. In what concerns the bootstrap parameters, for n_1 we always considered 3981, which corresponds to $n_1 = n^{1-\varepsilon}$ where $\varepsilon = .1$. For the number of bootstrap resamples we set, as before, $r = 200$.

In general we observe that our estimates of quantile and endpoint are reasonably close to the correspondent optimum. For quantile estimation, $\gamma < 0$, the estimates of k_0 are not so stable as for the cases when $\gamma > 0$. We show the case $G_{-.25}$ where we observe a larger variability of the estimates of k_0.

In Table 2.3 the reader also finds simulation results based on the same samples

n_1 (Interval to look for $k_0^*(n_1)$)	n_2 (Interval to look for $k_0^*(n_2)$)	$\hat{k}_0$ mean	mean/ true value	st. dev.	$\hat{x}_n$ ($x_n = 3.77$) mean	rootmse	rootmse/ min rootmse	% success
500 (10,260)	125 (10,54)	303.0	1.2	275.8	3.59	.56	1.75	60
750 (10,392)	281 (10,139)	347.6	1.3	281.1	3.54	.63	2.0	61
1000 (10,547)	500 (10,261)	323.8	.9	276.8	3.61	.55	1.7	65
1250 (10,691)	781 (10,420)	337.6	.9	296.1	3.58	.58	1.8	68
1500 (10,837)	1125 (10,624)	299.4	.8	278.9	3.55	.54	1.7	68
1750 (10,994)	1531 (10,853)	294.3	.8	271.4	3.60	.55	1.7	62

Table 2.2: Simulation results, bootstrap quantile estimation with 200 independent samples of size 2000 from $G_{-.25}$ and $r = 200$ (see text for details).

as before, but taking simply $k = [\sqrt{n}]$ for the intermediate sequence to use in the estimation.

Simulation results regarding quantile estimation, positive gamma, are omitted since they follow a similar trend.

It seems that the methods do not give satisfactory results for samples of size under, approximately, 2 000.

2.3.2 Application

The goal is to estimate the right endpoint of the distribution pertaining to the following data sets. The data consists of the total life span (in days) of the people who died as residents in the Netherlands, which were born between the years 1877 - 1881 (included) and were still alive on January 1, 1971. Evidence has been given to support that the distribution of the population under study has a finite endpoint and the extreme value index is between $-1/2$ and 0; for a brief discussion we refer to Aarssen and de Haan (1994), where the same samples are analysed after suitable preparation for statistical analysis. The sample size is 10391. Results are also displayed for the women and men data separately, corresponding to samples of size 6260 and 4131, respectively.

In Table 2.4 are the results obtained from the bootstrap endpoint estimation, as explained in Section 2.2.1. As in the simulations discussed previously, for the choice of n_1 we used $\varepsilon = .1$, hence $n_1 = n^{.9}$. For the number of bootstrap resamples we used $r = 500$. Below each bootstrap sample size, n_1 and n_2, in round brackets, is the range taken to look for the optimal $k_0^*(n_1)$ and $k_0^*(n_2)$, respectively. As for the lower bound we always took 10, as in simulations. For the upper bound we took $.8\, n_i$, $i = 1, 2$, also the same as in simulations when the data was all positive.

Recall that the bootstrap uses a initial estimate of gamma to calculate $\hat{k}_0$. The initial estimates given in the table were obtained from the diagram of estimates, that is, plotting the estimates of γ against k. We consider three possible choices for

QUANTILE (Based on the results given in Section 2.2.1.)

	$\hat{k}_0$			$\hat{x}_n(\hat{k}_0)$					$\hat{x}_n([\sqrt{n}])$			
	mean	mean/ true value	st. dev.	mean	mean/ true value	rootmse	rootmse/ min rootmse	% success	mean	mean/ true value	rootmse	rootmse/ min rootmse
$G_{.5}$	588.0	.96	637.8	797.7	1.32	426.8	1.19	64	789.9	1.30	486.6	1.36
Cauchy	1243.4	1.04	673.3	3.55×10^4	1.21	1.98×10^4	1.34	81	4.18×10^4	1.42	4.54×10^4	3.09
$G_{1.5}$	1075.0	1.13	595.0	3.04×10^7	1.63	2.64×10^7	1.33	81	3.88×10^7	2.08	7.38×10^7	3.72
$G_{-.25}$	928.3	92.83	1050.2	3.71	.98	.29	1.95	60	3.79	1.01	.18	1.22
$G^*_{-.25}$	1050.3	.58	1126.1	3.71	.98	.30	1.99	74	3.80	1.01	.19	1.25

ENDPOINT

	$\hat{k}_0$			$\hat{x}_0(\hat{k}_0)$					$\hat{x}_0([\sqrt{n}])$			
	mean	mean/ true value	st. dev.	mean	mean/ true value	rootmse	rootmse/ min rootmse	% success	mean	mean/ true value	rootmse	rootmse/ min rootmse
$G^{(2)}_{-.25}$	1724.6	1.33	922.6	3.77	.94	.33	1.52	95	4.21	1.05	1.31	6.04
$RB^{(1)}_{4,4,2}$	306.2	.88	287.1	.00	-**	.01	1.15	69	.00	-**	.02	1.22
$RB^{(2)}_{4,4,2}$	745.5	2.15	741.6	-.01	-**	.02	1.84	68	.00	-**	.02	1.22

* With lower bound equal to 20.
** Not defined since the true value equals zero.
(1) Based on the results given in Section 2.2.1.
(2) Based on the results given in Section 2.2.3.

Table 2.3: Simulation results on bootstrap quantile and endpoint estimation with $n = 10\,000$, $r = 200$ and 200 independent simulations (see text for details).

size of the bootstrap resamples	intermediate bootstrap results		initial estimates	final bootstrap results		
	$k_0^*(n_1)$ $(k_0^*(n_1)/n_1)$	$k_0^*(n_2)$ $(k_0^*(n_2)/n_2)$	$\hat{\gamma}_{n,1}$	$\hat{k}_0(n)$ $(\hat{k}_0/n)$	$\hat{\gamma}_{n,1}$	$\hat{x}_0$
men+women sample $n = 10391$						
$n_1 = 4120; n_2 = 1633$ (10,3296) ; (10,1306)	2606 (.63)	1272 (.78)	-.1	92 (.01)	-.28	111.6 years
$n_1 = 4120; n_2 = 1633$ (10,3296) ; (10,1306)	2606 (.63)	1272 (.78)	-.15	198 (.02)	-.14	116.6 years
$n_1 = 4120; n_2 = 1633$ (10,3296) ; (10,1306)	2606 (.63)	1272 (.78)	-.2	318 (.03)	-.17	114.4 years
women sample $n = 6260$						
$n_1 = 2611; n_2 = 1089$ (10,2088) ; (10,871)	1858 (.71)	863 (.79)	-.15	148 (.02)	-.20	112.9 years
$n_1 = 2611; n_2 = 1089$ (10,2088) ; (10,871)	1858 (.71)	863 (.79)	-.2	238 (.04)	-.14	115.9 years
men sample $n = 4131$						
$n_1 = 1796; n_2 = 780$ (10,1436) ; (10,624)	1391(.77)	623(.80)	-.1	53 (.01)	-.07	127.5 years
$n_1 = 1796; n_2 = 780$ (10,1436) ; (10,624)	1391(.77)	623(.80)	-.15	115 (.03)	-.21	112.8 years

Table 2.4: Results of bootstrap on endpoint estimation of life span of men and women.

the men+women sample and two for the men sample and the women sample.

In general we observe a reduction in the variability of the estimates of the end-point, given by the bootstrap, when comparing these with the corresponding diagram of estimates (see Aarssen and de Haan, 1994). The men's sample shows more variability than the others. We mention that the largest observations correspond in fact to two men, at the age of 109 and 111 (cf. Aarssen and de Haan, 1994). For the women+men sample we consider that the solutions of 116.6 years and 114.4 years are the most reasonable ones. Nonetheless we still give the result when the intermediate value of gamma equals $-.1$, enhancing that large variability in the data (the sample indicates the value $-.1$ when k is small) may still be reflected in the final outcome.

2.4 Proofs

We start with a number of auxiliary results. The first one has been taken from Draisma et al. (1999).

Lemma 2.4.1. *Assume $U(\infty) > 0$ and there exist functions $a(t) > 0$ and $A(t) \to 0$, with $A(t)$ not changing sign eventually, such that*

$$\frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} \to H_{\gamma,\rho}(x)$$

where

$$H_{\gamma,\rho}(x) = \frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma - 1}{\gamma}\right) \qquad (\rho < 0).$$

Suppose that $\gamma \neq \rho$. Then

$$\lim_{t\to\infty} \frac{\frac{a(t)}{U(t)} - \gamma_+}{A(t)} = c \in [-\infty, \infty]$$

where

$$c = \begin{cases} 0 & \text{if } \gamma < \rho \\ \frac{\gamma}{\gamma+\rho} & \text{if } \gamma > -\rho \\ \frac{\gamma}{\gamma+\rho} & \text{if } 0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) = 0 \\ \pm\infty & \text{if } \rho < \gamma \leq 0 \\ \pm\infty & \text{if } 0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) \neq 0 \\ \pm\infty & \text{if } \gamma = -\rho. \end{cases}$$

Furthermore

$$\frac{\frac{\log U(tx) - \log U(t)}{a(t)/U(t)} - \frac{x^{\gamma_-}-1}{\gamma_-}}{\tilde{A}(t)} \to H_{\gamma_-,\rho'}(x) \tag{2.4.1}$$

where

$$\tilde{A}(t) = \begin{cases} A(t) & \text{if } c = 0 \\ \gamma_+ - \frac{a(t)}{U(t)} & \text{if } c = \pm\infty \\ \rho A(t)/(\gamma+\rho) & \text{if } c = \gamma/(\gamma+\rho), \end{cases}$$

$|\tilde{A}(t)| \in RV_{\rho'}$,

$$\rho' = \begin{cases} -\gamma & \text{if } (0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) \neq 0) \\ \gamma & \text{if } \rho < \gamma \leq 0 \\ \rho & \text{if } (0 < \gamma < -\rho \text{ and } \lim_{t\to\infty}(U(t) - a(t)/\gamma) = 0) \\ & \quad \text{or } \gamma < \rho \text{ or } \gamma \geq -\rho. \end{cases}$$

Remark 2.4.1. Hence $\rho' = 0$ if $\gamma = 0$.

Lemma 2.4.2. *Suppose for some function $a(t) > 0$ and function $A(t)$ not changing sign, $\lim_{t\to\infty} A(t) = 0$,*

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

for all $x > 0$, with $\rho < 0$. Then we have

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

for all functions $x(t)$ with $x(t) \to \infty$ $(t \to \infty)$. The same holds with $\rho = 0$ and $\gamma < 0$. Moreover, for $\gamma < 0$,

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

Proof. From Drees' inequality (1998a) it follows that

$$\lim_{t\to\infty} \sup_{x\geq 1} x^{-\gamma-\rho-\varepsilon} \left[\frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma-1}{\gamma}}{A(t)} - H_{\gamma,\rho}(x)\right] = 0. \qquad (2.4.2)$$

for negative ρ and each positive ε. The first result follows by considering the cases $\gamma > 0$, $\gamma = 0$ and $\gamma < 0$ separately.

As to the second result, relation (2.11) and Remark 2(i) from de Haan and Stadtmüller (1996) imply: $\lim_{t\to\infty}(U(t) - a(t)/\gamma) = U(\infty)$ and

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

The result follows. □

Remark 2.4.2. If $\{U(tx)/U(t) - x^\gamma\}/\alpha(t) \to x^\gamma(x^\rho - 1)/\rho$ for some function $\alpha(t)$, with $\gamma > 0$ and $\rho < 0$, as $t \to \infty$, for all $x > 0$, then we have

$$\lim_{t\to\infty} \{x(t)^{-\gamma} U(tx(t))/U(t) - 1\}/\alpha(t) = -1/\rho$$

for all functions $x(t)$ with $x(t) \to \infty$ $(t \to \infty)$.

Take random variables $Y_1, Y_2, \dots$ i.i.d. with d.f. $1 - 1/y$, $y > 1$. Then $U(Y_1)$, $U(Y_2)$, $\dots$ are i.i.d. F.

Lemma 2.4.3. *Write*

$$M_j := \frac{M_n^{(j)} U^j(Y_{n-k,n})}{a^j(Y_{n-k,n})} - l_j$$

for $j = 1, 2, 3$ *with*

$$\begin{aligned} M_n^{(j)} &:= \frac{1}{k}\sum_{i=0}^{k-1}\{\log U(Y_{n-i,n}) - \log U(Y_{n-k,n})\}^j, \\ 1/l_1 &:= 1 - \gamma_- \\ 1/l_2 &:= (1-\gamma_-)(1-2\gamma_-)/2 \\ 1/l_3 &:= (1-\gamma_-)(1-2\gamma_-)(1-3\gamma_-)/6. \end{aligned}$$

Then under the conditions of Lemma 2.4.1, for $k = k(n) \to \infty$ *and* $k(n)/n \to 0$ $(n \to \infty)$

$$\begin{aligned} M_1 &= \frac{P_1}{\sqrt{k}} + d_1\tilde{A}(\frac{n}{k}) + o_p(\frac{1}{\sqrt{k}}) + o_p(\tilde{A}(\frac{n}{k})) \\ M_2 &= \frac{P_2}{\sqrt{k}} + d_2\tilde{A}(\frac{n}{k}) + o_p(\frac{1}{\sqrt{k}}) + o_p(\tilde{A}(\frac{n}{k})) \\ M_3 &= \frac{P_3}{\sqrt{k}} + d_3\tilde{A}(\frac{n}{k}) + o_p(\frac{1}{\sqrt{k}}) + o_p(\tilde{A}(\frac{n}{k})) \end{aligned}$$

where (P_1, P_2, P_3) *is normally distributed with mean vector zero and covariance matrix*

$$\begin{cases} EP_1^2 = \frac{1}{(1-\gamma_-)^2(1-2\gamma_-)} \\ EP_2^2 = \frac{4(5-11\gamma_-)}{(1-\gamma_-)^2(1-2\gamma_-)^2(1-3\gamma_-)(1-4\gamma_-)} \\ EP_3^2 = \frac{36(19-105\gamma_-+146\gamma_-^2)}{(1-\gamma_-)^2(1-2\gamma_-)^2(1-3\gamma_-)^2(1-4\gamma_-)(1-5\gamma_-)(1-6\gamma_-)} \\ E(P_1P_2) = \frac{4}{(1-\gamma_-)^2(1-2\gamma_-)(1-3\gamma_-)} \\ E(P_1P_3) = \frac{18}{(1-\gamma_-)^2(1-2\gamma_-)(1-3\gamma_-)(1-4\gamma_-)} \\ E(P_2P_3) = \frac{12(9-21\gamma_-)}{(1-\gamma_-)^2(1-2\gamma_-)^2(1-3\gamma_-)(1-4\gamma_-)(1-5\gamma_-)} \end{cases}$$

and

$$\begin{cases} d_1 = \frac{1}{(1-\gamma_-)(1-\rho'-\gamma_-)} \\ d_2 = \frac{2(3-2\rho'-4\gamma_-)}{(1-\gamma_-)(1-2\gamma_-)(1-\rho'-\gamma_-)(1-\rho'-2\gamma_-)} \\ d_3 = \frac{6(18\gamma_-^2-22\gamma_-+15\rho'\gamma_-+3\rho'^2-8\rho'+6)}{(1-\gamma_-)(1-2\gamma_-)(1-3\gamma_-)(1-\rho'-\gamma_-)(1-\rho'-2\gamma_-)(1-\rho'-3\gamma_-)}. \end{cases}$$

Proof. From Lemma 2.4.1 and Drees' inequality (1998a),

$$\lim_{t\to\infty} \sup_{x\geq 1} x^{-\gamma_- -\rho'-\varepsilon} \left[\frac{\frac{\log U(tx)-\log U(t)}{a(t)/U(t)} - \frac{x^{\gamma_-}-1}{\gamma_-}}{\tilde{A}(t)} - H_{\gamma_-,\rho'}(x) \right] = 0 \tag{2.4.3}$$

for negative ρ' and each positive ε. Hence from this uniform convergence it follows that

$$\{\frac{1}{k}\sum_{i=0}^{k-1} \log U(Y_{n-i,n}) - \log U(Y_{n-k,n})\}\ U(Y_{n-k,n})/a(Y_{n-k,n})$$

$$= \frac{1}{k}\sum_{i=0}^{k-1} \frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\gamma_-} - 1}{\gamma_-} + \tilde{A}(Y_{n-k,n})\frac{1}{k}\sum_{i=0}^{k-1} H_{\gamma_-,\rho'}\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) + o_p(\tilde{A}(\frac{n}{k}))$$

$$\stackrel{d}{=} \frac{1}{k}\sum_{i=0}^{k-1}(Y_i^{\gamma_-}-1)/\gamma_- + \tilde{A}(\frac{n}{k})\frac{1}{k}\sum_{i=0}^{k-1} H_{\gamma_-,\rho'}(Y_i) + o_p(\tilde{A}(\frac{n}{k}))$$

$$= E(Y^{\gamma_-}-1)/\gamma_- + \frac{P_1}{\sqrt{k}} + \tilde{A}(\frac{n}{k})EH_{\gamma_-,\rho'}(Y) + o_p(\tilde{A}(\frac{n}{k})) + o_p(\frac{1}{\sqrt{k}})$$

with $Y, Y_1, Y_2, \ldots$ i.i.d. with d.f. $1-1/y$, $y>1$, and P_1 the normal limit random variable of

$$\sqrt{k}[\frac{1}{k}\sum_{i=1}^{k}(Y_i^{\gamma_-}-1)/\gamma_- - E(Y^{\gamma_-}-1)/\gamma_-].$$

Similarly for $M_n^{(j)}, j=2,3$; note that by Lemma 2.4.1

$$\left(\frac{\log U(tx)-\log U(t)}{a(t)/U(t)}\right)^j$$
$$= \left(\frac{x^{\gamma_-}-1}{\gamma_-}\right)^j + j\tilde{A}(t)\left(\frac{x^{\gamma_-}-1}{\gamma_-}\right)^{j-1} H_{\gamma_-,\rho'}(x) + o(\tilde{A}(t)),$$

hence

$$
\begin{aligned}
&M_n^{(j)}\{U(Y_{n-k,n})/a(Y_{n-k,n})\}^j\\
&\stackrel{d}{=} \frac{1}{k}\sum_{i=0}^{k-1}\left\{\frac{Y_i^{\gamma_-}-1}{\gamma_-}\right\}^j + j\tilde{A}(\frac{n}{k})\frac{1}{k}\sum_{i=0}^{k-1}\left\{\frac{Y_i^{\gamma_-}-1}{\gamma_-}\right\}^{j-1} H_{\gamma_-,\rho'}(Y_i) + o_p(\tilde{A}(\frac{n}{k}))\\
&= E\left\{\frac{Y^{\gamma_-}-1}{\gamma_-}\right\}^j + j\tilde{A}(\frac{n}{k})E\left\{\left(\frac{Y^{\gamma_-}-1}{\gamma_-}\right)^{j-1} H_{\gamma_-,\rho'}(Y)\right\} + \frac{P_j}{\sqrt{k}}\\
&\quad + o_p(\tilde{A}(\frac{n}{k})) + o_p(\frac{1}{\sqrt{k}}).
\end{aligned}
$$

□

Lemma 2.4.4. *Under the given conditions*

$$\hat{\gamma}_{n,1}^{+}(k) = M_n^{(1)} = \gamma_+ + \gamma_+ M_1 + q_{\gamma,\rho} l_1 \tilde{A}(\frac{n}{k}) + o_p(\tilde{A}(\frac{n}{k}))$$

$$\hat{\gamma}_{n,2}^{+}(k) = (M_n^{(2)}/2)^{1/2} = \gamma_+ + \frac{\gamma_+}{4}M_2 + q_{\gamma,\rho}(l_2/2)^{1/2}\tilde{A}(\frac{n}{k}) + o_p(\tilde{A}(\frac{n}{k})) + o_p(\frac{1}{\sqrt{k}})$$

with

$$q_{\gamma,\rho} = \lim_{t\to\infty}\frac{a(t)/U(t)-\gamma_+}{\tilde{A}(t)} = \begin{cases} 0 & \text{if } \gamma < \rho \\ \gamma/\rho & \text{if } (\lim_{t\to\infty} U(t) - a(t)/\gamma_+ = 0 \\ & \quad \text{and } 0 < \gamma < -\rho) \text{ or } \gamma > -\rho \\ -1 & \text{if } (\lim_{t\to\infty} U(t) - a(t)/\gamma_+ \neq 0 \\ & \quad \text{and } 0 < \gamma < -\rho) \\ & \quad \text{or } \rho < \gamma \leq 0 \text{ or } \gamma = -\rho \end{cases}$$

Proof.

$$M_n^{(1)} = a(Y_{n-k,n})/U(Y_{n-k,n})\{l_1 + M_1\}$$

$$= \frac{\frac{a(Y_{n-k,n})}{U(Y_{n-k,n})} - \gamma_+}{\tilde{A}(Y_{n-k,n})}\frac{\tilde{A}(Y_{n-k,n})}{\tilde{A}(\frac{n}{k})}\tilde{A}(\frac{n}{k})\{l_1 + M_1\} + \gamma_+\{l_1 + M_1\}$$

$$= q_{\gamma,\rho} l_1 \tilde{A}(\frac{n}{k}) + \gamma_+ + \gamma_+ M_1 + o_p(\tilde{A}(\frac{n}{k})).$$

$$\{M_n^{(2)}/2\}^{1/2} = \frac{a(Y_{n-k,n})}{U(Y_{n-k,n})}\{l_2/2 + M_2/2\}^{1/2}$$

$$= \frac{\frac{a(Y_{n-k,n})}{U(Y_{n-k,n})} - \gamma_+}{\tilde{A}(Y_{n-k,n})}\ \frac{\tilde{A}(Y_{n-k,n})}{\tilde{A}(\frac{n}{k})}\tilde{A}(\frac{n}{k})\{(l_2/2)^{1/2} + \frac{1}{4}\frac{M_2}{(l_2/2)^{1/2}}\}$$

$$+\gamma_+\{(l_2/2)^{1/2}+\frac{1}{4}\frac{M_2}{(l_2/2)^{1/2}}\}+o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}})$$

$$=q_{\gamma,\rho}(l_2/2)^{1/2}\tilde{A}(\frac{n}{k})+\gamma_++\gamma_+M_2/4+o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}}).$$

□

Lemma 2.4.5. *Under the given conditions*

$$\begin{aligned}
\hat{\gamma}^-_{n,1}(k) &= 1-\frac{1}{2}\{1-(M_n^{(1)})^2/M_n^{(2)}\}^{-1}\\
&= \gamma_- - \frac{4}{l_1l_2}M_1+\frac{2}{l_2^2}M_2+o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}})\\
&= \gamma_-+\frac{1}{2}(1-\gamma_-)^2(1-2\gamma_-)\{-4M_1+(1-2\gamma_-)M_2\}\\
&\quad +o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}}).\\
\hat{\gamma}^-_{n,2}(k) &= 1-\frac{2}{3}\left\{1-\frac{M_n^{(1)}M_n^{(2)}}{M_n^{(3)}}\right\}^{-1}\\
&= \gamma_- - \frac{3l_2}{2l_1^2l_3}M_1-\frac{3}{2l_1l_3}M_2+\frac{3l_2}{2l_1l_3^2}M_3+o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}})\\
&= \gamma_-+\frac{(1-\gamma_-)^2(1-3\gamma_-)}{12}\{-6M_1-3(1-2\gamma_-)M_2+\\
&\quad (1-2\gamma_-)(1-3\gamma_-)M_3\}+o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}}).
\end{aligned}$$

Remark 2.4.3. Hence for $\gamma>0$

$$\begin{aligned}
\hat{\gamma}^-_{n,1}(k) &= -2M_1+\frac{1}{2}M_2+o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}})\\
\hat{\gamma}^-_{n,2}(k) &= -\frac{1}{2}M_1-\frac{1}{4}M_2+\frac{1}{12}M_3+o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}}).
\end{aligned}$$

Proof of Lemma 2.4.5. For the expansion of $\hat{\gamma}^-_{n,1}(k)$ see Dekkers et al. (1989), proof of Corollary 3.2. Next we consider $\hat{\gamma}^-_{n,2}(k)$:

$$\begin{aligned}
\frac{M_n^{(1)}M_n^{(2)}}{M_n^{(3)}}-\frac{1-3\gamma_-}{3(1-\gamma_-)} &= \frac{(l_1+M_1)(l_2+M_2)}{l_3+M_3}-\frac{l_1l_2}{l_3}\\
&= \frac{l_2}{l_3}M_1+\frac{l_1}{l_3}M_2-\frac{l_1l_2}{l_3^2}M_3+o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}}).
\end{aligned}$$

Write $\pi:=M_n^{(1)}M_n^{(2)}/M_n^{(3)}$ and $\nu:=(1-3\gamma_-)/\{3(1-\gamma_-)\}$. Hence

$$\hat{\gamma}^-_{n,2}(k)-\gamma_-=1-\frac{2}{3}\frac{1}{1-\pi}-1+\frac{2}{3}\frac{1}{1-\nu}=\frac{2}{3}\frac{\nu-\pi}{(1-\nu)(1-\pi)},$$

that is,

$$
\begin{aligned}
&\hat{\gamma}^{-}_{n,2}(k) - \gamma_- \\
&= \frac{2}{3}(\nu - \pi)/(1-\nu)^2 \\
&= -\frac{2}{3}\{\frac{3}{2}(1-\gamma_-)\}^2 \left[\frac{l_2}{l_3}M_1 + \frac{l_1}{l_3}M_2 - \frac{l_1 l_2}{l_3^2}M_3 + o_p(\tilde{A}(\frac{n}{k})) + o_p(\frac{1}{\sqrt{k}})\right] \\
&= -\frac{3}{2}\left[\frac{l_2}{l_1^2 l_3}M_1 + \frac{1}{l_1 l_3}M_2 - \frac{l_2}{l_1 l_3^2}M_3 + o_p(\tilde{A}(\frac{n}{k})) + o_p(\frac{1}{\sqrt{k}})\right].
\end{aligned}
$$

□

Lemma 2.4.6. *Let $\hat{b}(n/k) = U(Y_{n-k,n})$. Under the given conditions*

$$\frac{\hat{b}(\frac{n}{k}) - U(\frac{n}{k})}{a(\frac{n}{k})} = \frac{B}{\sqrt{k}} + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k}))$$

with B a standard normal random variable, independent of P_1, P_2 and P_3.

Proof. We use the second order conditions for U.

$$
\begin{aligned}
\frac{U(Y_{n-k,n}) - U(\frac{n}{k})}{a(\frac{n}{k})} &= \frac{(\frac{k}{n}Y_{n-k,n})^\gamma - 1}{\gamma} + A(\frac{n}{k})H_{\gamma,\rho}(\frac{k}{n}Y_{n-k,n}) + o(A(\frac{n}{k})) \\
&= (\frac{k}{n}Y_{n-k,n} - 1) + o_p(\frac{k}{n}Y_{n-k,n} - 1) + A(\frac{n}{k})o_p(1) + o(A(\frac{n}{k})) \\
&= \frac{B}{\sqrt{k}} + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k}))
\end{aligned}
$$

□

Remark 2.4.4. No bias term comes into play.

Lemma 2.4.7. *Under the given conditions*

$$
\begin{aligned}
\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1 &= \frac{l_2 + 4l_1}{l_1 l_2}M_1 - \frac{2l_1}{l_2^2}M_2 + \gamma\frac{B}{\sqrt{k}} + o_p(\tilde{A}(\frac{n}{k})) + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k})) \\
&= (1-\gamma_-)(3-4\gamma_-)M_1 - \frac{1}{2}(1-\gamma_-)(1-2\gamma_-)^2 M_2 + \gamma\frac{B}{\sqrt{k}} \\
&\quad + o_p(\tilde{A}(\frac{n}{k})) + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k}))
\end{aligned}
$$

and

$$\begin{aligned}\frac{\hat{a}_2(\frac{n}{k})}{a(\frac{n}{k})}-1 &= \frac{2l_3+3l_2}{2l_1l_3}M_1+\frac{3}{2l_3}M_2-\frac{3l_2}{2l_3^2}M_3+\gamma\frac{B}{\sqrt{k}}\\ &\quad +o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k}))\\ &= \frac{3}{2}(1-\gamma_-)^2M_1+\frac{1}{4}(1-\gamma_-)(1-2\gamma_-)(1-3\gamma_-)M_2\\ &\quad -\frac{1}{12}(1-\gamma_-)(1-2\gamma_-)(1-3\gamma_-)^2M_3+\gamma\frac{B}{\sqrt{k}}\\ &\quad +o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k})).\end{aligned}$$

Proof.

$$\begin{aligned}\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} &= \frac{X_{n-k,n}M_n^{(1)}(1-\hat{\gamma}_{n,1}^-(k))}{a(\frac{n}{k})}\\ &= \frac{(1-\gamma_-)M_n^{(1)}U(Y_{n-k,n})}{a(Y_{n-k,n})}\,\frac{a(Y_{n-k,n})}{a(\frac{n}{k})}\,\frac{1-\hat{\gamma}_{n,1}^-(k)}{1-\gamma_-}.\end{aligned}$$

Now by the second order conditions for U

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

locally uniformly for $x>0$, hence

$$\frac{a(Y_{n-k,n})}{a(\frac{n}{k})}-1=\left(\frac{k}{n}Y_{n-k,n}\right)^\gamma-1+A(\frac{n}{k})(\frac{k}{n}Y_{n-k,n})^\gamma\frac{(\frac{k}{n}Y_{n-k,n})^\rho-1}{\rho}+o(A(\frac{n}{k}))$$

$$=\gamma(\frac{k}{n}Y_{n-k,n}-1)+o_p(\frac{k}{n}Y_{n-k,n}-1)+A(\frac{n}{k})o_p(1)+o(A(\frac{n}{k}))$$

$$=\gamma\frac{B}{\sqrt{k}}+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k})).$$

Consequently

$$\begin{aligned}\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} &= (1+(1-\gamma_-)M_1)(1+\gamma\frac{B}{\sqrt{k}}+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k}))(1-\frac{\hat{\gamma}_{n,1}^-(k)-\gamma_-}{1-\gamma_-})\\ &= 1+(1-\gamma_-)M_1+\gamma\frac{B}{\sqrt{k}}+\frac{4l_1}{l_1l_2}M_1-\frac{2l_1}{l_2^2}M_2\\ &\quad +o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k})).\end{aligned}$$

Similarly for

$$\begin{aligned}
\frac{\hat{a}_2(\frac{n}{k})}{a(\frac{n}{k})} &= \frac{X_{n-k,n}M_n^{(1)}(1-\hat{\gamma}_{n,2}^{-}(k))}{a(\frac{n}{k})} \\
&= (1+(1-\gamma_-)M_1)(1+\gamma\frac{B}{\sqrt{k}}+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k})))(1-\frac{\hat{\gamma}_{n,2}^{-}(k)-\gamma_-}{1-\gamma_-}).
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{\hat{a}_2(\frac{n}{k})}{a(\frac{n}{k})}-1 &= (1-\gamma_-)M_1+\gamma\frac{B}{\sqrt{k}}+\frac{3l_2}{2l_1l_3}M_1+\frac{3}{2l_3}M_2-\frac{3l_2}{2l_3^2}M_3 \\
&\quad +o_p(\tilde{A}(\frac{n}{k}))+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k})).
\end{aligned}$$

□

Proof of Theorem 2.2.1. Write $a_n := k/(np_n)$. As in de Haan and Rootzén (1993, p.7) we write

$$\begin{aligned}
\hat{x}_{n,1}(k)-x_n &= \frac{a_n^{\hat{\gamma}_{n,1}(k)}-1}{\hat{\gamma}_{n,1}(k)}\hat{a}_1(\frac{n}{k})+\hat{b}(\frac{n}{k})-U(\frac{1}{p_n}) \\
&= \left(\frac{a_n^{\hat{\gamma}_{n,1}(k)}-1}{\hat{\gamma}_{n,1}(k)}-\frac{a_n^{\gamma}-1}{\gamma}\right)\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})}a(\frac{n}{k}) \\
&\quad +\frac{a_n^{\gamma}-1}{\gamma}\left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})}-1\right)a(\frac{n}{k}) \\
&\quad +\frac{\hat{b}(\frac{n}{k})-U(\frac{n}{k})}{a(\frac{n}{k})}a(\frac{n}{k})-\left\{\frac{U(\frac{1}{p_n})-U(\frac{n}{k})}{a(\frac{n}{k})}-\frac{a_n^{\gamma}-1}{\gamma}\right\}a(\frac{n}{k}).
\end{aligned}$$

We have asymptotic expansions for $\hat{\gamma}_{n,1}(k), \hat{a}_1(n/k), \hat{b}(n/k)$ and also for the last term (the bias term) but not for $(a_n^{\hat{\gamma}_{n,1}(k)}-1)/\hat{\gamma}_{n,1}(k)$. So we want to simplify the expression (as in de Haan and Rootzén, 1993).

First suppose $\gamma > 0$. Hence

$$\begin{aligned}
\hat{x}_{n,1}(k)-x_n &\sim a(\frac{n}{k})a_n^{\gamma}\left[\left\{(1-a_n^{-\gamma})(\frac{1}{\hat{\gamma}_{n,1}(k)}-\frac{1}{\gamma})+\frac{a_n^{\hat{\gamma}_{n,1}(k)-\gamma}-1}{\hat{\gamma}_{n,1}(k)}\right\}\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})}\right. \\
&\quad +\frac{1-a_n^{-\gamma}}{\gamma}(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})}-1)+a_n^{-\gamma}\frac{\hat{b}(\frac{n}{k})-U(\frac{n}{k})}{a(\frac{n}{k})} \\
&\quad \left.-a_n^{-\gamma}\left\{\frac{U(\frac{1}{p_n})-U(\frac{n}{k})}{a(\frac{n}{k})}-\frac{a_n^{\gamma}-1}{\gamma}\right\}\right]
\end{aligned}$$

$$\sim a(\frac{n}{k})a_n^{\gamma}\left[\frac{1}{\hat{\gamma}_{n,1}(k)} - \frac{1}{\gamma} + \frac{a_n^{\hat{\gamma}_{n,1}(k)-\gamma} - 1}{\hat{\gamma}_{n,1}(k)} + \frac{1}{\gamma}(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1) - \frac{1}{\gamma}\frac{-1}{\rho + \gamma_-}A(\frac{n}{k})\right] \tag{2.4.4}$$

plus terms of lower order by the lemmas above for any intermediate sequence $k(n)$ and $n \to \infty$.

Next note that $a_n^{\hat{\gamma}_{n,1}(k)-\gamma}$ converges to one in probability, since $\tilde{A}\left(\frac{n}{k}\right)\sqrt{k} \to \lambda \in (-\infty, \infty)$ ensures that $\hat{\gamma}_{n,1}(k) - \gamma = O_p(\frac{1}{\sqrt{k}})$, and $\log a_n/\sqrt{k} \to 0$. Hence we may replace $(a_n^{\hat{\gamma}_{n,1}(k)-\gamma} - 1)/\hat{\gamma}_{n,1}$ by $\log a_n(\hat{\gamma}_{n,1}(k) - \gamma)/\hat{\gamma}_{n,1}$ in (2.4.4). Finally note that this term dominates all the other terms, as $n \to \infty$. The result now follows from Lemmas 2.4.4 and 2.4.5.

Next suppose $\gamma < 0$. Note that

$$\begin{aligned}
\hat{x}_{n,1}(k) - x_n &= a(\frac{n}{k})\left[(a_n^{\gamma} - 1)(\frac{1}{\hat{\gamma}_{n,1}(k)} - \frac{1}{\gamma}) + \frac{a_n^{\hat{\gamma}_{n,1}(k)} - a_n^{\gamma}}{\hat{\gamma}_{n,1}(k)}\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})}\right.\\
&\quad \left. + \frac{a_n^{\gamma} - 1}{\gamma}\left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1\right) + \frac{\hat{b}(\frac{n}{k}) - U(\frac{n}{k})}{a(\frac{n}{k})} - \frac{1}{\gamma}\frac{-1}{\rho + \gamma_-}A(\frac{n}{k})\right]\\
&= a(\frac{n}{k})\left[\frac{1}{\gamma} - \frac{1}{\hat{\gamma}_{n,1}(k)} + \frac{a_n^{\hat{\gamma}_{n,1}(k)} - a_n^{\gamma}}{\hat{\gamma}_{n,1}(k)} - \frac{1}{\gamma}\left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1\right) + \right.\\
&\quad \left. \frac{\hat{b}(\frac{n}{k}) - U(\frac{n}{k})}{a(\frac{n}{k})} - \frac{1}{\gamma(\rho + \gamma_-)}A(\frac{n}{k})\right]
\end{aligned}$$

plus terms of lower order, for any intermediate sequence $k(n)$.

Now

$$\frac{a_n^{\hat{\gamma}_{n,1}(k)} - a_n^{\gamma}}{\hat{\gamma}_{n,1}(k) - \gamma} = \frac{\log\ a_n}{\hat{\gamma}_{n,1}(k) - \gamma}\int_{\gamma}^{\hat{\gamma}_{n,1}(k)} a_n^s ds \le (\log\ a_n)a_n^{\max(\hat{\gamma}_{n,1}(k),\gamma)} \to 0$$

$(n \to \infty)$. Hence the second term $(a_n^{\hat{\gamma}_{n,1}(k)} - a_n^{\gamma})/\hat{\gamma}_{n,1}(k)$ is of smaller order than the first term $1/\gamma - 1/\hat{\gamma}_{n,1}(k)$. We find $(n \to \infty)$

$$\begin{aligned}
\hat{x}_{n,1}(k) - x_n &\sim a(\frac{n}{k})\left[\frac{\hat{\gamma}_{n,1}(k) - \gamma}{\gamma^2} - \frac{1}{\gamma}\left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1\right) + \frac{\hat{b}(\frac{n}{k}) - U(\frac{n}{k})}{a(\frac{n}{k})}\right.\\
&\quad \left. - \frac{1}{\gamma(\rho + \gamma_-)}A(\frac{n}{k})\right]\\
&\sim a(\frac{n}{k})\left[\frac{\hat{\gamma}_{n,1}(k) - \gamma}{\gamma^2} - \frac{1}{\gamma}\left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1\right) + \frac{\hat{b}(\frac{n}{k}) - U(\frac{n}{k})}{a(\frac{n}{k})}\right.\\
&\quad \left. - \frac{1_{\{\gamma < \rho\}}}{\gamma(\rho + \gamma_-)}\tilde{A}(\frac{n}{k})\right]
\end{aligned} \tag{2.4.5}$$

by Lemma 2.4.1. The result follows from the previous lemmas. □

Proof of Theorem 2.2.2. Since we are dealing with the *asymptotic* second moment it makes sense to first consider the limit behaviour in distribution rather than in L_2.

First suppose $\gamma > 0$. Note that under our conditions $a(\frac{n}{k})a_n^\gamma \sim c_1 p_n^{-\gamma}$. Consider (2.4.4) for the sequence $\tilde{k}(n) = [n^{-2\rho'/(1-2\rho')}]$. Then by the expressions of Lemmas 2.4.4 and 2.4.5 we have $\hat{\gamma}_{n,1}(k) - \gamma = O_p((\tilde{k}(n))^{-1/2})$ (see also Draisma et al., 1999). Hence, since $\log p_n = o(\sqrt{\tilde{k}(n)})$, $(\hat{\gamma}_{n,1}(k) - \gamma)\log a_n$ converges to zero for the sequence $\tilde{k}(n)$, and in fact the entire expression in square brackets tends to zero. This must then also be the case for the as yet unknown optimal sequence. Hence we may replace $(a_n^{\hat{\gamma}_{n,1}(k)-\gamma} - 1)/\hat{\gamma}_{n,1}(k)$ by $\log a_n(\hat{\gamma}_{n,1}(k) - \gamma)/\hat{\gamma}_{n,1}(k)$ in the minimization procedure. Since $(\log\ a_n)(\hat{\gamma}_{n,1}(k) - \gamma)/\hat{\gamma}_{n,1}(k)$ dominates all the other terms we find $(n \to \infty)$

$$\inf_k \text{ as. } E(\hat{x}_{n,1}(k) - x_n)^2 \sim \left(\frac{c_1 p_n^{-\gamma}}{\gamma}\right)^2 \inf_k \text{ as. } E(\log a_n)^2(\hat{\gamma}_{n,1}(k) - \gamma)^2. \quad (2.4.6)$$

By Lemmas 2.4.4 and 2.4.5, disregarding terms that are $o_p(\frac{1}{\sqrt{k}})$ or $o_p(\tilde{A}(\frac{n}{k}))$,

$$\begin{aligned}
\text{as. } E(\hat{\gamma}_{n,1}(k) - \gamma)^2 &= E\{(\gamma_+ - 2)M_1 + \frac{1}{2}M_2 + q_{\gamma,\rho}\tilde{A}(\frac{n}{k})\}^2 \\
&= E\{(\gamma_+ - 2)(\frac{P_1}{\sqrt{k}} + d_1\tilde{A}(\frac{n}{k})) + \frac{1}{2}(\frac{P_2}{\sqrt{k}} + d_2\tilde{A}(\frac{n}{k})) + q_{\gamma,\rho}\tilde{A}(\frac{n}{k})\}^2 \\
&= (\gamma_+ - 2)^2\frac{EP_1^2}{k} + \frac{1}{4}\frac{EP_2^2}{k} + (\gamma_+ - 2)\frac{EP_1P_2}{k} \\
&\quad + \{(\gamma_+ - 2)d_1 + \frac{1}{2}d_2 + q_{\gamma,\rho}\}^2\tilde{A}^2(\frac{n}{k}) \\
&=: \frac{\gamma^2 c_3(\gamma_+)}{c_1^2 k} + \frac{\gamma^2 c_4(\gamma_+, \rho')}{c_1^2}\tilde{A}^2(\frac{n}{k}).
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{\gamma^2}{c_1^2} as.\ E(\hat{x}_{n,1}(k) - x_n)^2 &\sim (\log\ a_n)^2 p_n^{-2\gamma}\left\{\frac{c_3(\gamma_+)}{k} + c_4(\gamma_+, \rho')\tilde{A}^2(\frac{n}{k})\right\} \\
&\sim p_n^{-2\gamma}(\log(\frac{k}{np_n}))^2\left\{\frac{c_3(\gamma_+)p_n^{-1}}{(k/(np_n))}\,\frac{1}{n} + c_4(\gamma_+, \rho')\tilde{c}_2(\frac{n}{k})^{2\rho'}\right\} \\
&\sim p_n^{-2\gamma-2\rho'}(\log(\frac{k}{np_n}))^2\left\{\frac{c_3(\gamma_+)p_n^{2\rho'-1}}{(k/(np_n))}\,\frac{1}{n} + c_4(\gamma_+, \rho')\tilde{c}_2(\frac{np_n}{k})^{2\rho'}\right\}.
\end{aligned}$$

So we are looking for

$$\arg\min_u p_n^{-2\gamma-2\rho'}\left\{(\log u)^2\frac{c_3 p_n^{2\rho'-1}}{u}\,\frac{1}{n} + c_4\tilde{c}_2(\log u)^2 u^{-2\rho'}\right\}.$$

Write $s := (\log u)^2/u$. Then $u \sim s^{-1}(\log s)^2 (u \to \infty)$ and we are dealing with

$$\arg\min_s p_n^{-2\gamma-2\rho'} \left\{ \frac{c_3 p_n^{2\rho'-1}}{n} s + c_4 \tilde{c}_2 s^{2\rho'} (\log s)^{2(1-2\rho')} \right\}.$$

This can be minimized by setting the derivative equal to zero. The result is

$$\begin{aligned} \frac{c_3}{\tilde{c}_2 c_4 (-2\rho')} \frac{p_n^{2\rho'-1}}{n} &= s^{2\rho'-1} (\log s)^{2(1-2\rho')} + \frac{1-2\rho'}{\rho'} s^{2\rho'-1} (\log\ s)^{2(1-2\rho')-1} \\ &\sim s^{2\rho'-1} (\log s)^{2(1-2\rho')}. \end{aligned}$$

That is,

$$\frac{1}{u} \sim \frac{s}{(\log\ s)^2} = \left(\frac{c_4 \tilde{c}_2 (-2\rho')}{c_3} \right)^{\frac{1}{1-2\rho'}} p_n n^{\frac{1}{1-2\rho'}}.$$

Note that the right hand side tends to zero since $np_n \to c$ (finite, ≥ 0). Now, replacing u by $k/(np_n)$, we get

$$\frac{k}{np_n} \sim \left(\frac{c_4 \tilde{c}_2 (-2\rho')}{c_3} \right)^{\frac{-1}{1-2\rho'}} p_n^{-1} n^{\frac{-1}{1-2\rho'}}$$

or

$$k_0(n) \sim \left(\frac{c_3}{c_4 \tilde{c}_2 (-2\rho')} \right)^{\frac{1}{1-2\rho'}} n^{\frac{-2\rho'}{1-2\rho'}}.$$

Note that $k_0(n)$ does not depend on p_n.

Next we consider as. $E(\hat{x}_{n,1}(k) - x_n)^2$ for $\gamma < 0$. From (2.4.5), disregarding terms which are $o_p(\frac{1}{\sqrt{k}})$ or $o_p(\tilde{A}(\frac{n}{k}))$,

$$\begin{aligned} \hat{x}_{n,1}(k) - x_n &= \\ &= a(\frac{n}{k}) \left[\frac{q_{\gamma,\rho} l_1}{\gamma^2} \tilde{A}(\frac{n}{k}) - \frac{4}{\gamma^2 l_1 l_2} M_1 + \frac{2}{\gamma^2 l_2^2} M_2 \right. \\ &\quad \left. - \frac{1}{\gamma} \left\{ \left(\frac{1}{l_1} + \frac{4}{l_2} \right) M_1 - \frac{2l_1}{l_2^2} M_2 \right\} - \frac{B}{\sqrt{k}} + \frac{B}{\sqrt{k}} - \frac{1_{\{\gamma<\rho\}}}{\gamma(\rho+\gamma_-)} \tilde{A}(\frac{n}{k}) \right] \\ &= a(\frac{n}{k}) \left[\left(\frac{-4}{\gamma^2 l_1 l_2} - \frac{1}{\gamma l_1} - \frac{4}{\gamma l_2} \right) M_1 + \left(\frac{2}{\gamma^2 l_2^2} + \frac{2l_1}{\gamma l_2^2} \right) M_2 \right. \\ &\quad \left. + \left(\frac{q_{\gamma,\rho} l_1}{\gamma^2} - \frac{1_{\{\gamma<\rho\}}}{\gamma(\rho+\gamma_-)} \right) \tilde{A}(\frac{n}{k}) \right]. \end{aligned}$$

Hence

$$\begin{aligned}
\frac{\text{as. } E(\hat{x}_{n,1}(k)-x_n)^2}{a^2(\frac{n}{k})} &= \\
&= \left(-\frac{4}{\gamma^2 l_1 l_2}-\frac{1}{\gamma l_1}-\frac{4}{\gamma l_2}\right)^2 \frac{EP_1^2}{k}+\left(\frac{2}{\gamma^2 l_2^2}+\frac{2l_1}{\gamma l_2^2}\right)^2 \frac{EP_2^2}{k} \\
&\quad+\left(-\frac{4}{\gamma^2 l_1 l_2}-\frac{1}{\gamma l_1}-\frac{4}{\gamma l_2}\right)\left(\frac{2}{\gamma^2 l_2^2}+\frac{2l_1}{\gamma l_2^2}\right)\frac{EP_1P_2}{k} \\
&\quad+\left\{\left(-\frac{4}{\gamma^2 l_1 l_2}-\frac{1}{\gamma l_1}-\frac{4}{\gamma l_2}\right)d_1+\left(\frac{2}{\gamma^2 l_2^2}+\frac{2l_1}{\gamma l_2^2}\right)d_2\right. \\
&\quad\left.+\left(\frac{q_{\gamma,\rho}l_1}{\gamma^2}-\frac{1_{\{\gamma<\rho\}}}{\gamma(\rho+\gamma_-)}\right)\right\}^2 \tilde{A}^2(\frac{n}{k}) \\
&=: \frac{c_5(\gamma_-)}{k}+c_6(\gamma_-,\rho')\tilde{A}^2(\frac{n}{k})=\frac{c_5(\gamma_-)}{k}+\tilde{c}_2c_6(\gamma_-,\rho')(\frac{n}{k})^{2\rho'}.
\end{aligned}$$

Hence

$$\begin{aligned}
\text{as. } E(\hat{x}_{n,1}(k)-x_n)^2 &= c_1^2(\frac{n}{k})^{2\gamma}\left\{\frac{c_5}{k}+c_6\tilde{c}_2(\frac{n}{k})^{2\rho'}\right\} \\
&= c_1^2 n^{2\gamma}\left\{\frac{c_5}{k^{1+2\gamma}}+c_6\tilde{c}_2\frac{n^{2\rho'}}{k^{2\gamma+2\rho'}}\right\}.
\end{aligned}$$

By assumption $1+2\gamma>0$. Write $t:=k^{-(1+2\gamma)}$. We want to minimize

$$tc_5+\tilde{c}_2c_6n^{2\rho'}t^{\frac{2\rho'+2\gamma}{1+2\gamma}}.$$

Setting the derivative equal to zero yields

$$k^{1-2\rho'}=t^{\frac{2\rho'+2\gamma}{1+2\gamma}-1}=\frac{c_5}{\tilde{c}_2c_6}n^{-2\rho'}\frac{1+2\gamma}{-2\rho'-2\gamma},$$

i.e. $(n\to\infty)$

$$k_0(n)\sim\left(\frac{1+2\gamma}{-2\rho'-2\gamma}\,\frac{c_5}{\tilde{c}_2c_6}\right)^{\frac{1}{1-2\rho'}}n^{\frac{-2\rho'}{1-2\rho'}}.$$

□

Proof of Corollary 2.2.1. From Theorem 2.2.2, $|\tilde{A}(n/k_0)|\sqrt{k_0}\sim\sqrt{\tilde{c}_2}(n/k_0)^{\rho'}\sqrt{k_0}\sim\sqrt{\tilde{c}_2}(h(\gamma_+,\gamma_-,\rho'))^{(1-2\rho')/2}>0$. The result follows easily from Theorem 2.2.1. □

Proposition 2.4.1. *Under the conditions of Theorem 2.2.2, as $n\to\infty$,*

$$\overline{k}_0(n)\sim\begin{cases}\left(\frac{\overline{c}_3}{\overline{c}_4\tilde{c}_2(-2\rho')}\right)^{\frac{1}{1-2\rho'}}n^{\frac{-2\rho'}{1-2\rho'}} & \textit{for } \gamma>0\\ \left(\frac{1+2\gamma}{-2\rho'-2\gamma}\,\frac{\overline{c}_5}{\tilde{c}_2\overline{c}_6}\right)^{\frac{1}{1-2\rho'}}n^{\frac{-2\rho'}{1-2\rho'}} & \textit{for } \gamma<0\end{cases}$$

$$= \overline{h}(\gamma_+, \gamma_-, \rho')\, n^{\frac{-2\rho'}{1-2\rho'}}$$

where $\overline{k}_0(n) := \arg\inf_k$ *as.* $E(\hat{x}_{n,1}(k) - \hat{x}_{n,2}(k))^2$,

$$\hat{x}_{n,2}(k) := X_{n-k,n} + \hat{a}_2(\frac{n}{k}) \frac{(\frac{k}{np_n})^{\hat{\gamma}_{n,2}(k)} - 1}{\hat{\gamma}_{n,2}(k)},$$

and $\tilde{c}_2$ *from* $|\tilde{A}(t)| \sim \sqrt{\tilde{c}_2} t^{\rho'}$ $(t \to \infty)$.

Proof. For $\gamma > 0$, neglecting terms which are $o_p(\frac{1}{\sqrt{k}})$ or $o_p(\tilde{A}(\frac{n}{k}))$, and by similar arguments as in the proof of Theorem 2.2.2, the dominant term in the expansion of $\hat{x}_{n,1}(k) - \hat{x}_{n,2}(k)$ turns out to be

$$\begin{aligned} & c_1 p_n^{-\gamma} \gamma^{-1} (\log a_n)(\hat{\gamma}_{n,1}(k) - \hat{\gamma}_{n,2}(k)) \\ \sim \;& c_1 p_n^{-\gamma} \gamma^{-1} \log a_n \left[(\gamma_+ - \frac{3}{2}) M_1 + \frac{1}{4}(3 - \gamma_+) M_2 - \frac{1}{12} M_3 \right]. \end{aligned}$$

Hence

$$\begin{aligned} & \text{as. } E(\hat{x}_{n,1}(k) - \hat{x}_{n,2}(k))^2 \\ =: \;& c_1^2 p_n^{-2\gamma} \gamma^{-2} \left[\overline{c}_3(\gamma_+) \frac{(\log a_n)^2}{k} + \overline{c}_4(\gamma_+, \rho')(\log a_n)^2 \tilde{A}^2(\frac{n}{k}) \right]. \end{aligned}$$

The result follows similarly as in Theorem 2.2.2.

Next suppose $\gamma < 0$. Then, similarly as before (neglecting terms which are $o_p(\frac{1}{\sqrt{k}})$ or $o_p(\tilde{A}(\frac{n}{k}))$),

$$\begin{aligned} & \hat{x}_{n,1}(k) - \hat{x}_{n,2}(k) = \\ = \;& a(\frac{n}{k}) \left\{ \frac{\hat{\gamma}_{n,1}(k) - \hat{\gamma}_{n,2}(k)}{\gamma^2} - \frac{1}{\gamma} \left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1 \right) + \frac{1}{\gamma} \left(\frac{\hat{a}_2(\frac{n}{k})}{a(\frac{n}{k})} - 1 \right) \right\} \\ = \;& a(\frac{n}{k}) \left[\frac{1}{\gamma^2} \left\{ q_{\gamma,\rho} l_1 \tilde{A}(\frac{n}{k}) - q_{\gamma,\rho} (\frac{l_2}{2})^{1/2} \tilde{A}(\frac{n}{k}) + \left(-\frac{4}{l_1 l_2} + \frac{3 l_2}{2 l_1^2 l_3} \right) M_1 \right. \right. \\ & + \left(\frac{2}{l_2^2} + \frac{3}{2 l_1 l_3} \right) M_2 - \frac{3 l_2}{2 l_1 l_3^2} M_3 \Big\} \\ & \left. + \frac{1}{\gamma} \left\{ \left(-\frac{4}{l_2} + \frac{3 l_2}{2 l_1 l_3} \right) M_1 \left(\frac{2 l_1}{l_2^2} + \frac{3}{2 l_3} \right) M_2 - \frac{3 l_2}{2 l_3^2} M_3 \right\} \right] \end{aligned}$$

Hence

$$\begin{aligned} \text{as. } E(\hat{x}_{n,1}(k) - \hat{x}_{n,2})^2 \;&=:\; a^2(\frac{n}{k}) \left(\overline{c}_5(\gamma_-) \frac{1}{k} + \overline{c}_6(\gamma_-, \rho') \tilde{A}^2(\frac{n}{k}) \right) \\ &=\; c_1^2 (\frac{n}{k})^{2\gamma} \left(\frac{\overline{c}_5}{k} + \overline{c}_6 \tilde{c}_2 (\frac{n}{k})^{2\rho'} \right). \end{aligned}$$

The result follows similarly as in Theorem 2.2.2. □

Proof of Theorem 2.2.3. We proceed as in the previous theorems and proposition but there are two differences. Firstly we need an expansion for the empirical tail function

$$U_n := \left(\frac{1}{1-F_n}\right)^{\leftarrow}$$

instead of the real one U. Secondly we are no longer dealing with asymptotic mean square error but with real finite sample mean square error. We can deal with both difficulties exactly as in Draisma et al. (1999): inequalities for U_n are obtained by using well-known results for the uniform quantile process, which we combine with the results of Theorem 2.2.2 and Proposition 2.4.1, and the mean square error becomes tractable by restricting to a certain set.

Let $k_{0,1}(n)$ be such that $E\left(\hat{x}_{n,1}(k)-\hat{x}_{n,2}(k)\right)^2 \mathbf{1}_{(|\hat{x}_{n,1}(k)-\hat{x}_{n,2}(k)|\le k^\delta)}$, $\delta > -1/2$, is minimal. The statements, $k_{0,1}(n)/\bar{k}_0(n) \to 1$ and

$$\frac{E\left(\hat{x}_{n,1}(k_{0,1})-\hat{x}_{n,2}(k_{0,1})\right)^2 \mathbf{1}_{\left(|\hat{x}_{n,1}(k_{0,1})-\hat{x}_{n,2}(k_{0,1})|\le k_{0,1}^\delta\right)}}{as.E\left(\hat{x}_{n,1}(\bar{k}_0)-\hat{x}_{n,2}(\bar{k}_0)\right)^2} \to 1,$$

as $n \to \infty$, follow as in the proof of Theorem 3.3. of Draisma et al. (1999). The expansions of the bootstrap estimators are similar to the ones for the estimators themselves, now using the inequality for U_n from Lemma 5.3 of Draisma et al. (1999), instead of using (2.4.3) for $\log U$. Some details are in the proof of Theorem 3.4 of the aforementioned paper, where also it becomes clear that the condition $n_1 = O(n^{1-\varepsilon})$, implies that the bootstrap set-up does not result in extra (unwanted) terms in the expansions.

Finally we sketch why, when replacing $k_0(n)$ by $\hat{k}_0(n)$ in Corollary 2.2.1, the same asymptotic normal distribution obtains.

Let $t \in [t_1, t_2]$, $0 < t_1 < t_2 < \infty$. Proceeding as in the proof of Lemma 2.4.3 one gets, e.g. if $\tilde{A}(n/k)\sqrt{k} \to \lambda$, hence $\tilde{A}(n/[kt])\sqrt{k} \to \lambda t^{-\rho'}$, as $n \to \infty$,

$$M_k(t) := \sqrt{k}\left(\frac{U(Y_{n-[kt],n})}{a(Y_{n-[kt],n})}\,\frac{1}{[kt]}\sum_{i=0}^{[kt]-1}\left\{\log U(Y_{n-i,n}) - \log U(Y_{n-[kt],n})\right\} - l_1\right)$$

converges in distribution to

$$\frac{1}{(1-\gamma_-)\sqrt{1-2\gamma_-}}\,\frac{B(t)}{t} + \frac{\lambda t^{-\rho'}}{(1-\gamma_-)(1-\rho'-\gamma_-)}$$

in $D[t_1, t_2]$, with $B(t)$ standard Brownian motion. Similarly we can give a process version of the other asymptotic quantities involved.

According to Skorohod's theorem (Billingsley, 1971) we can change the sample space and we can replace convergence in distribution by almost sure convergence, not only for the various asymptotically normal quantities but also for the relation

$\hat{k}_0(n)/k_0(n) \to 1$. Then on this sample space, e.g.,

$$M_{k_0}\left(\frac{\hat{k}_0(n)}{k_0(n)}\right) \to \frac{1}{(1-\gamma_-)\sqrt{1-2\gamma_-}}\,B(1) + \frac{\lambda}{(1-\gamma_-)(1-\rho'-\gamma_-)},$$

as $n \to \infty$, almost surely hence in distribution.

Using this device everywhere one sees that indeed the result of Corollary 2.2.1 holds when $k_0(n)$ is replaced by $\hat{k}_0(n)$ throughout. □

Proof of Theorem 2.2.4. By the preceding lemmas, apart from terms $o_p(\frac{1}{\sqrt{k}})$ and $o_p(\tilde{A}(\frac{n}{k}))$,

$$\begin{aligned}
\hat{x}_{0,1}(k) - x_0 &= \hat{b}(\frac{n}{k}) - \hat{a}_1(\frac{n}{k})\,\frac{1}{\hat{\gamma}^-_{n,1}(k)} - U(\infty) \\
&= a(\frac{n}{k})\left[\frac{\hat{b}(\frac{n}{k})-b(\frac{n}{k})}{a(\frac{n}{k})} - \left(\frac{1}{\hat{\gamma}^-_{n,1}(k)} - \frac{1}{\gamma_-}\right)\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - \frac{1}{\gamma_-}\left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1\right)\right. \\
&\quad \left. + \frac{b(\frac{n}{k}) - U(\infty)}{a(\frac{n}{k})} - \frac{1}{\gamma_-}\right] \\
&= a(\frac{n}{k})\left[\frac{\hat{b}(\frac{n}{k})-b(\frac{n}{k})}{a(\frac{n}{k})} + \frac{\hat{\gamma}^-_{n,1}(k)-\gamma_-}{\gamma_-^2} - \frac{1}{\gamma_-}\left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1\right) + \frac{1}{\gamma_-(\gamma_-+\rho)}A(\frac{n}{k})\right].
\end{aligned}$$

Hence

$$\begin{aligned}
\text{as. } E(\hat{x}_{0,1}(k) - x_0)^2 &=: a^2(\frac{n}{k})\left\{\frac{c_7(\gamma_-)}{k} + c_8(\gamma_-,\rho')\tilde{A}^2(\frac{n}{k})\right\} \\
&= c_1^2(\frac{n}{k})^{2\gamma_-}\left\{\frac{c_7}{k} + c_8\tilde{c}_2(\frac{n}{k})^{2\rho'}\right\}.
\end{aligned}$$

□

Proposition 2.4.2. *Under the conditions of Theorem 2.2.4, as $n \to \infty$,*

$$\overline{k}_0(n) \sim \left(\frac{1+2\gamma_-}{-2\rho'-2\gamma_-}\,\frac{\overline{c}_7}{\tilde{c}_2\overline{c}_8}\right)^{\frac{1}{1-2\rho'}} n^{\frac{-2\rho'}{1-2\rho'}} = \overline{g}(\gamma_-,\rho')\,n^{\frac{-2\rho'}{1-2\rho'}}$$

where $\overline{k}_0(n) := \arg\inf_k$ *as.* $E(\hat{x}_{0,1}(k) - \hat{x}_{0,2}(k))^2$, *and* $\tilde{c}_2$ *from* $|\tilde{A}(t)| \sim \sqrt{\tilde{c}_2}t^{\rho'}(t \to \infty)$.

Proof.

$$\hat{x}_{0,1}(k) - \hat{x}_{0,2}(k) = X_{n-k,n} - \frac{\hat{a}_1(\frac{n}{k})}{\hat{\gamma}^-_{n,1}(k)} - \left\{X_{n-k,n} - \frac{\hat{a}_2(\frac{n}{k})}{\hat{\gamma}^-_{n,2}(k)}\right\}$$

$$= a(\frac{n}{k})\left[-\frac{1}{\hat{\gamma}^-_{n,1}(k)}\left(\frac{\hat{a}_1(\frac{n}{k})}{a(\frac{n}{k})} - 1\right) + \frac{1}{\hat{\gamma}^-_{n,2}(k)}\left(\frac{\hat{a}_2(\frac{n}{k})}{a(\frac{n}{k})} - 1\right) - \left(\frac{1}{\hat{\gamma}^-_{n,1}(k)} - \frac{1}{\hat{\gamma}^-_{n,2}(k)}\right)\right].$$

Hence

$$\begin{aligned}\text{as. } E(\hat{x}_{0,1}(k) - \hat{x}_{0,2}(k))^2 &=: a^2(\frac{n}{k})\left\{\frac{\overline{c}_7(\gamma_-)}{k} + \overline{c}_8(\gamma_-,\rho')\tilde{A}^2(\frac{n}{k})\right\} \\ &= c_1^2(\frac{n}{k})^{2\gamma_-}\left\{\frac{\overline{c}_7}{k} + \overline{c}_8\tilde{c}_2(\frac{n}{k})^{2\rho'}\right\}.\end{aligned}$$

□

Proof of Theorem 2.2.5. The result follows similarly as in the proof of Theorem 2.2.3. □

Proof of Theorem 2.2.6. With the obvious changes, following the same arguments as in the proof of Theorem 2.2.2 for positive γ, we know that for the optimal sequence

$$\hat{x}_{n,1}^{+}(k) - x_n \sim \frac{c_1 p_n^{-\gamma_+}}{\gamma_+} \log a_n \left(\hat{\gamma}_{n,1}^{+}(k) - \gamma_+\right)$$

that is,

$$\inf_k \text{as. E}\left(\hat{x}_{n,1}^{+}(k) - x_n\right)^2 \sim \left(\frac{c_1 p_n^{-\gamma_+}}{\gamma_+}\right)^2 \inf_k \text{as. E}\left(\log a_n\right)^2\left(\hat{\gamma}_{n,1}^{+}(k) - \gamma_+\right)^2.$$

Note that from Lemma 2.4.4

$$\text{E}\left(\hat{\gamma}_{n,1}^{+}(k) - \gamma_+\right)^2 \sim \frac{\gamma_+^2}{k} + \frac{\gamma_+^2}{\rho'^2(1-\rho')^2}\tilde{A}^2\left(\frac{n}{k}\right).$$

□

Proposition 2.4.3. *Under the conditions of Theorem 2.2.6, as $n \to \infty$,*

$$\bar{k}_0(n) \sim \left(\frac{(1-\rho')^4}{-2\rho'\tilde{c}_2}\right)^{1/(1-2\rho')} n^{\frac{-2\rho'}{1-2\rho'}} = \bar{l}(\gamma_+,\rho')\, n^{\frac{-2\rho'}{1-2\rho'}}$$

where $\bar{k}_0(n) := \arg\inf_k$ *as.* $E\left(\hat{x}_{n,1}^{+}(k) - \hat{x}_{n,2}^{+}(k)\right)^2$ *and* $\tilde{c}_2$ *from* $|\tilde{A}(t)| \sim \sqrt{\tilde{c}_2}t^{\rho'}\,(t \to \infty)$.

Proof. Following similar arguments as before, we know that for the optimal sequence

$$\hat{x}_{n,1}^{+}(k) - \hat{x}_{n,2}^{+}(k) \sim \frac{c_1 p_n^{-\gamma_+}}{\gamma_+} \log a_n \left(\hat{\gamma}_{n,1}^{+}(k) - \hat{\gamma}_{n,2}^{+}(k)\right)$$

where from Lemma 2.4.4, neglecting terms which are $o_p(\frac{1}{\sqrt{k}})$ or $o_p(\tilde{A}(\frac{n}{k}))$,

$$\hat{\gamma}_{n,1}^{+}(k) - \hat{\gamma}_{n,2}^{+}(k) = \frac{\gamma_+ P}{\sqrt{k}} - \frac{\gamma_+}{4}\frac{Q}{\sqrt{k}} - \frac{\gamma_+}{2\,(1-\rho')^2}\tilde{A}\left(\frac{n}{k}\right)$$

so that, as $n \to \infty$,

$$\mathrm{E}\left(\hat{\gamma}^+_{n,1}(k) - \hat{\gamma}^+_{n,2}(k)\right)^2 \sim \frac{\gamma_+^2}{4k} + \frac{\gamma_+^2}{4(1-\rho')^4}\tilde{A}^2\left(\frac{n}{k}\right).$$

□

Proof of Theorem 2.2.7. The result follows similarly as in the proof of Theorem 2.2.3. □

Proof of Theorems 2.2.8 and 2.2.9. Set

$$M_j := \frac{N_n^{(j)}}{a^j(Y_{n-k,n})} - l_j,$$

for $j = 1, 2, 3$ with l_j as in Lemma 2.4.3. Since we are dealing with $N_n^{(j)}$, $j = 1, 2, 3$, we shall use the second order condition (2.2.9), instead of (2.4.1). Then from ($t \to \infty$)

$$\left(\frac{U(tx) - U(t)}{a(t)}\right)^j = \left(\frac{x^{\gamma_-} - 1}{\gamma_-}\right)^j + j\,A(t)\left(\frac{x^{\gamma_-} - 1}{\gamma_-}\right)^{j-1} H_{\gamma_-,\rho}(x) + o(A(t)),$$

the results of Lemma 2.4.3 follow similarly as in the proof of this lemma, with $\tilde{A}$ replaced by A and ρ' replaced by ρ. The rest of the proof is the same as before. □

Acknowledgments. Ana Ferreira's and Laurens de Haan's research were partly supported by PRAXIS XXI/FEDER.

Chapter 3

Optimal asymptotic estimation of small exceedance probabilities

J. Statist. Plann. Inference 2002, 104, 83-102

Abstract. For the estimation of the probability of a tail set beyond the range of the observations an estimator based on Pareto tails can be used. We calculate the optimum number of upper order statistics used for this estimator, in the mean square error sense. Moreover an adaptive procedure is given to find this optimum in a practical situation.

3.1 Introduction

Let $X_1, X_2, \dots, X_n$ be a sample of n i.i.d. random variables (r.v.), with common (but unknown) distribution function (d.f.) F. The aim is to estimate an extreme exceedance probability, that is, given a 'high' value x one wants to estimate $1-F(x)$.

On the one hand, if x is well into the sample range then it is known that $1-F(x)$ can be estimated via the empirical d.f. (Einmahl, 1990). On the other hand, if x is at the boundary or outside the range of the observations (and then we shall call it a 'high' value) then alternative approaches have to be considered. Empirically this means $P(X > x) \leq 1/n$, and hence we will denote x by x_n and define $p_n = P(X > x_n)$. Therefore in this paper we consider the cases $np_n \to c \geq 0$ where c is a finite real constant, as $n \to \infty$. Note that 'well into the sample range' means $np_n \to \infty$ and in this case the use of the empirical d.f. to estimate p_n is preferred.

For the main conditions we assume that F belongs to the domain of attraction of the generalised extreme value d.f. for some real extreme value index γ (Gnedenko, 1943), shortly $F \in \mathcal{D}_\gamma(GEV)$, $\gamma \in \mathbb{R}$.

Equivalently, one may write it in terms of tail probabilities,

$$\lim_{t\to\infty} t\{1 - F(b(t) + xa(t))\} = 1 - H_\gamma(x) \tag{3.1.1}$$

for all x for which $0 < H_\gamma(x) < 1$, where $a(t) > 0$, $b(t)$ are suitable normalising functions; $H_\gamma(x)$ is the generalised Pareto d.f. given by

$$H_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0, x \in \mathbb{R}^+, \gamma \in \mathbb{R}\,. \tag{3.1.2}$$

For the exceedance probability estimator we use (as in Dekkers et al., 1989; Dijk and de Haan, 1992),

$$\hat{p}_n(k) = \frac{k}{n} \max\left\{0, \left(1 + \hat{\gamma}_n(k) \frac{x_n - \hat{b}(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\right)\right\}^{-1/\hat{\gamma}_n(k)} \tag{3.1.3}$$

where $\hat{a}(t)$, $\hat{b}(t)$ and $\hat{\gamma}_n(k)$ are estimators of $a(t)$, $b(t)$ and γ, respectively, n is the sample size and k is an intermediate sequence i.e., $k = k(n)$ such that $k(n)/n \to 0$ and $k(n) \to \infty$, as $n \to \infty$. A motivation for (3.1.3) may be given as follows. Relations (3.1.1) and (3.1.2) suggest, for large x,

$$1 - F(x) \approx \frac{1}{t}\left\{1 + \gamma \frac{x - b(t)}{a(t)}\right\}^{-1/\gamma}, \quad \text{as} \quad t \to \infty\,. \tag{3.1.4}$$

Then take for t the quantity n/k. Since in practice the condition $\{1 + \hat{\gamma}_n(k)(x_n - \hat{b}(\frac{n}{k}))\,/\hat{a}(\frac{n}{k})\} \geq 0$ may not be fulfilled, the maximum value between this quantity and zero is taken.

Our main result concerns the characterisation of an optimal rate, in the asymptotic sense, for the number of upper order statistics k in (3.1.3). The reasoning is in the same line as in Danielsson et al. (2001) and Draisma et al. (1999) on extreme value index estimation, and Ferreira et al. (1999) on endpoint and high quantiles estimation. We obtain an optimal rate of order n to some positive power (cf. Theorem 3.2.1). However this asymptotic result might not be adequate for practical applications. Firstly, it contains second order parameters that can not be estimated with sufficient accuracy. Secondly, even in cases when c_2 and ρ are known, the asymptotic optimal result may be far from the real optimum. Still, our optimal rate may be applied in the adapted bootstrap procedure, as suggested in the aforementioned papers, to optimise the performance of $\hat{p}_n(k)$.

In this paper we focus on models with a power expansion as in (3.2.6) below, therefore excluding the case $\gamma = 0$. For the gamma estimator we consider, the case $\gamma = 0$ requires slowly varying - type conditions, regarded as beyond the scope of this work.

An alternative approach to our problem is given in the work by Hall and Weissman (1997). However they concentrate only on models with positive γ. Moreover, our result allows smaller exceedance probabilities, therefore covering more interesting situations. Other related work on tail estimation includes Davis and Resnick (1984) and Smith (1987).

The paper is organised as follows. Section 3.2 deals with our main result, namely Theorem 3.2.1. In Section 3.3 is a simulation study, including results from the adaptive bootstrap. In Section 3.4 are the proofs. A short explanation of the adaptive bootstrap method is postponed to the Appendix.

Finally some notation. Let a_n and b_n be two sequences, then $a_n \sim b_n$ means $\lim_{n\to\infty} a_n/b_n = 1$. Let $\gamma_- = \min(\gamma, 0)$ and $\gamma_+ = \max(\gamma, 0)$. Denote the right endpoint of a d.f. F by x_0 i.e., $x_0 = \sup\{x : F(x) < 1\}$. Under the assumption $F \in \mathcal{D}_\gamma(GEV)$, x_0 is finite if $\gamma < 0$ and infinite if $\gamma > 0$. For a real function f write $f^{\leftarrow}$ for its generalised inverse function.

3.2 Asymptotic optimal rate

On the basis of an i.i.d. sample $X_1, X_2, \dots, X_n$, from a d.f. F, the estimators used in (3.1.3) are (as in Dekkers et al., 1989; Dijk and de Haan, 1992)

$$\hat{\gamma}_n(k) = M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1} \tag{3.2.1}$$

and

$$\hat{a}\left(\frac{n}{k}\right) = X_{n-k,n} M_n^{(1)} (1 - \hat{\gamma}_n^-(k)), \tag{3.2.2}$$

where

$$M_n^{(j)} = \frac{1}{k}\sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n})^j, \qquad j = 1, 2\,, \tag{3.2.3}$$

and

$$\hat{\gamma}_n^-(k) = 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1} = \hat{\gamma}_n(k) - M_n^{(1)}; \tag{3.2.4}$$

and

$$\hat{b}\left(\frac{n}{k}\right) = X_{n-k,n}. \tag{3.2.5}$$

Our main result is valid for the following model. Let $U = (1/(1-F))^{\leftarrow}$. Then, as $t \to \infty$, we assume

$$U(t) = \begin{cases} c_0 t^\gamma \left(1 + c_2' t^{\rho'} + o(t^{\rho'})\right) & ,\gamma > 0 \text{ and } \rho' < 0 \\ c_0 + c_0 c_1 t^\gamma \left(1 + c_2' t^{\rho'} + o(t^{\rho'})\right) & ,-1/2 < \gamma < 0 \text{ and } \rho' < 0 \end{cases} \tag{3.2.6}$$

where $c_0 > 0$, $c_1 < 0$ and $c_2' \neq 0$ are real constants (also we shall need to impose $c_2' - c_1/2 \neq 0$ in the $\gamma = \rho'$ case). Then

$$\log U(t) = \begin{cases} \gamma \log t + \log c_0 + c_2 t^\rho + o(t^\rho) & ,\gamma > 0 \text{ and } \rho < 0 \\ \log c_0 + c_1 t^\gamma \left(1 + c_2 t^\rho + o(t^\rho)\right) & ,-1/2 < \gamma \le \rho < 0 \end{cases} \tag{3.2.7}$$

where γ, c_0 and c_1 are the same as before but ρ and $c_2 \neq 0$ may differ from ρ' and c_2', respectively. In fact if $\gamma > 0$ and $\rho' < 0$, or $-1/2 < \gamma < \rho' < 0$, (3.2.6) and (3.2.7) are equivalent, with $\rho' = \rho$ and $c_2' = c_2$. For the other cases note that (3.2.6) with $\rho' \leq \gamma < 0$ implies, as $t \to \infty$,

$$\log U(t) = \log c_0 + c_1 t^{\gamma} \left(1 + c_2 t^{\gamma} + o(t^{\gamma})\right) \tag{3.2.8}$$

(where $c_2 = c_2' - c_1/2$ if $\rho' = \gamma$ and $c_2 = -c_1/2$ if $\rho' < \gamma$). But if (3.2.8) holds then one does not get back (3.2.6) with $\rho' \leq \gamma$ but only the case $-1/2 < \gamma = \rho' < 0$ with $c_2' = c_2 + c_1/2$.

The motivation for such a model comes from the second order regular variation condition given in e.g. de Haan and Stadtmüller (1996), together with the requirement that their second order auxiliary function $A(t)$ must be asymptotic to $ct^{\rho'}$, as $t \to \infty$, where $c \neq 0$ and $\rho' < 0$ are real constants. Moreover, in general a second order representation of the type (3.2.6) holds for U iff a second order representation in terms of the underlying d.f. F, of the type $1 - F(x) = dx^{-1/\gamma}\{1 + \delta x^{\alpha} + o(x^{\alpha})\}$ if $\gamma > 0$; $1 - F(x_0 - x^{-1}) = dx^{1/\gamma}\{1 + \delta x^{\alpha} + o(x^{\alpha})\}$ if $\gamma < 0$, as $x \to \infty$ ($d > 0$, $\delta \neq 0$, $\alpha < 0$), holds. These include the well-known Hall model (see e.g. Hall and Weissman, 1997).

We restrict attention to $\gamma > -1/2$, since otherwise the extrapolation should be based on extreme rather than intermediate order statistics (cf. Aarssen and de Haan, 1994).

Let $r_\gamma(k,n) = 1/\log(k/(np_n))$ if $\gamma > 0$, $(k/(np_n))^{\gamma}$ if $\gamma < 0$ and $C = c_2\rho^2/\gamma$ if $\gamma > 0$, $c_2(\gamma + \rho)\rho/\gamma$ if $\gamma < 0$.

Lemma 3.2.1. *Assume* (3.2.6) *and* $np_n \to$ *constant (finite,* ≥ 0*), as* $n \to \infty$*. Let* $k(n)$ *be an intermediate sequence.*

(i) If $(n/k)^{\rho}\sqrt{k} \to \lambda \geq 0$ *and* $r_\gamma(k,n)\sqrt{k} \to \infty$*, then*

$$R_{\gamma,\rho}(k,n)\left(\frac{\hat{p}_n(k)}{p_n} - 1\right) := r_\gamma(k,n)\sqrt{k}\left(\frac{\hat{p}_n(k)}{p_n} - 1\right) \tag{3.2.9}$$

converges in distribution to a normal r.v., N *say, with mean* $\lambda\, C\, bias_{\gamma,\rho}$ *where*

$$bias_{\gamma,\rho} = \begin{cases} \frac{-\gamma\rho+\gamma+\rho}{\gamma\rho(1-\rho)^2} & \gamma > 0 \\ \frac{(1-\gamma)\rho}{\gamma^2(1-\gamma-\rho)(\gamma+\rho)(1-2\gamma-\rho)} & , -1/2 < \gamma < \rho'(=\rho) < 0 \\ \frac{1-3\gamma^2}{\gamma(1-\gamma)(1-2\gamma)(1-3\gamma)} & , \gamma(=\rho) > \rho' \text{ and } -1/2 < \gamma < 0 \\ \frac{c_1(1-4\gamma+\gamma^2+6\gamma^3)+2c_2(1-\gamma)^2}{4c_2\gamma^2(1-\gamma)(1-2\gamma)(1-3\gamma)} & , -1/2 < \gamma = \rho'(=\rho) < 0, \end{cases} \tag{3.2.10}$$

and variance

$$var_\gamma = \begin{cases} 1 + 1/\gamma^2 & , \gamma > 0 \\ \frac{(1-\gamma)^2(1-3\gamma+4\gamma^2)}{\gamma^4(1-2\gamma)(1-3\gamma)(1-4\gamma)} & , -1/2 < \gamma < 0. \end{cases} \tag{3.2.11}$$

(ii) if $(n/k)^{\rho}\sqrt{k} \to \infty$ *and* $r_\gamma(k,n)(n/k)^{-\rho} \to \infty$*, then*

$$R_{\gamma,\rho}(k,n)\left(\frac{\hat{p}_n(k)}{p_n} - 1\right) := r_\gamma(k,n)\left(\frac{n}{k}\right)^{-\rho}\left(\frac{\hat{p}_n(k)}{p_n} - 1\right) \qquad (3.2.12)$$

converges in probability to $N = C\, bias_{\gamma,\rho}$*.*

Remark 3.2.1. Note that: (i) $(n/k)^{\rho}\sqrt{k} \to \lambda \geq 0$ and $r_\gamma(k,n)\sqrt{k} \to \infty$ imply $r_\gamma(k,n)(n/k)^{-\rho} \to \infty$; (ii) $(n/k)^{\rho}\sqrt{k} \to \infty$ and $r_\gamma(k,n)(n/k)^{-\rho} \to \infty$ imply $r_\gamma(k,n)\sqrt{k} \to \infty$.

In the sequel we shall restrict ourselves to intermediate sequences $k(n)$ for which $(n/k)^{\rho}\sqrt{k}$ converges to a finite or infinite constant.

From Lemma 3.2.1 we have that the best rate of convergence of $(\hat{p}_n(k)/p_n - 1)$ is attained in the case (i) with $\lambda > 0$. To see this just note the following. Let $\gamma < 0$. If $\lambda > 0$, k is of order $n^{-2\rho/(1-2\rho)}$ (say k_0) and the rate of convergence is $k_0^{\gamma+1/2}/(np_n)^{\gamma}$. Now let k_1 be any other sequence such that $(n/k_1)^{\rho}\sqrt{k_1} \to 0$, which means $k_1 = o(k_0)$. Then, the ratio of the correspondent rates of convergence is $(k_0/k_1)^{\gamma+1/2}$, which goes to infinite since $\gamma > -1/2$. Next take k_2 such that $(n/k_2)^{\rho}\sqrt{k_2} \to \infty$. Then, the ratio of the rates equals $k_0^{\gamma+1/2}/(k_2^{\gamma+\rho}n^{-\rho}) = (k_0/k_2)^{\gamma+\rho}k_0^{1/2-\rho}/n^{-\rho}$, which again converges to infinite. The case $\gamma > 0$ is similar.

Also from Lemma 3.2.1, we have that $(\hat{p}_n(k)/p_n - 1)$ is asymptotic to $R^{-1}_{\gamma,\rho}(k,n)N$, where N is a r.v. Our goal is to find the optimal sequence $k_0(n)$ such that the mean square error (mse, say) of the approximating r.v. $R^{-1}_{\gamma,\rho}(k,n)N$ is minimal, that is

$$k_0(n) = \arg\inf_k\; r_\gamma^{-2}(k,n)\left(\frac{var_\gamma}{k} + bias^2_{\gamma,\rho}C^2\left(\frac{n}{k}\right)^{2\rho}\right). \qquad (3.2.13)$$

Note that from the previous considerations it follows that this optimal rate is, asymptotically, proportional to n to some power between zero and one.

Theorem 3.2.1. *Suppose* (3.2.6) *and, as* $n \to \infty$, $np_n \to$ *constant (finite,* ≥ 0*) and*

$$\log x_n = o\left(n^{\frac{-\rho}{1-2\rho}}\right),\ \text{if } \gamma > 0\,; \quad (x_0 - x_n)^{-1} = o\left(n^{-\frac{\gamma+\rho}{1-2\rho}}\right),\ \text{if } \gamma < 0. \qquad (3.2.14)$$

Then, $k_0(n)$ *(cf.* (3.2.13)*) is asymptotically equal to*

$$\begin{cases} \left(\frac{var_\gamma}{bias^2_{\gamma,\rho}}\frac{\gamma^2}{-2c_2^2\rho^5}\right)^{\frac{1}{1-2\rho}} n^{\frac{-2\rho}{1-2\rho}} & ,\ \text{if } \gamma > 0 \\ \\ \left(\frac{var_\gamma}{bias^2_{\gamma,\rho}}\frac{(1+2\gamma)\gamma^2}{-2c_2^2\rho^2(\gamma+\rho)^3}\right)^{\frac{1}{1-2\rho}} n^{\frac{-2\rho}{1-2\rho}} & ,\ \text{if } -1/2 < \gamma < 0. \end{cases} \qquad (3.2.15)$$

Remark 3.2.2. Our theorem is valid, up to changes in constants, for any estimators of $(\hat{\gamma}, \hat{a}, \hat{b})$ satisfying the following two properties: (i) if $(n/k)^{\rho}\sqrt{k} \to \lambda \geq 0$, then

$\sqrt{k}\left(\hat{\gamma}-\gamma, \hat{a}/a-1, (\hat{b}-b)/a\right)$ converges in distribution to a $3-d$ normal r.v., with specified parameters; (ii) if $(n/k)^{\rho}\sqrt{k} \to \infty$, then $(n/k)^{-\rho}\left(\hat{\gamma}-\gamma, \hat{a}/a-1, (\hat{b}-b)/a\right)$ converges to a known $3-d$ (non-zero) constant vector. Therefore, the moment estimator (3.2.1) is, in fact, one possible estimator we used (cf. Lemma 3.4.2). For instance, most of these conditions have been established for the maximum likelihood estimator (cf. Smith, 1987). Or, if one wants to restrict to the particular case of $\gamma > 0$, the Hill (1975) estimator could be another possibility.

Remark 3.2.3. Note that $k_0(n)$ in the theorem does not depend on x_n.

Remark 3.2.4. Concerning the conditions (3.2.14) in Theorem 3.2.1:

(1) Since ρ and γ are unknown (3.2.14) may alternatively be written as

$$\log x_n = o\left(n^{\epsilon}\right), \epsilon > 0, \text{ if } \gamma > 0; \quad \left(x_0 - x_n\right)^{-1} = o\left(n^{\epsilon}\right), \epsilon > 0, \text{ if } \gamma < 0. \tag{3.2.16}$$

(2) In fact (3.2.14) is only required when $np_n \to 0$ as $n \to \infty$.

(3) Conditions (3.2.14) are equivalent to $\log p_n / n^{-\rho/(1-2\rho)} \to 0$, if $\gamma > 0$ and $p_n^{\gamma}/n^{-(\gamma+\rho)/(1-2\rho)} \to 0$ if $\gamma < 0$. Or, equivalently,

$$\lim_{n\to\infty} \frac{\log(k_0(n)/(np_n))}{\sqrt{k_0(n)}} = 0 \qquad \text{if } \gamma > 0 \tag{3.2.17}$$

$$\lim_{n\to\infty} \frac{(k_0(n)/(np_n))^{-\gamma}}{\sqrt{k_0(n)}} = 0 \qquad \text{if } \gamma < 0. \tag{3.2.18}$$

Note that if $\gamma > 0$, (3.2.17) implies (3.2.18); conversely if $\gamma < 0$, (3.2.18) implies (3.2.17). Therefore our optimal sequence satisfy condition (2.10) in Dijk and de Haan (1992).

Example 3.2.1. Generalised extreme value d.f. Let $GEV_{\gamma}(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}$, $1+\gamma x > 0$, $\gamma \in \mathbb{R}$. Then $U(t) = ((-\log(1-1/t))^{-\gamma} - 1)/\gamma$, $t > 1$. Hence, as $t \to \infty$, one gets $U(t) = -1/\gamma + t^{\gamma}/\gamma(1 - \gamma t^{-1}/2 + o(t^{-1}))$, $-1/2 < \gamma < 0$; $t^{\gamma}/\gamma(1 - t^{-\gamma} + o(t^{-\gamma}))$, $0 < \gamma < 1$; $t(1 - 3t^{-1}/2 + o(t^{-1}))$, $\gamma = 1$; $t^{\gamma}/\gamma(1 - \gamma t^{-1}/2 + o(t^{-1}))$, $\gamma > 1$.

Example 3.2.2. Reversed Burr d.f. A r.v. Y has Burr d.f. with parameters β, λ and τ if $F_Y(y) = 1 - \beta^{\lambda}/(\beta + y^{\tau})^{\lambda}$, $y > 0$, $\beta, \lambda, \tau > 0$. We shall denote the d.f. of $X = -Y^{-1}$ by Reversed Burr, say $RB_{\beta,\lambda,\tau}$, which is given by $F_X(x) = 1 - \beta^{\lambda}/(\beta + (-x)^{-\tau})^{\lambda}$, $x < 0 = x_0$, $\beta, \lambda, \tau > 0$. This r.v. has been referred to in financial applications. Note that in order to properly use this model with the suggested methods it must be shifted by a positive constant, say a, so that $x_0 = a > 0$. Therefore in this case $U(t) = a - \beta^{-1/\tau}(t^{1/\lambda} - 1)^{-1/\tau}$, $t > 1$. Hence, as $t \to \infty$, $U(t) = a - \beta^{-1/\tau} t^{-1/(\lambda\tau)}(1 + t^{-1/\lambda}/\tau + o(t^{-1/\lambda}))$. The extreme value parameters are $\gamma = -1/(\lambda\tau)$ and $\rho' = -1/\lambda$.

3.2.1 Consequences of applying a positive shift in our model.

Our estimators require strictly positive data. It is straightforward from its definition, that the moment estimator (3.2.1) is not shift invariant. We shall now discuss the effect of a positive shift, independent of the sample size, has on our results. Namely, on the asymptotic behaviour of the estimator in general, and on the minimal mse of the approximating r.v. in particular. E.g. in Rosen and Weissman (1996) is a simulation study which includes the moment estimator and the shift was taken into account.

Let $a > 0$ be a positive constant. Denote the shifted r.v. by $X^* = X + a$, F^* the respective d.f. and its right-endpoint by $x_0^* = x_0 + a$. Then $F^*(x) = F(x-a)$, and so $U^*(t) = a + U(t)$. We discuss the cases $\gamma > 0$ and $\gamma < 0$ separately.

In the case $\gamma > 0$, after a shift a in (3.2.6), c_0 and γ remain the same but the second order parameters may change. More specifically for (3.2.7), it remains the same in case $\rho > -\gamma$, whereas in the other cases it changes to

$$\log U^*(t) = \begin{cases} \gamma \log t + \log c_0 + \frac{a}{c_0} t^{-\gamma} + o(t^{-\gamma}) & , \rho < -\gamma \\ \gamma \log t + \log c_0 + \left(\frac{a}{c_0} + c_2\right) t^{-\gamma} + o(t^{-\gamma}) & , \rho = -\gamma \text{ and } \frac{a}{c_0} + c_2 \neq 0. \end{cases}$$

For simplicity, in the following we will assume $a/c_0 \neq -c_2$ when $\rho = -\gamma$. Otherwise, it may happen or not that such an expansion still holds for U^*, depending on whether U admits an expansion with higher order terms. For instance, such is true for GEV d.f. (cf. Example 3.2.3).

Hence, when $\gamma > 0$ the second order parameter ρ changes to $\rho^* = \max(-\gamma, \rho)$. Consequently, if $\rho < -\gamma$ the optimal rate of convergence becomes worse, since it is mainly governed by $\sqrt{k_0}$, which becomes asymptotically proportional to $n^{\gamma/(1+2\gamma)}$. This is more serious when γ approaches zero. Otherwise, if $\rho \geq -\gamma$, the rate of convergence does not change.

In what concerns the minimal mse of the approximate limit distribution, it approaches zero with a rate mainly governed by k_0^{-1}. Hence, the same considerations as for the optimal rate apply. That is, the rate gets worse when $\rho < -\gamma$. The minimum mse is exactly the same if $\rho > -\gamma$, since the model is the same. If $\rho = -\gamma$ the order is the same. In this case, $k_0^* \sim (c_2/(a/c_0 + c_2))^{2/(1+2\gamma)}\, k_0$ and the ratio of the minimum mse's, minmse$(k_0)/$ minmse(k_0^*) say, is asymptotic to $(c_2/(a/c_0 + c_2))^{2/(1+2\gamma)}$. This is larger than one when $a \in (0, -2c_0c_2)$ and $c_2 < 0$.

Now we turn to the case $\gamma < 0$. After a shift, (3.2.7) changes to $\log U^*(t) = \log(c_0 + a) + c_0\, c_1/(c_0 + a)\, t^\gamma\, (1 + c_2^* t^\rho + o(t^\rho))$ where

$$c_2^* = \begin{cases} c_2 & , -1/2 < \gamma < \rho'(= \rho) < 0 \\ -\frac{c_0 c_1}{2(c_0+a)} & , \gamma(= \rho) > \rho' \text{ and } -1/2 < \gamma < 0 \\ c_2 + \frac{a c_1}{2(c_0+a)} & , \rho' = \gamma \text{ and } c_2 + \frac{a c_1}{2(c_0+a)} \neq 0. \end{cases}$$

Similarly as before we only consider the cases $c_2 \neq -ac_1/(2(c_0 + a))$ if $\rho' = \gamma$.

Therefore γ and the second order parameter ρ do not change. Consequently the rate of convergence and the rate of minimal mse remain the same.

If $\gamma < \rho'$ the minimal mse is exactly the same, since c_2 does not change. If $\gamma(=\rho) > \rho'$, then $k_0^* \sim (c_0/(c_0+a))^{-2/(1-2\gamma)}\, k_0$. After a few simplifications we have that the ratio minmse$(k_0)/$ minmse(k_0^*) is asymptotic to $(c_0/(c_0+a))^{-2(1+2\gamma)/(1-2\gamma)}$. This is always larger than one. For the other situation, $\gamma = \rho'$, we do not analyse it here since it would be rather technical and considered not so relevant for the present study (note the dependence of $bias_{\gamma,\rho}$ on c_1 and c_2).

Notice that in any case the variance remains the same since γ does not change.

Example 3.2.3. Consider the GEV d.f., with $0 < \gamma < 1$, and a such that $a/c_0 + c_2 = a\gamma - 1 = 0$. Then, expanding U up to third order terms one gets $U^*(t) = U(t) + a = t^\gamma/\gamma(1 - \gamma t^{-1}/2 + o(t^{-1}))$. That is, the second order parameter ρ changes to $\rho^* = -1$, implying a better rate of convergence. Therefore, in general, one has that for the cases where such expansion exists, the second order parameter becomes smaller. This means an improvement in the optimal rate of convergence, and in the minimum mse as well.

3.3 Simulation results

For the simulations we use the distribution families presented in the examples in Section 3.2, generalised extreme value d.f. and Reversed Burr d.f. Specifically $RB_{4,4,2}$, $GEV_{-.1}$, $GEV_{.5}$ and GEV_1. We opted by fixing two exceedance probabilities to be estimated: $p_n = 1/(n\log n)$ and $p_n = 1/n$. Then to each d.f. they correspond to a different quantile given by,

$$x_n = \frac{(-\log(1-p_n))^{-\gamma} - 1}{\gamma} + a \quad \text{and} \quad x_n = -\left(\frac{\beta}{p_n^{1/\lambda}} - \beta\right)^{-1/\tau} + a\,,$$

for GEV_γ and $RB_{\beta,\lambda,\tau}$, respectively. The parameter a stands for a positive shift in the data set, in order the sample be constituted of positive values. In our simulation study, for each d.f. a is the minimum positive integer such that all generated values are positive. Three methods to estimate the exceedance probability were considered: *1)* (3.1.3) with k determined by the adaptive bootstrap method resumed in the Appendix; *2)* (3.1.3) with k equal to the intermediate sequence $\sqrt{n}$; *3)* empirical d.f., by calculating the number of values in the sample greater than the respective quantile. Of course the last approach is only used in the case $p_n = 1/n$.

The simulation results are resumed in Tables 3.1, 3.2, 3.3 and 3.4, and Figures 3.1, 3.2 and 3.3. They are based on 200 independent replications of samples of size $n = 10\,000$ and $n = 2\,000$. For each n, the generated samples are based on the same pseudo-random sample.

In the tables are the sample bias, i.e. sample mean minus true value, root mean square error (rmse) and the number of simulations with valid solutions in 200 (n. solut.). The second line of values for each d.f. (in italic), corresponds to these descriptive measures calculated only on those samples where all procedures yielded a valid estimate. We classify as a non-valid solution $\hat{p}_n(k) = 0$ or $\hat{\gamma}_n(k) < -1/2$,

	1 - bootstrap bias ($\times 10^4$)	rootmse ($\times 10^4$)	n. of solut.	*2 - $k(n) = \sqrt{n}$* bias ($\times 10^4$)	rootmse ($\times 10^4$)	n. of solut.
$GEV_{-.1}$ (a=4)	.108	.259	162	.094	.243	172
	.123	*.268*	*150*	*.086*	*.229*	*150*
$RB_{4,4,2}$ (a=1375)	.061	.202	70	.076	.254	147
	.066	*.207*	*67*	*.142*	*.308*	*67*
$GEV_{.5}$ (a=2)	-.017	.046	191	.008	.110	200
	-.017	*.046*	*191*	*-.001*	*.085*	*191*
GEV_1 (a=1)	.030	.061	193	.006	.094	200
	.030	*.061*	*193*	*.006*	*.095*	*193*

Table 3.1: Simulation results based on samples of size $n = 10\,000$ and 200 independent repetitions. Estimation of $p_{10\,000} = 1/(10\,000 \log 10\,000) \approx .108 \times 10^{-4}$.

when using the estimation methods *1)* or *2)*. Moreover in the bootstrap method (for the technical details, namely the explanation of the quantities $k_0^*(n_1)$ and $k_0^*(n_2)$, see the Appendix) sometimes happens that $k_0^*(n_1)$ is less or equal to $k_0^*(n_2)$, or that the intermediate consistent estimate of p_n is equal to zero, or that $\hat{k}_0(n)$ is equal to 0,1 or it is greater than the sample size; all of these also considered non-valid simulations.

When estimating $p_n = 1/n$, the empirical d.f. always yields a valid solution. But, from Tables 3.2 and 3.4 the obtained estimates have the largest rmse. Hence, our small simulation study indicates that the estimates from the empirical d.f. are the poorest, when any of the other two alternative methods is giving a solution.

For all simulations where $\gamma < 0$, and when comparing the methods *1)* and *2)*, none of each is systematically better than the other, either if looking at the rmse or at the bias, or from the histograms and boxplots. Nonetheless the bootstrap procedure fails more often. The worst cases happen with $RB_{4,4,2}$. In particular when $n = 2000$ we only got approximately 25% of valid solutions.

When $\gamma > 0$, it is clear that the bootstrap procedure is giving the most accurate results. The corresponding estimates always have the smallest rmse. The histograms and boxplots confirm the smaller variance of these estimates.

Hence we conclude that the bootstrap procedure is giving reasonable estimates. In this short simulation study these are among the best. The case $\gamma < 0$ needs further work.

3.4 Proofs

Let $Y_1, Y_2, \ldots$ be i.i.d. r.v. with d.f. $1 - 1/y$, $y > 1$. Then $U(Y_1)$, $U(Y_2)$, $\ldots$ are i.i.d. F.

As mentioned in Section 3.2, our model (3.2.6) implies (3.2.7) which in turn

	1 - bootstrap			*2 - $k(n)=\sqrt{n}$*			*3 - empirical d.f.*		
	bias ($\times 10^3$)	rootmse ($\times 10^3$)	n. of solut.	bias ($\times 10^3$)	rootmse ($\times 10^3$)	n. of solut.	bias ($\times 10^3$)	rootmse ($\times 10^3$)	n. of solut.
$GEV_{-.1}$	-.007	.068	158	.005	.070	192	.008	.096	200
$(a{=}4)$	*-.005*	*.068*	*155*	*.003*	*.062*	*155*	*.010*	*.094*	*155*
$RB_{4,4,2}$	-.006	.086	91	.000	.073	183	.008	.096	200
$(a{=}1375)$	*-.006*	*.086*	*91*	*.030*	*.080*	*91*	*.027*	*.110*	*91*
$GEV_{.5}$	.001	.029	177	-.003	.056	200	.008	.096	200
$(a{=}2)$	*.001*	*.029*	*177*	*-.009*	*.049*	*177*	*.005*	*.091*	*177*
GEV_1	.016	.039	192	-.002	.052	200	.008	.096	200
$(a{=}1)$	*.016*	*.039*	*192*	*-.002*	*.052*	*192*	*.009*	*.097*	*192*

Table 3.2: Simulation results based on samples of size $n = 10\,000$ and 200 independent repetitions. Estimation of $p_{10\,000} = 1/10\,000 = .1 \times 10^{-3}$.

	1 - bootstrap			*2 - $k(n)=\sqrt{n}$*		
	bias ($\times 10^4$)	rootmse ($\times 10^4$)	n. of solut.	bias ($\times 10^4$)	rootmse ($\times 10^4$)	n. of solut.
$GEV_{-.1}(a{=}4)$	.291	1.007	133	.402	1.127	162
	.355	*1.037*	*123*	*.393*	*1.135*	*123*
$RB_{4,4,2}(a{=}1375)$	.619	1.402	47	.562	1.280	119
	.619	*1.402*	*47*	*.842*	*1.417*	*47*
$GEV_{.5}(a{=}2)$	-.052	.378	179	.024	.666	199
	-.051	*.380*	*178*	*.022*	*.663*	*178*
$GEV_1(a{=}1)$	.244	.424	188	.027	.598	200
	.244	*.424*	*188*	*.033*	*.592*	*188*

Table 3.3: Simulation results based on samples of size $n = 2\,000$ and 200 independent repetitions. Estimation of $p_{2\,000} = 1/(2\,000 \log 2\,000) \approx .658 \times 10^{-4}$.

	1 - bootstrap			*2 - $k(n)=\sqrt{n}$*			*3 - empirical d.f.*		
	bias ($\times 10^3$)	rootmse ($\times 10^3$)	n. of solut.	bias ($\times 10^3$)	rootmse ($\times 10^3$)	n. of solut.	bias ($\times 10^3$)	rootmse ($\times 10^3$)	n. of solut.
$GEV_{-.1}$	-.001	.345	147	-.023	.317	189	.032	.506	200
$(a{=}4)$	*-.001*	*.345*	*147*	*.005*	*.313*	*147*	*.068*	*.513*	*147*
$RB_{4,4,2}$	-.053	.343	56	-.015	.336	169	.032	.506	200
$(a{=}1375)$	*-.053*	*.343*	*56*	*.161*	*.345*	*56*	*.214*	*.579*	*56*
$GEV_{.5}$	.040	.193	179	-.044	.295	200	.032	.506	200
$(a{=}2)$	*.040*	*.193*	*179*	*-.036*	*.292*	*179*	*.042*	*.501*	*179*
GEV_1	.102	.207	178	-.030	.281	200	.032	.506	200
$(a{=}1)$	*.102*	*.207*	*178*	*-.027*	*.282*	*178*	*.039*	*.496*	*178*

Table 3.4: Simulation results based on samples of size $n = 2\,000$ and 200 independent repetitions. Estimation of $p_{2\,000} = 1/2\,000 = .5 \times 10^{-3}$.

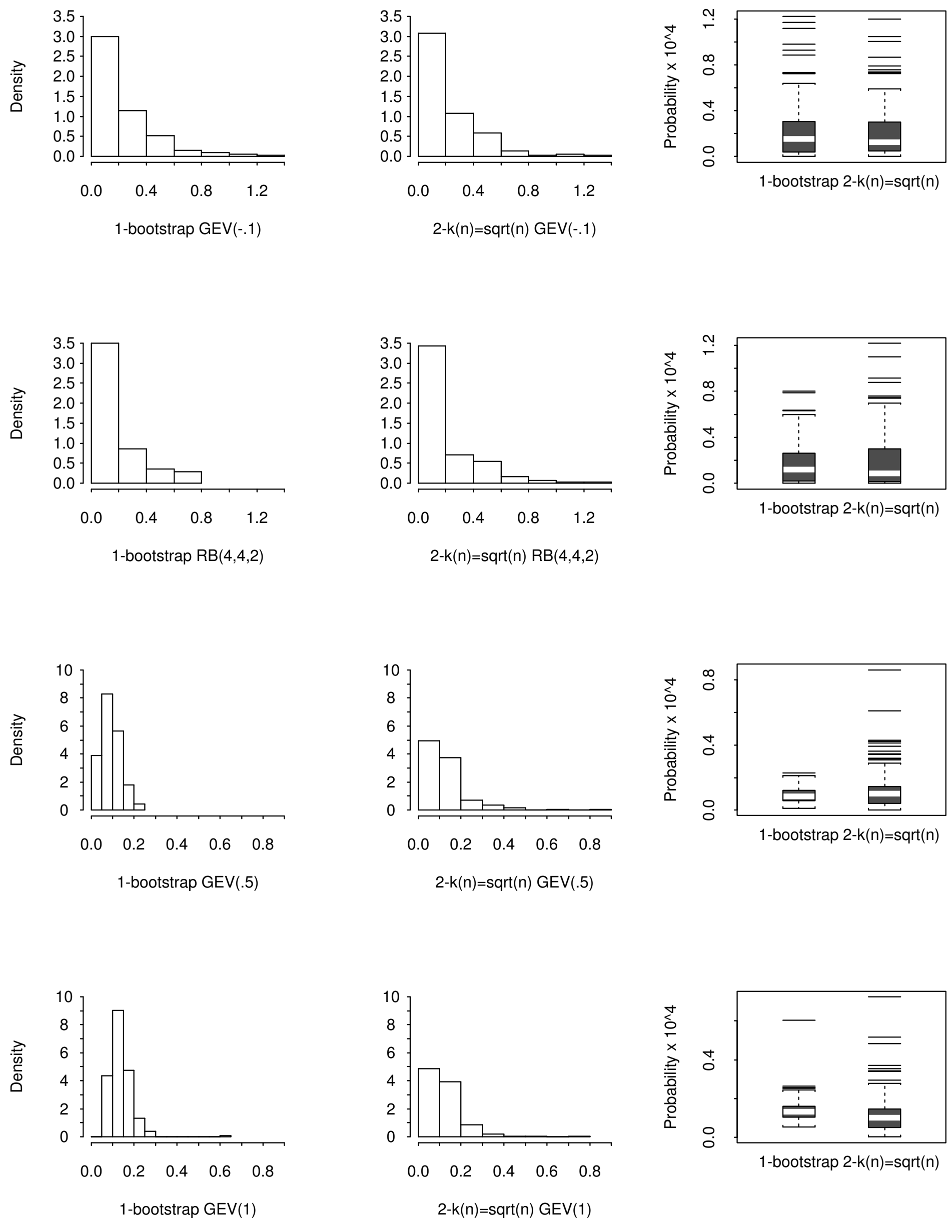


Figure 3.1: Histograms, with area equal to 1, and boxplots of the estimates of $p_{10\,000} \approx .108 \times 10^{-4}$.

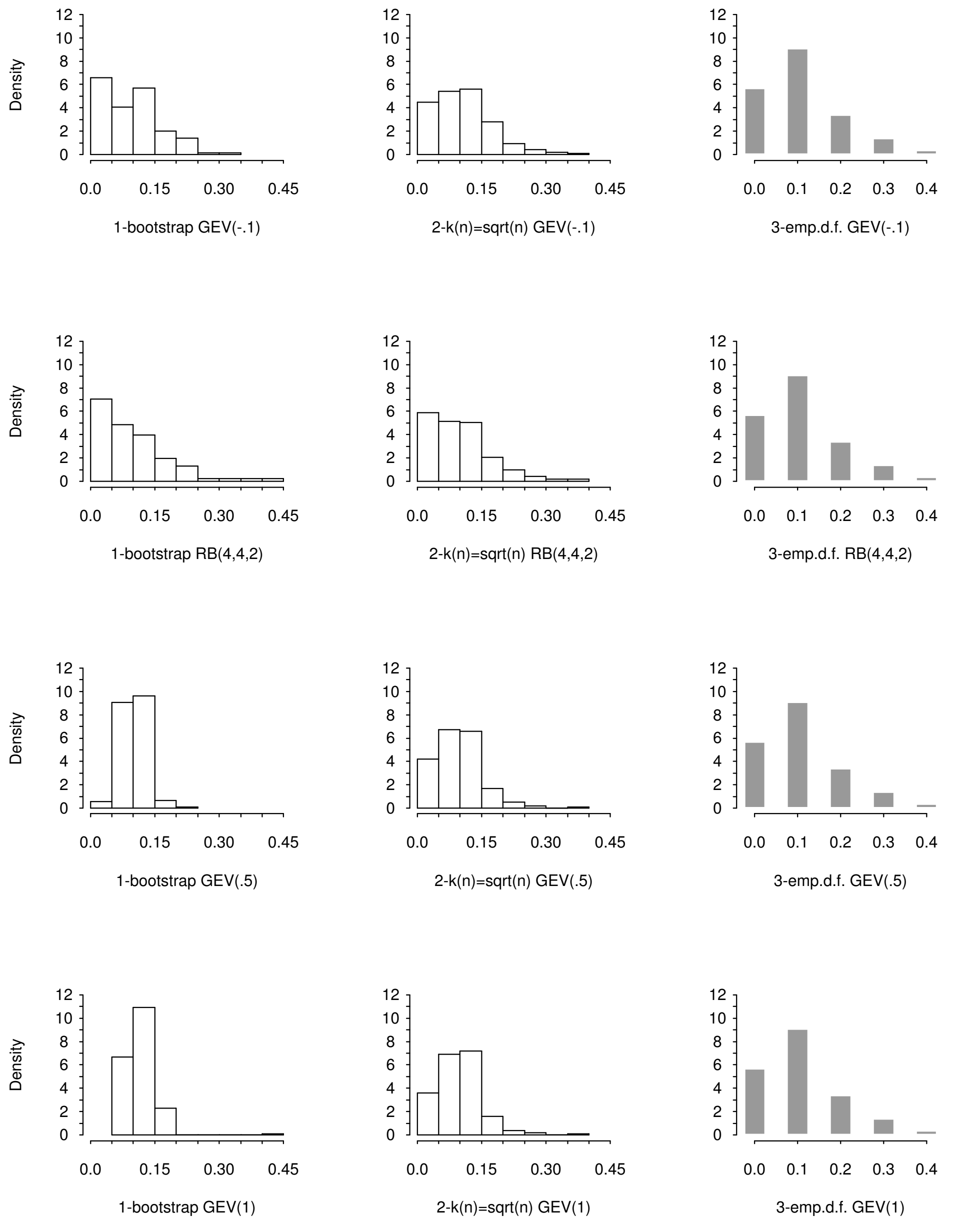


Figure 3.2: Histograms, with area equal to 1, of the estimates of $p_{10\,000} = .1 \times 10^{-3}$.

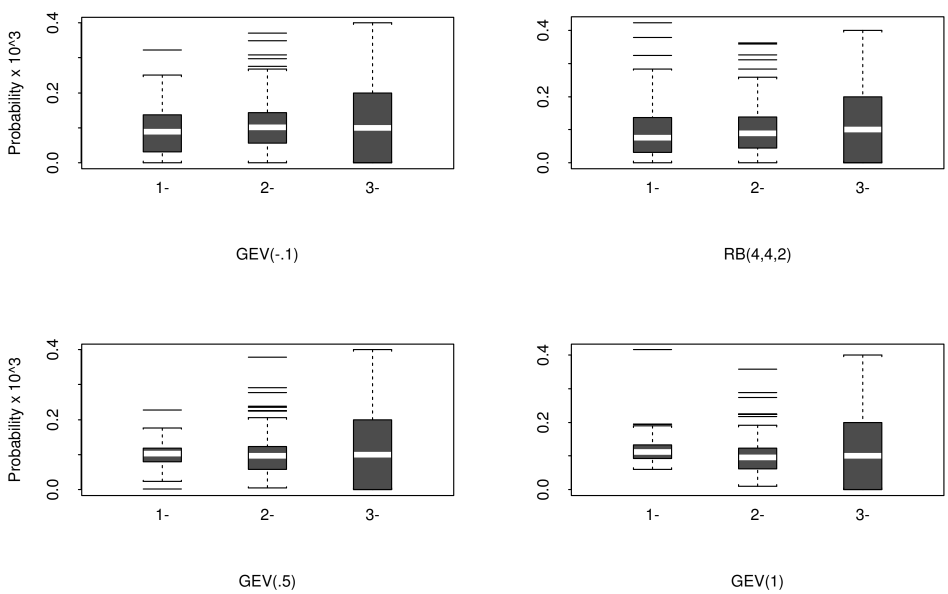


Figure 3.3: Boxplots of the estimates of $p_{10\,000} = .1 \times 10^{-3}$; 1-bootstrap; 2-$k(n) = \sqrt{n}$; 3-empirical d.f.

verifies the second order condition considered in e.g. Ferreira et al. (1999),

$$\lim_{t\to\infty} \frac{\frac{\log U(tx)-\log U(t)}{a(t)/U(t)} - \frac{x^{\gamma_-}-1}{\gamma_-}}{A(t)} = \frac{1}{\rho}\left[\frac{x^{\gamma_-+\rho}-1}{\gamma_-+\rho} - \frac{x^{\gamma_-}-1}{\gamma_-}\right] \tag{3.4.1}$$

where $a(t) > 0$ and $A(t) \to 0$ (as $t \to \infty$) are real functions. Note, however, that here we do not exclude the case $\gamma = \rho'$ (although in this case one still has to impose $c_2 = c_2' - c_1/2 \neq 0$ in (3.2.6)). Simple examples of the auxiliary functions are

$$a(t) = \begin{cases} \gamma c_0 t^\gamma \left[1 + (\gamma+\rho)\frac{c_2}{\gamma}t^\rho\right] & , \gamma > 0 \text{ and } \rho < 0 \\ \gamma c_0 c_1 t^\gamma \left[1 + (\gamma+\rho)\frac{c_2}{\gamma}t^\rho\right] & , -1/2 < \gamma < \rho < 0 \\ \gamma c_0 c_1 t^\gamma \left[1 + (c_1 + 2c_2)t^\gamma\right] & , -1/2 < \gamma = \rho < 0 \end{cases} \tag{3.4.2}$$

and

$$A(t) = C\,t^\rho = \begin{cases} \frac{c_2}{\gamma}\rho^2 t^\rho & , \gamma > 0 \text{ and } \rho < 0 \\ \frac{c_2}{\gamma}(\gamma+\rho)\rho t^\rho & , -1/2 < \gamma \leq \rho < 0. \end{cases} \tag{3.4.3}$$

Note that $a(t) \in RV_\gamma$ and $A(t) \in RV_\rho$.

Lemma 3.4.1. *Let* $\log U$ *be as in* (3.2.7) *and*

$$M_j = \frac{M_n^{(j)} U^j(Y_{n-k,n})}{a^j(Y_{n-k,n})} - l_j\,, \qquad j = 1, 2$$

where $a(t)$ *is a positive function such that* (3.4.1) *holds and*

$$\begin{aligned} M_n^{(j)} &= \frac{1}{k}\sum_{i=0}^{k-1}\{\log U(Y_{n-i,n}) - \log U(Y_{n-k,n})\}^j, \\ 1/l_1 &= 1 - \gamma_-\,, \\ 1/l_2 &= (1-\gamma_-)(1-2\gamma_-)/2\,. \end{aligned}$$

Also choose a function $A(t) \to 0$ *such that* (3.4.1) *holds. Then for* $k = k(n) \to \infty$ *and* $k(n)/n \to 0$ $(n \to \infty)$

$$M_j = \frac{P_j}{\sqrt{k}} + d_j A(\frac{n}{k}) + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k})), \quad j = 1, 2$$

where (P_1, P_2) *is normally distributed with mean vector zero and covariance matrix*

$$\begin{cases} EP_1^2 = \frac{1}{(1-\gamma_-)^2(1-2\gamma_-)} \\ EP_2^2 = \frac{4(5-11\gamma_-)}{(1-\gamma_-)^2(1-2\gamma_-)^2(1-3\gamma_-)(1-4\gamma_-)} \\ E(P_1P_2) = \frac{4}{(1-\gamma_-)^2(1-2\gamma_-)(1-3\gamma_-)} \end{cases}$$

and

$$\begin{cases} d_1 = \frac{1}{(1-\gamma_-)(1-\rho-\gamma_-)} \\ d_2 = \frac{2(3-2\rho-4\gamma_-)}{(1-\gamma_-)(1-2\gamma_-)(1-\rho-\gamma_-)(1-\rho-2\gamma_-)}\,. \end{cases}$$

Proof. This lemma is an immediate consequence of Lemma 4.5 in Ferreira et al. (1999), since (3.4.1) holds. □

As a particular case, the following is a compilation of Lemmas 4.6 - 4.11 in Ferreira et al. (1999). Namely they state the consistency of the estimators $\hat{\gamma}_n(k)$ and $\hat{a}(\frac{n}{k})$.

Lemma 3.4.2. *Assume* (3.2.6) *and choose auxiliary functions* $a(t) > 0$ *and* $A(t) \to 0$ *such that* (3.4.1) *holds. Then, for* $k = k(n) \to \infty$ *and* $k(n)/n \to 0$ $(n \to \infty)$

$$\frac{\hat{\gamma}_n(k)}{\gamma} = 1 + \left(\frac{\gamma_+}{\gamma} - \frac{4}{\gamma l_1 l_2}\right) M_1 + \frac{2}{\gamma l_2^2} M_2 + \frac{q_{\gamma,\rho} l_1}{\gamma} A(\frac{n}{k})$$

with

$$q_{\gamma,\rho} = \lim_{t\to\infty} \frac{a(t)/U(t) - \gamma_+}{A(t)} = \begin{cases} \gamma/\rho & ,\gamma > 0 \text{ and } \rho < 0 \\ 0 & ,-1/2 < \gamma < \rho < 0 \\ \frac{c_1}{2c_2} & ,-1/2 < \gamma = \rho < 0, \end{cases}$$

$$\frac{\hat{a}(\frac{n}{k})}{a(\frac{n}{k})} = 1 + \frac{l_2 + 4l_1}{l_1 l_2} M_1 - \frac{2l_1}{l_2^2} M_2 + \gamma \frac{B}{\sqrt{k}} + o_p(A(\frac{n}{k}))$$

with B *a standard normal r.v., independent of* (P_1, P_2)*, and*

$$\frac{\hat{b}(\frac{n}{k}) - U(\frac{n}{k})}{a(\frac{n}{k})} = \frac{B}{\sqrt{k}} + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k})) .$$

Proof. Our conditions are just particular cases of the more general situation considered in Ferreira et al. (1999). Note that if $0 > \gamma > \rho'$ then $c_2 = -c_1/2$ and so $q_{\gamma,\rho} = -1$, as expected from the general result. To get $q_{\gamma,\rho}$ in the particular case $-1/2 < \gamma = \rho' = \rho < 0$, take w.l.g. the functions given in (3.4.2) and (3.4.3). □

The next lemma is presented in a somewhat technical way, directed to what we need for the proof of Theorem 3.2.1. For a more general result we refer to Ferreira et al. (1999), although their conditions are slightly different.

Lemma 3.4.3. *Suppose* (3.2.6) *and choose* $a(t) > 0$ *and* $A(t) \to 0$ *such that* (3.4.1) *holds. Then*

$$\lim_{\substack{t\to\infty \\ x\to\infty}} \frac{\frac{U(tx)-U(t)}{a(t)} \frac{\gamma}{x^\gamma - 1} - 1}{A(t)} = \begin{cases} -\frac{\gamma+\rho}{\rho^2} & \gamma > 0 \\ -\frac{1}{\rho+\gamma}\left(1 + \frac{c_1}{2c_2} 1_{\{\gamma=\rho\}}\right) & -1/2 < \gamma < 0. \end{cases}$$

Proof. Again, our conditions are just particular cases of the more general situation considered in Ferreira et al. (1999); if $0 > \gamma > \rho'$ then $c_2 = -c_1/2$ and so the above limit equals zero, as expected from the aforementioned general results. Note, however, that under our conditions and taking w.l.g. $a(t)$ and $A(t)$ as in (3.4.2) and (3.4.3), respectively, the result also follows. □

We shall now prove Theorem 3.2.1. Then we will see that it also proves Lemma 3.2.1.

Proof of Theorem 3.2.1. Let $a_n = k/(np_n)$. Then

$$\hat{p}_n(k) = \frac{k}{n}\max\left\{0, \left(1+\hat{\gamma}_n(k)\frac{x_n - \hat{b}(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\right)\right\}^{-1/\hat{\gamma}_n(k)}$$

$$= \frac{k}{n}\max\left\{0, \left(1+\gamma\frac{\hat{\gamma}_n(k)}{\gamma}\frac{a(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\left[\frac{U(\frac{n}{k}a_n)-U(\frac{n}{k})}{a(\frac{n}{k})}+\frac{U(\frac{n}{k})-U(Y_{n-k,n})}{a(\frac{n}{k})}\right]\right)\right\}^{-1/\hat{\gamma}_n(k)}$$

which, by the previous lemmas, for large values of n, has the same limit behaviour as

$$\frac{k}{n}\left\{1+\gamma\frac{\hat{\gamma}_n(k)}{\gamma}\frac{a(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\right.$$
$$\left[\frac{a_n^\gamma-1}{\gamma}-\frac{a_n^\gamma-1}{\gamma}\left(\frac{(\gamma+\rho)1_{\{\gamma>0\}}}{\rho^2}+\frac{1_{\{\gamma<0\}}}{\gamma+\rho}\left(1+\frac{c_1 1_{\{\gamma=\rho\}}}{2c_2}\right)\right)A(\frac{n}{k})\right.$$
$$\left.\left.+\frac{a_n^\gamma-1}{\gamma}o(A(\frac{n}{k}))-\frac{B}{\sqrt{k}}+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k}))\right]\right\}^{-1/\hat{\gamma}_n(k)}. \qquad (3.4.4)$$

The reasoning is divided in the two cases, γ positive and γ negative. First suppose $\gamma>0$. Then $a_n^{-\gamma}\to 0$ as $n\to\infty$ and so (3.4.4) becomes

$$\frac{k}{n}\left\{1+\gamma\frac{\hat{\gamma}_n(k)}{\gamma}\frac{a(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\right.$$

$$\left.\left[\frac{a_n^\gamma-1}{\gamma}+a_n^\gamma\left(-\frac{\gamma+\rho}{\gamma\rho^2}A(\frac{n}{k})+o(A(\frac{n}{k}))\right)-\frac{B}{\sqrt{k}}+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k}))\right]\right\}^{-1/\hat{\gamma}_n(k)}.$$

Using the expansions given in Lemma 3.4.2 this becomes

$$\frac{k}{n}\left\{1+\gamma\left[1+\left(-\frac{l_2+4l_1}{l_1l_2}+\frac{\gamma_+}{\gamma}-\frac{4}{\gamma l_1 l_2}\right)M_1+\left(\frac{2l_1}{l_2^2}+\frac{2}{\gamma l_2^2}\right)M_2-\frac{\gamma B}{\sqrt{k}}+\frac{l_1}{\rho}A(\frac{n}{k})\right]\right.$$

$$\left.\left[\frac{a_n^\gamma-1}{\gamma}+a_n^\gamma\left(-\frac{\gamma+\rho}{\gamma\rho^2}A(\frac{n}{k})+o(A(\frac{n}{k}))\right)-\frac{B}{\sqrt{k}}+o_p(\frac{1}{\sqrt{k}})+o_p(A(\frac{n}{k}))\right]\right\}^{-1/\hat{\gamma}_n(k)}$$

that is,

$$p_n\left(\frac{k}{np_n}\right)\left\{a_n^\gamma+a_n^\gamma\left[g_1M_1+g_2M_2-\frac{\gamma B}{\sqrt{k}}+g_3A(\frac{n}{k})\right]\right\}^{-1/\hat{\gamma}_n(k)}$$
$$= \quad p_n\left\{a_n^{\gamma-\hat{\gamma}_n(k)}\left(1+g_1M_1+g_2M_2-\frac{\gamma B}{\sqrt{k}}+g_3A(\frac{n}{k})\right)\right\}^{-1/\hat{\gamma}_n(k)} \qquad (3.4.5)$$

where g_1, g_2 and g_3 are non zero real constants depending on γ and ρ.

In the optimal case $a_n^{\gamma-\hat{\gamma}_n(k)}$ must converge to one in probability. Note that the second factor in the main brackets converges to one. In fact, we know that there exists a sequence $k = k(n)$ such that $a_n^{\gamma-\hat{\gamma}_n(k)} \to 1$ $(n \to \infty)$ in probability: take for $k_0(n)$, for example, the optimal one in extreme value index estimation (for this sequence we have $\hat{\gamma}_n(k_0) - \gamma = O((k_0(n))^{-1/2})$ and $\log(k_0/(np_n))/\sqrt{k_0} \to 0$ - for the later see Remark 3.2.4.3.). Note that the power $-1/\hat{\gamma}_n(k)$ has no influence since $-1/\hat{\gamma}_n(k) \to -1/\gamma$ $(n \to \infty)$ in probability. Therefore, as $n \to \infty$, if $a_n^{\gamma-\hat{\gamma}_n(k)} \to 1$, then $(a_n^{\gamma-\hat{\gamma}_n(k)} - 1)/((\gamma - \hat{\gamma}_n(k)) \log a_n) \to 1$ in probability and so, at least in the optimal case, the second factor in (3.4.5) does not contribute asymptotically. Hence the simplified expansion

$$p_n \{1 + (\gamma - \hat{\gamma}_n(k)) \log a_n + o((\gamma - \hat{\gamma}_n(k)) \log a_n)\}^{-1/\hat{\gamma}_n(k)} .$$

Disregarding terms of smaller order, we get

$$p_n \left\{1 - \frac{1}{\gamma} \frac{\gamma}{\hat{\gamma}_n(k)} (\gamma - \hat{\gamma}_n(k)) \log a_n \right\}$$

or,

$$p_n - \frac{p_n}{\gamma} (\gamma - \hat{\gamma}_n(k)) \log a_n .$$

Consequently, the second moment of the approximating r.v. of $(\hat{p}_n(k)/p_n - 1)$ is asymptotic to

$$\begin{aligned} &(\log a_n)^2 E(\gamma - \hat{\gamma}_n(k))^2 \\ \sim \quad &(\log \frac{k}{np_n})^2 \left(\frac{var_\gamma}{k} + bias_{\gamma,\rho}^2 \frac{c_2^2}{\gamma^2} \rho^4 \, (\frac{n}{k})^{2\rho} \right), \end{aligned}$$

as $n \to \infty$, where we have used $A(n/k) \sim c_2 \rho^2/\gamma \; (n/k)^\rho$ and, var_γ and $bias_{\gamma,\rho}$ are obtained from the expantion given in Lemma 3.4.2 (and from the covariance matrix given in Lemma 3.4.1). Thus taking the derivative with respect to k in the last expression and equating it to zero, one gets the result. For more details see Ferreira et al. (1999), e.g. proof of Proposition 4.12.

Next suppose $\gamma < 0$. Then $a_n^\gamma \to 0$ as $n \to \infty$ and so, by Lemma 3.4.2 relation (3.4.4) leads to

$$\frac{k}{n} \left\{ 1 + \gamma \left[1 - \frac{4}{\gamma l_1 l_2} M_1 + \frac{2}{\gamma l_2^2} M_2 + \frac{c_1 l_1 \, 1_{\{\gamma=\rho\}}}{2\gamma c_2} A(\frac{n}{k}) \right] \right.$$

$$\left[1 - \frac{l_2 + 4l_1}{l_1 l_2} M_1 + \frac{2l_1}{l_2^2} M_2 - \gamma \frac{B}{\sqrt{k}} + o_p(A(\frac{n}{k})) \right]$$

$$\left. \left[\frac{a_n^\gamma - 1}{\gamma} + \frac{1 + \frac{c_1}{2c_2} 1_{\{\gamma=\rho\}}}{\gamma(\rho+\gamma)} A(\frac{n}{k}) - \frac{B}{\sqrt{k}} + o_p(\frac{1}{\sqrt{k}}) + o_p(A(\frac{n}{k})) \right] \right\}^{-1/\hat{\gamma}_n(k)}$$

or

$$\frac{k}{n}\left\{a_n^{\gamma}+\left(\frac{l_2+4l_1}{l_1l_2}+\frac{4}{\gamma l_1 l_2}\right)M_1-\left(\frac{2l_1}{l_2^2}+\frac{2}{\gamma l_2^2}\right)M_2+\frac{1+\frac{c_1}{2c_2}1_{\{\gamma=\rho\}}}{\rho+\gamma}A(\frac{n}{k})\right.$$
$$\left.-\frac{c_1l_1\,1_{\{\gamma=\rho\}}}{2\gamma c_2}A(\frac{n}{k})\right\}^{-1/\hat{\gamma}_n(k)}$$

that is,

$$\begin{aligned}
&p_n\left(\frac{k}{np_n}\right)\left\{a_n^{\gamma}+g_4M_1+g_5M_2+g_6A(\frac{n}{k})\right\}^{-1/\hat{\gamma}_n(k)}\\
&=\ p_n\left\{a_n^{\gamma-\hat{\gamma}_n(k)}+a_n^{-\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6A(\frac{n}{k})\right)\right\}^{-1/\hat{\gamma}_n(k)}\\
&=\ p_n\left\{1+\left(a_n^{\gamma-\hat{\gamma}_n(k)}-1\right)+a_n^{-\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6A(\frac{n}{k})\right)\right\}^{-1/\hat{\gamma}_n(k)}
\end{aligned}$$

where g_4 and g_5 are non zero real constants depending on γ and ρ; g_6 also depends on c_1 and c_2 in the case $\gamma=\rho'$.

Next we prove that

$$\frac{a_n^{\gamma-\hat{\gamma}_n(k)}-1}{a_n^{-\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6A(\frac{n}{k})\right)}\to 0 \qquad (n\to\infty) \tag{3.4.6}$$

in probability. Note that $\gamma-\hat{\gamma}_n(k)=O_p(g_4M_1+g_5M_2+g_6A(\frac{n}{k}))$. Therefore (3.4.6) is of the same order as

$$\begin{aligned}
&\left|\frac{a_n^{\gamma-\hat{\gamma}_n(k)}-1}{a_n^{-\hat{\gamma}_n(k)}\left(\gamma-\hat{\gamma}_n(k)\right)}\right|=\left|\frac{a_n^{\gamma}-a_n^{\hat{\gamma}_n(k)}}{\gamma-\hat{\gamma}_n(k)}\right|=\\
&=\ \left|\frac{\log a_n}{\gamma-\hat{\gamma}_n(k)}\int_{\hat{\gamma}_n(k)}^{\gamma}a_n^s ds\right|\le(\log\ a_n)a_n^{\max(\hat{\gamma}_n(k),\gamma)}\to 0\ \text{ as }\ n\to\infty\ .
\end{aligned}$$

Hence the numerator in (3.4.6) is of smaller order than the denominator, and so (3.4.4) simplifies to

$$p_n\left\{1+a_n^{-\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6A(\frac{n}{k})\right)\right\}^{-1/\hat{\gamma}_n(k)}\ . \tag{3.4.7}$$

In the optimal case $a_n^{-\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6A(\frac{n}{k})\right)$ must converge to zero in probability and in fact there exists a sequence such that this holds (take for example the optimal one in extreme value index estimation together with condition (3.2.14)). So, expanding (3.4.7) again we get, neglecting terms of lower order,

$$p_n\left\{1-\frac{a_n^{-\gamma}}{\hat{\gamma}_n(k)}\left(g_4M_1+g_5M_2+g_6A(\frac{n}{k})\right)\right\}$$

or,

$$p_n - \frac{p_n}{\gamma} a_n^{-\gamma} \left(g_4 M_1 + g_5 M_2 + g_6 A(\frac{n}{k}) \right) .$$

Consequently, the second moment of the approximating r.v. of $(\hat{p}_n(k)/p_n - 1)$ is asymptotic to

$$\begin{aligned} & \frac{p_n^{2\gamma}}{\gamma^2} \left(\frac{n}{k}\right)^{2\gamma} E \left(g_4 M_1 + g_5 M_2 + g_6 A(\frac{n}{k}) \right)^2 \\ \sim \quad & p_n^{2\gamma} \left(\frac{n}{k}\right)^{2\gamma} \left(\frac{var_\gamma}{k} + bias_{\gamma,\rho}^2 \frac{c_2^2(\gamma+\rho)^2\rho^2}{\gamma^2} (\frac{n}{k})^{2\rho} \right) \end{aligned}$$

as $n \to \infty$, where we have used $A(n/k) \sim c_2(\gamma+\rho)\rho/\gamma\ (n/k)^\rho$ and, var_γ and $bias_{\gamma,\rho}$ are obtained from Lemmas 3.4.1 and 3.4.2. Thus taking the derivative with respect to k in the last expression and equating it to zero, one gets the result. For more details see Ferreira et al. (1999), e.g. proof of Proposition 4.12. Note that in order to assure that a minimum is in fact attained one must assume $\gamma > -1/2$. □

Proof of Lemma 3.2.1. Let $a_n = k/(np_n)$. The lemma follows directly from the proof of the Theorem. See (3.4.5) for $\gamma > 0$ and (3.4.7) for $\gamma < 0$, and the expansions afterwards. Note that the conditions on $r_\gamma(a_n)$, ensure that $a_n^{\gamma - \hat{\gamma}_n(k)}$ converges to one in probability for the case $\gamma > 0$, and that $a_n^{-\hat{\gamma}_n(k)} \left(O_p(1/\sqrt{k}) + O_p((n/k)^\rho) \right)$ converges to zero in probability in the case $\gamma < 0$. □

Appendix. Adaptive bootstrap on exceedance probability estimation

Without going into details, we remark that theoretically the adaptive bootstrap method to estimate $k_0(n)$ is still valid on exceedance probability estimation. The proof follows the same line as in, e.g. Draisma et al. (1999) on extreme value index estimation and Ferreira et al. (1999) on endpoint and high quantiles estimation. One of the main steps is Theorem 3.2.1 in Section 3.2. In the following we shall just give the main steps necessary to implement this method.

We want the value of k minimising $E[(\hat{p}_n(k)/p_n - 1)^2]$ (although this is only meant in an asymptotic sense, i.e. second moment of the approximating asymptotic distribution), where p_n and the underlying d.f. are unknown. The ideia is to replace these unknown quantities such that the optimal $k_0(n)$ is still attained. That is, the proposed bootstrap method is based on minimising, over $k = k(n)$, the sample mse of $(\hat{p}_n^*(k)/\hat{\bar{p}}_n^*(k) - 1)$. From a bootstrap sample of size n, $\hat{p}_n^*(k)$ is calculated from (3.1.3) and $\hat{\bar{p}}_n^*(k)$ is an alternative estimator of p_n. For $\hat{\bar{p}}_n(k)$ we still use (3.1.3) but with

$$\hat{\bar{\gamma}}_n(k) = \sqrt{M_n^{(2)}/2} + 1 - \frac{2}{3}(1 - \frac{M_n^{(1)} M_n^{(2)}}{M_n^{(3)}})^{-1}, \tag{3.A.1}$$

in place of (3.2.1); similarly for (3.2.2). For $M_n^{(3)}$ just take $j = 3$ in (3.2.3). Then, similarly as in Theorem 3.2.1, we need the variance and the bias of the limit distribution of $r_\gamma(k,n)\sqrt{k}\,(\hat{p}_n(k) - \hat{\bar{p}}_n(k))/p_n$. Following a similar reasoning as in the proof of Theorem 3.2.1 one gets

$$\overline{var}_\gamma = \begin{cases} \frac{1}{4}\left(1+\frac{1}{\gamma^2}\right) & ,\gamma > 0 \\ \frac{(1-\gamma)^2(1-6\gamma+35\gamma^2-78\gamma^3+72\gamma^4)}{4\gamma^4(1-2\gamma)(1-3\gamma)(1-4\gamma)(1-5\gamma)(1-6\gamma)} & ,\gamma < 0 \end{cases} \tag{3.A.2}$$

and

$$\overline{bias}_{\gamma,\rho} = \begin{cases} \frac{-\rho-\gamma+\gamma\rho}{2\gamma(1-\rho)^3} & ,\gamma > 0 \\ \frac{(\gamma-1)\rho}{2\gamma^2(1-\gamma-\rho)(1-2\gamma-\rho)(1-3\gamma-\rho)} & ,-1/2 < \gamma < \rho'(=\rho) < 0 \\ \frac{-2+17\gamma-50\gamma^2+47\gamma^3}{2\gamma^2(1-\gamma)(1-2\gamma)(1-3\gamma)(1-4\gamma)} & \\ -\frac{2-20\gamma+70\gamma^2-100\gamma^3+48\gamma^4}{2\gamma^2(1-\gamma)(1-2\gamma)(1-3\gamma)(1-4\gamma)\sqrt{(1-\gamma)(1-2\gamma)}} & ,\gamma(=\rho) > \rho' \text{ and} \\ & \quad -1/2 < \gamma < 0. \end{cases} \tag{3.A.3}$$

The bootstrap procedure follows: *Step 1)* Select randomly and independently n_1 times ($n_1 = O(n^{1-\varepsilon})$, $0 < \varepsilon < 1/2$) a member from the sample $\{X_1, X_2, \ldots, X_n\}$. Indicate the result by $X_1^*, X_2^*, \ldots X_{n_1}^*$. Form the order statistics $X_{1,n_1}^* \le X_{2,n_1}^* \le \cdots \le X_{n_1,n_1}^*$ and compute the quantities $\hat{p}_n(k)$ and $\hat{\bar{p}}_n(k)$. We denote the resulting quantities by $\hat{p}_n^*(k)$ and $\hat{\bar{p}}_n^*(k)$ for $k = 1, 2, \ldots, n_1 - 1$. Form $q_{n_1,k}^* = (\hat{p}_n^*(k)/\hat{\bar{p}}_n^*(k) - 1)^2$ on the basis of these bootstrap estimators; *Step 2)* Repeat step 1, r times independently. This results in a sequence $q_{n_1,k,s}^*$, $k = 1, 2, \ldots, n_1 - 1$ and $s = 1, 2, \ldots, r$. Calculate $\frac{1}{r}\sum_{s=1}^r q_{n_1,k,s}^*$; *Step 3)* Minimise $\frac{1}{r}\sum_{s=1}^{r} q_{n_1,k,s}^*$ with respect to k but reject values which are very small or very near to n_1. Denote the value of k where the minimum is obtained by $k_0^*(n_1)$; *Step 4)* Repeat step 1 up to 3 independently with the number n_1 replaced by $n_2 = (n_1)^2/n$. So n_2 is smaller than n_1. This results in $k_0^*(n_2)$; *Step 5)* Calculate

$$\hat{k}_0(n) = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)} \left(\frac{var_{\hat{\gamma}}\ \overline{bias}^2_{\hat{\gamma},\hat{\rho}}}{\overline{var}_{\hat{\gamma}}\ bias^2_{\hat{\gamma},\hat{\rho}}} \right)^{\frac{1}{1-2\hat{\rho}}}$$

with $\hat{\gamma}$ any consistent estimator of γ (we have used (3.1.3) with $k = \sqrt{n}$) and, $\hat{\rho} = \hat{\rho}_{n_1}(k_0^*) = \log k_0^*(n_1)/(-2\log n_1 + 2\log k_0^*(n_1))$ a consistent estimator of ρ. This $\hat{k}_0(n)$, which is obtained adaptively, is asymptotically as good as the optimal number of order statistics in (3.2.15) (for further details we refer to Ferreira et al., 1999).

Acknowledgments. The author gratefully acknowledges Prof. Laurens de Haan for his support throughout the work presented in this paper. The author also thanks a referee for her/his suggestions that simplified and improved the paper. The author's research was partially supported by PRAXIS XXI/FEDER.

Chapter 4

On optimising confidence intervals for the tail index and high quantiles

Co-author: Casper de Vries

Abstract. The aim of this paper is to obtain confidence intervals for the tail index and high quantiles with their optimal rate. When obtaining confidence intervals for these quantities, the common approach that we find in the literature is to use the normal distribution approximation with a non-optimal rate. We propose to use the optimal rate, but then additional problems arise, since a bias term with unkown sign has to be estimated. We provide an estimator for this sign and the full programme to obtain these optimal confidence intervals. Moreover, we demonstrate the gain in coverage, and show the relevance of these confidence intervals by calculating the reduction in capital requirements in a Value at Risk exercise. Simulation results are also presented.

4.1 Introduction

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from some unknown distribution function F, and denote the order statistics by $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. Suppose F satisfies the maximum domain of attraction condition (Fisher and Tippett, 1928; Gnedenko, 1943), with positive extreme value index. In terms of regularly varying functions, this is equivalent to that for some $\gamma > 0$

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}, \quad \text{for all } x > 0. \tag{4.1.1}$$

Then F is said to have a regularly varying tail with index $-1/\gamma$ (i.e. $1-F \in \mathrm{RV}_{-1/\gamma}$).

For any non-decreasing function f, let $f^{\leftarrow}$ denote its left-continuous inverse, that is $f^{\leftarrow}(y) = \inf\{x : f(x) \geq y\}$. Let $U = (1/(1-F))^{\leftarrow}$. Consider the following refinement of condition (4.1.1) (e.g. de Haan, 1994; de Haan and Stadtmüller, 1996). Suppose there exists a function a, with constant sign near infinity and $\lim_{t\to\infty} a(t) = 0$, such that

$$\lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{a(t)} = \frac{x^{\rho} - 1}{\rho}, \quad \text{for all } x > 0,\ \rho < 0. \tag{4.1.2}$$

The following estimator of γ was proposed by Hill (1975),

$$\hat{\gamma}(k) = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}. \tag{4.1.3}$$

It is well known that Hill's estimator has, is general, large variance for small values of k and large bias for large values of k. Hence, when estimating γ, one usually looks for k balancing this trade-off.

Let k_n be an intermediate sequence, i.e. $k_n \to \infty$ and $k_n/n \to 0$, as $n \to \infty$, and

$$H_{n,k_n} = \sqrt{k_n}\left(\frac{\hat{\gamma}(k_n)}{\gamma} - 1\right). \tag{4.1.4}$$

Under condition (4.1.2) and if $a(n/k_n)\sqrt{k_n} \to \lambda$, $\lambda \in (-\infty, \infty)$, H_{n,k_n} converges in distribution to a normal random variable with mean $\lambda/(\gamma(1-\rho))$ and variance 1 (e.g. Hall, 1982; Dekkers et al., 1989). The best rate of convergence is attained when $a(n/k_n)\sqrt{k_n} \to \lambda \neq 0$, and in this case the limiting distribution has non-zero mean.

Usually when using Hill's estimator in applied problems, one simply uses (4.1.4) with $\lambda = 0$ in order to construct a confidence interval for γ (e.g. Cheng and Peng, 2001 and Caers et al. 1998). In this paper we construct a confidence interval for gamma, using (4.1.4) but with the sequence k_n being the optimal rate of convergence, in the sense of minimizing the asymptotic mean square error. This will be done in Section 4.2. In order to implement this confidence interval, several problems have to be solved: one needs an adaptive way to obtain the optimal sequence k_n, moreover one needs to estimate two new parameters consistently, the second order parameter ρ and the sign of the asymptotic bias. For the adaptive choice of the optimal sequence k_n we follow Danielsson et al. (2001). For the estimation of the parameter ρ we follow Danielsson et al. (2001) and Fraga Alves et al. (2001). For the sign of the asymptotic bias, in Section 4.3 we introduce a new estimator and show its consistency. In Section 4.4 we obtain optimal confidence intervals for high quantiles. From the results it follows that the reverse problem of estimation of tail probabilities can be similarly worked out. We leave it to the reader. In Section 4.5 are simulation results and in Section 4.6 is a Value at Risk data analysis.

Related papers on confidence interval estimation are Caers et al. (1998) and Cheng and Peng (2001). In the first paper the authors also use the bootstrap methodology to obtain the optimal k_n, though in a rather different way from what we consider here, but then they end up considering $\lambda = 0$ to obtain the confidence intervals. In the later paper the authors try to find k_n optimising the confidence interval but their criterion is quite different from ours. They look for the optimal sequence in the non-optimal range of values satisfying $a(n/k_n)\sqrt{k_n} \to 0$. Later on we discuss that even with regard to the coverage probability, to consider $\lambda = 0$ is a non-optimal choice. The authors also point out the importance of the sign of the asymptotic bias but they do not discuss explicitly its estimation.

We restrict ourselves to the case $\rho < 0$, since optimality results for the choice of k_n are well established for this case.

4.2 Optimal confidence interval for the tail index

Let $a(n/k_n)\sqrt{k_n} \to \lambda \in \mathbb{R}$. Denote by Φ the standard normal distribution function and $z_\alpha = \Phi^{\leftarrow}(1-\alpha)$. As a first approach to construct a confidence interval with significance level α for γ, based on (4.1.4) and its limiting distribution, if we solve $-z_\alpha < H_{n,k_n} - \lambda/(\gamma(1-\rho)) < z_\alpha$ in γ we get

$$\frac{\hat{\gamma}(k_n)\sqrt{k_n} - \frac{\lambda}{1-\rho}}{z_{\alpha/2} + \sqrt{k_n}} < \gamma < \frac{\hat{\gamma}(k_n)\sqrt{k_n} - \frac{\lambda}{1-\rho}}{-z_{\alpha/2} + \sqrt{k_n}}, \tag{4.2.1}$$

provided $\sqrt{k_n} - z_{\alpha/2} > 0$.

Let k_n^0 denote the 'optimal' sequence, in the sense of minimizing the mean square error of the limiting distribution (Hall and Welsh, 1985; Dekkers and de Haan, 1993). Under our conditions this sequence is easily calculated if one assumes, moreover, that the regularly varying function a behaves, asymptotically, as a power function,

$$a(t) \sim \; ct^{\rho}, \quad c \neq 0, \tag{4.2.2}$$

as $t \to \infty$. Then, our assumptions are equivalent to assuming Hall's model

$$1 - F(x) = \mathrm{C}x^{-1/\gamma}\left(1 + \mathrm{D}x^{\rho/\gamma} + o(x^{\rho/\gamma})\right), \quad \mathrm{C} > 0, \mathrm{D} \neq 0, \quad x \to \infty$$

where, from (4.1.2) and (4.2.2) we have $D = c\gamma^{-1}\rho^{-1}\mathrm{C}^{\rho}$. Therefore (Hall and Welsh (1985))

$$k_n^0 \sim \left(\frac{\gamma^2(1-\rho)^2}{-2\rho\, c^2}\right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}. \tag{4.2.3}$$

Then it is easy to see that the value λ for this sequence k_n^0, is asymptotic to $\mathrm{sign}(c)\gamma(1-\rho)/\sqrt{-2\rho}$. Then, in this case (4.2.1) simplifies to

$$\frac{\hat{\gamma}(k_n^0)\sqrt{k_n^0}}{z_{\alpha/2} + \mathrm{sign}(c)/\sqrt{-2\rho} + \sqrt{k_n^0}} < \gamma < \frac{\hat{\gamma}(k_n^0)\sqrt{k_n^0}}{-z_{\alpha/2} + \mathrm{sign}(c)/\sqrt{-2\rho} + \sqrt{k_n^0}}.$$

Now, in order to obtain a confidence interval from this inequality, we need to approximate adaptively k_n^0, and estimate ρ and sign(c). From Hall and Welsh (1985) - Theorem 4.1, an adaptive choice $\hat{k}_n^0$ can be used for which

$$\frac{\hat{k}_n^0}{k_n^0} \xrightarrow{P} 1. \tag{4.2.4}$$

For ρ we need a consistent estimator and for sign(c) an estimator $\widehat{\text{sign}}$ satisfying

$$\mathrm{P}\{\widehat{\text{sign}} = \text{sign}(c)\} \to 1 \qquad (n \to \infty). \tag{4.2.5}$$

Then the following theorem holds.

Theorem 4.2.1. *Suppose* (4.1.2) *and* (4.2.2). *Let* $\hat{k}_n^0$ *satisfy* (4.2.4), $\widehat{sign}$ *satisfy* (4.2.5) *and let* $\hat{\rho}$ *be a consistent estimator for* ρ. *Then, as* $n \to \infty$,

$$\sqrt{\hat{k}_n^0}\left(\frac{\hat{\gamma}(\hat{k}_n^0)}{\gamma} - 1\right) - \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}}$$

converges, in distribution, to a standard normal random variable. Therefore, as $n \to \infty$,

$$P\left(\frac{\hat{\gamma}(\hat{k}_n^0)\sqrt{\hat{k}_n^0}}{z_{\alpha/2} + \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}} + \sqrt{\hat{k}_n^0}} < \gamma < \frac{\hat{\gamma}(\hat{k}_n^0)\sqrt{\hat{k}_n^0}}{-z_{\alpha/2} + \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}} + \sqrt{\hat{k}_n^0}}\right) \to 1 - \alpha \tag{4.2.6}$$

which gives an asymptotic confidence interval for γ, *with confidence coefficient* $1-\alpha$.

Note that for the cases where the true γ is near zero a one-sided confidence interval can alternatively be considered. The extension of our results to this case is obvious.

Accuracy of the confidence interval. Denote the confidence interval based on (4.2.1), where k_n is such that $\lambda = 0$ and ρ is replaced by a consistent estimator, by $(\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})$. In the following we shall see that the confidence interval (4.2.6) is more accurate than $(\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})$.

Fix

$$\lim_{n\to\infty} P(\gamma \in (\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n})) = 1 - \alpha,$$

for each $\gamma > 0$. Then, the probabilities of covering the wrong value γ',

$$P(\gamma' \in [\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n}]) \tag{4.2.7}$$

should be as small as possible, for each $\gamma' > 0$. In fact for $\gamma' \neq \gamma$ and all sequences k_n with $a(n/k_n)\sqrt{k_n} \to \lambda$ this probability converges to zero, since the lower and upper limits of the confidence interval converge to γ in probability. Therefore next we compare the probabilities of wrong coverage when $\gamma_n'/\gamma \to 1$ $(n \to \infty)$.

For the confidence interval (4.2.6) the probability of wrong coverage equals

$$P\Big(- z_{\alpha/2}\frac{\gamma_n'}{\gamma} \\ \leq \sqrt{\hat{k}_n^0}\left(\frac{\hat{\gamma}(\hat{k}_n^0)}{\gamma} - 1\right) - \frac{\widehat{\text{sign}}}{\sqrt{-2\hat{\rho}}} + \left(1 - \frac{\gamma_n'}{\gamma}\right)\left(\sqrt{\hat{k}_n^0} - \frac{\widehat{\text{sign}}}{\sqrt{-2\hat{\rho}}}\right) \leq z_{\alpha/2}\frac{\gamma_n'}{\gamma}\Big).$$

Hence for sequences γ_n' with $\sqrt{k_n^0}(1-\gamma_n'/\gamma) \to \nu \neq 0, \pm\infty$, this probability converges to $\Phi(z_{\alpha/2} - \nu) - \Phi(-z_{\alpha/2} - \nu)$. Now take k_n with $a(n/k_n)\sqrt{k_n} \to 0$. Then for sequences γ_n^* with $\sqrt{k_n}(1 - \gamma_n^*/\gamma) \to \nu^* \neq 0, \pm\infty$, probability (4.2.7) equals

$$P\left(-z_{\alpha/2}\frac{\gamma_n^*}{\gamma} + \sqrt{k_n}\left(\frac{\gamma_n^*}{\gamma} - 1\right) \leq \sqrt{\hat{k}_n}\left(\frac{\hat{\gamma}(k_n)}{\gamma} - 1\right) \leq z_{\alpha/2}\frac{\gamma_n^*}{\gamma} + \sqrt{k_n}\left(\frac{\gamma_n^*}{\gamma} - 1\right)\right)$$

which converges to $\Phi(z_{\alpha/2} - \nu^*) - \Phi(-z_{\alpha/2} - \nu^*)$. Therefore in order to compare the two probabilities take a common sequence, e.g. γ_n^*. Then in the first case the probability of covering the wrong values converges to zero whilst in the second case it is equal to $\Phi(z_{\alpha/2} - \nu^*) - \Phi(-z_{\alpha/2} - \nu^*) > 0$.

We have attempted to improve the standard results for obtaining confidence intervals, by using the first term in the Edgeworth expansion for Hill's estimator. This has been done by Cheng and Pan (1998) - Section 4.1, in the situation without asymptotic bias (i.e. with the number of order statistics used of smaller order than k_n^0). Even we have attempted to use an Edgeworth expansion for the distribution of Hill's estimator, approximated not by the normal but by an appropriate gamma distribution, similar to Cheng and de Haan (2001).

For these two extensions we need the Edgeworth expansion in the situation of a non-null asymptotic bias. This includes the case when the optimal number of upper order statistics is used. In case of an approximation by the normal distribution this Edgeworth expansion has been given by Cuntz and Haeusler (2001). The Edgeworth expansion in case of an approximation by a gamma distribution is presented in Appendix B below, but it turns out that in the case of non-zero asymptotic bias the approximation by a gamma distribution does not lead to an improved rate. We decided not to pursue this further refinement since it leads to the need to estimate some new parameters and it seems not easy to do that in a way that can be useful for applications.

4.3 Estimation of the sign of the bias of Hill's estimator

For convenience, in this section we shall use the Hill process parameterised continuously. The following result was taken from Drees et al. (2000), Corollary 1.

Lemma 4.3.1. *Let k_n denote an arbitrary intermediate sequence. Under condition* (4.1.2)*, there exists a probability space carrying $X_1, X_2, \ldots$ and a sequence of Brownian motions W_n, such that*

$$\sup_{t_n \le t \le 1} t^{1/2} \left| \hat{\gamma}([k_n t]) - \left(\gamma + \frac{\gamma}{\sqrt{k_n}} \frac{W_n(t)}{t} + a(\frac{n}{k_n}) \frac{t^{-\rho}}{1-\rho} \right) \right| = o_p \left(k_n^{-1/2} + a(\frac{n}{k_n}) \right)$$

for all $t_n \to 0$, satisfying $k_n t_n \to \infty$.

In this expansion for the Hill process, we call the term $t^{-\rho}/(1-\rho)a(n/k_n)\sqrt{k_n}$, the bias of Hill's estimator. Note that if $a(n/k_n)\sqrt{k_n} \to \lambda \neq 0$, its asymptotic sign is determined by the sign of the function $a(n/k_n)$, which equals sign(c) provided (4.2.2) holds. For instance if $t = 1$, the sign of the expected value of the limiting variable of $\sqrt{k_n}(\hat{\gamma} - \gamma)$ equals sign$(\lambda/(1-\rho))$ = sign(c).

Let a_n, b_n and c_n be intermediate sequences such that

$$a_n < b_n \le c_n \text{ for all } n, \text{ and } a_n/b_n \to \nu \in [0,1). \tag{4.3.1}$$

We suggest the following estimator for the sign of the bias,

$$\widehat{\text{sign}} = \text{sign} \left(\hat{\gamma}(c_n) - \frac{1}{b_n - a_n + 1} \sum_{i=a_n}^{b_n} \hat{\gamma}(i) \right). \tag{4.3.2}$$

Theorem 4.3.1. *Assume* (4.1.2), (4.3.1) *and moreover that b_n satisfies $a(\frac{n}{b_n})\sqrt{b_n} \to \infty$. Then*

$$P\{\widehat{sign} = sign(c)\} \to 1, \qquad n \to \infty.$$

Proof. Lemma 4.3.1 implies that

$$\int_{a_n/b_n}^{1} \left| \hat{\gamma}([b_n t]) - \left(\gamma + \frac{\gamma}{\sqrt{b_n}} \frac{W_n(t)}{t} + a(\frac{n}{b_n}) \frac{t^{-\rho}}{1-\rho} \right) \right| dt = o_p(b_n^{-1/2} + a(\frac{n}{b_n})).$$

Therefore,

$$\begin{aligned} \lim_{n\to\infty} \frac{\hat{\gamma}(c_n) - \frac{1}{b_n - a_n + 1} \sum_{i=a_n}^{b_n} \hat{\gamma}(i)}{a(\frac{n}{b_n})} &= \lim_{n\to\infty} \frac{\hat{\gamma}(c_n) - \frac{b_n}{b_n - a_n} \int_{a_n/b_n}^{1} \hat{\gamma}_n([b_n t]) dt}{a(\frac{n}{b_n})} \\ = \lim_{n\to\infty} &\left\{ \frac{\gamma}{a(\frac{n}{b_n})} \left(\frac{W_n(1)}{\sqrt{c_n}} - \frac{b_n}{\sqrt{b_n}(b_n - a_n)} \int_{a_n/b_n}^{1} \frac{W_n(t)}{t} dt \right) \right. \\ &\left. + \frac{1}{1-\rho} \left(\frac{a(\frac{n}{c_n})}{a(\frac{n}{b_n})} - \frac{b_n}{b_n - a_n} \int_{a_n/b_n}^{1} t^{-\rho} dt \right) + o_p(\frac{1}{a(\frac{n}{b_n})\sqrt{b_n}}) \right\}. \end{aligned}$$

Since $a_n/b_n \to \nu \in [0,1)$, we have that $(b_n/(b_n - a_n)) \int_{a_n/b_n}^{1} W_n(t)/t \, dt = O_p(1)$. Hence taking $b_n \le c_n$ such that $a(\frac{n}{b_n})\sqrt{b_n} \to \infty$, the first and last factors in the last equality go to zero, in probability. For the second factor, just note that under the given conditions $a(n/c_n)/a(n/b_n) \geq 1$ and that $(1/(1-\nu))\int_\nu^1 t^{-\rho} dt < 1$, for all $\nu \in [0,1)$ and $\rho < 0$. Hence we have that the second factor converge in probability to some positive constant. □

Remark 4.3.1. Although we have excluded the case $\rho = 0$ in condition (4.1.2), the results of this section are still valid for $\rho = 0$, provided $b_n < c_n$ in (4.3.1).

Remark 4.3.2. Other proposals to estimate the sign could be using two consistent estimators of γ, for instance $\hat{\gamma}_1 = \hat{\gamma}$ and $\hat{\gamma}_2 = (2k_n)^{-1/2} \sqrt{\sum_{i=0}^{k_n-1}(\log X_{n-i,n} - \log X_{n-k,n})^2}$. Both admit expansions of the type $\hat{\gamma}_i = \gamma + c_i k_n^{-1/2} P_i + d_i a(n/k_n) + o_p(k_n^{-1/2}) + o_p(a(n/k_n))$, where c_i, d_i are some known constants and P_i are normal (0,1) random variables. Hence, for large k_n (i.e. $a(n/k_n)k_n^{-1/2} \to \infty$) and with $a(n/k_n) \sim c\,(n/k_n)^\rho$ we have

$$\operatorname{sign}\left(\frac{\hat{\gamma}_1}{\hat{\gamma}_2} - 1\right) = c\,\frac{d_1 - d_2}{\gamma}.$$

But typically one finds two kind of problems with this sort of estimators. First of all, they are very sensitive to the choice of k. Commonly a plot of $\operatorname{sign}(\hat{\gamma}_1\hat{\gamma}_2^{-1} - 1)$ versus k frequently alter sign. Secondly, since these estimators have similar behaviour, e.g. they have the same sign of the bias and predominance of one bias over the other, in a plot of $\operatorname{sign}(\hat{\gamma}_1\hat{\gamma}_2^{-1} - 1)$ these features turn out to be the most predominant.

4.4 Optimal confidence interval for high quantile

Suppose one is given a small probability p and one wants to estimate the quantile $x : P(X > x) = p$. We are interested in studying the situations where p is indeed very small, for instance where this small probability corresponds to an event that has never been observed. More specifically, $p = p_n$ must depend on n (size of the sample), since we use asymptotic theory, and we shall assume $np_n \to$ constant (≥ 0).

As before let k_n be an intermediate sequence. To estimate a high quantile $x_n = F^{\leftarrow}(1 - p_n)$ we suggest the following estimator (Dekkers and de Haan, 1989; Ferreira et al., 1999)

$$\hat{x}(k_n) = X_{n-k_n,n}\left(\frac{k_n}{np_n}\right)^{\hat{\gamma}(k_n)}. \tag{4.4.1}$$

The following result is from Ferreira et al. (1999).

Lemma 4.4.1. *Assume* (4.1.2), (4.2.2) *and* $np_n \to$*constant* (≥ 0). *Let* k_n *be an intermediate sequence such that* $a(n/k_n)\sqrt{k_n} \to \lambda \in (-\infty, \infty)$ *and* $\log(k_n/np_n)/\sqrt{k_n} \to 0$. *Then*

$$\frac{\sqrt{k_n}}{\gamma \log(\frac{k_n}{np_n})}\left(\frac{\hat{x}(k_n)}{x_n} - 1\right) \tag{4.4.2}$$

converges in distribution to a normal random variable, with mean $\lambda/(\gamma(1-\rho))$ *and variance 1.*

Let ${}^q k_n^0$ denote the sequence k_n minimizing the mean square error of the limiting distribution of (4.4.2). Then from Ferreira et al. (1999) it follows that (when $\gamma > 0$) ${}^q k_n^0 \sim k_n^0$ where k_n^0 is from (4.2.3). Hence a consistent estimator of ${}^q k_n^0$ is $\hat{k}_n^0$ from Section 4.2, that is,

$$\frac{\hat{k}_n^0}{{}^q k_n^0} \xrightarrow{P} 1. \tag{4.4.3}$$

Moreover, also from this paper it follows that Lemma 4.4.1 still holds if k_n is replaced by $\hat{k}_n^0$.

Therefore, similarly as in Section 4.2, the following theorem holds. For more details on the proof we refer to Ferreira et al. (1999). Motivated by an application on VAR estimation, we consider here a one-sided confidence interval. The changes to obtain a similar result for a two-sided confidence interval are obvious.

Theorem 4.4.1. *Suppose* (4.1.2), (4.2.2), *and that* $np_n \to$ *constant* (≥ 0) *and* $\log(p_n) = o(n^{-\rho/(1-2\rho)})$, *as* $n \to \infty$. *Let* $\hat{k}_n^0$ *satisfy* (4.4.3), $\widehat{sign}$ *satisfy* (4.2.5) *and let* $\hat{\rho}$ *be a consistent estimator for* ρ. *Then, as* $n \to \infty$,

$$\frac{\sqrt{\hat{k}_n^0}}{\hat{\gamma}(\hat{k}_n^0)\log(\frac{\hat{k}_n^0}{np_n})}\left(\frac{\hat{x}(\hat{k}_n^0)}{x_n} - 1\right) - \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}}$$

converges, in distribution, to a standard normal random variable. Therefore, as $n \to \infty$,

$$P\left(x_n < \hat{x}(\hat{k}_n^0)\left(1 + \frac{\hat{\gamma}(\hat{k}_n^0)\log(\frac{\hat{k}_n^0}{np_n})}{\sqrt{\hat{k}_n^0}}\left(-z_\alpha + \frac{\widehat{sign}}{\sqrt{-2\hat{\rho}}}\right)\right)^{-1}\right) \to 1-\alpha \tag{4.4.4}$$

which gives a one-sided asymptotic confidence interval for x_n, *with confidence coefficient* $1-\alpha$.

For a discussion on the accuracy of the confidence intervals, note that it follows similarly as in Section 4.2.

4.5 Simulations

We evaluate the performance of the confidence intervals for the tail index and high quantiles. For the estimation of the optimal sequence k_n we follow the bootstrap algorithm proposed by Danielsson et al. (2001). This bootstrap procedure also provides an estimator of the second order parameter, say $\hat{\rho}_1$. In appendix A we mention the main ideas behind this bootstrap procedure.

To estimate ρ we also use the following estimator proposed by Fraga Alves et al. (2001). Let $M_n^{(\alpha)}(k_n) = (1/k_n)\sum_{i=0}^{k_n-1}(\log X_{n-i,n} - \log X_{n-k,n})^\alpha$ and

$$T_n^{(1,2,3,0)}(k_n) = \frac{\log M_n^{(1)}(k_n) - \log\left(\frac{M_n^{(2)}(k_n)}{2}\right)/2}{\log\left(\frac{M_n^{(2)}(k_n)}{2}\right)/2 - \log\left(\frac{M_n^{(3)}(k_n)}{6}\right)/3}.$$

The estimator is given by

$$\hat{\rho}_2 = 3\frac{T_n^{(1,2,3,0)}(k_n) - 1}{T_n^{(1,2,3,0)}(k_n) - 3},$$

provided $1 \le T_n^{(1,2,3,0)} < 3$ (otherwise we shall say it is not defined). They proved consistency under condition (4.1.2) and for k_n (upper order statistics to use) satisfying $a(\frac{n}{k_n})\sqrt{k_n} \to \infty$.

To estimate the sign of the asymptotic bias we use (4.3.2) with $a_n = \log n$ and $b_n = c_n = n/\log\log n$.

We considered i.i.d. pseudo random numbers from the following distributions:

(1) Student-t distribution with $\nu = 1$ and 4 degrees of freedom, for which $\gamma = 1/\nu$, $\rho = -2/\nu$, $a(t) \sim (2/3)\Pi^2 t^{-2}$ if $\nu = 1$ and $a(t) \sim (5/24)\sqrt{16/3}\,t^{-1/2}$ if $\nu = 4$ (for the general formulas to obtain the scale constant of the function a we refer to Martins, 2000). Hence the sign of the bias is positive.

(2) Fréchet distribution, $F_{\mu,\sigma}(x) = \exp\{-((x-\mu)/\sigma)^{-1/\gamma}\}$, for which we have, if $\mu \neq 0$ and $0 < \gamma < 1$ then $\rho = -\gamma$ and $a(t) \sim -(\mu\gamma/\sigma)t^{-\gamma}$; if $\gamma = 1$ then $\rho = -1$ and $a(t) \sim (1/2 - \mu/\sigma)t^{-1}/\sigma$; if $\mu = 0$ or $\gamma > 1$ then $\rho = -1$ and $a(t) \sim (\gamma/2)t^{-1}$. We shall consider (μ, σ, γ) equal to (0,1,1) and (1,1,1). Then note that when $\gamma = 1$ the sign of the bias equals the sign$(1/2 - \mu/\sigma)$.

4.5.1 Simulations for the tail index

In Tables 4.1 and 4.2 are simulation results based on 500 samples of size $n = 2\,000$. In Table 4.1 are the boostrap results, namely mean and standard deviation of $\hat{k}_n^0$ and mean and root mean square error of the estimates of γ and ρ, and percentage of times that the estimator of the sign of the bias yielded the correct sign.

In Table 4.2 are the results on confidence intervals of size 98%, 96% and 90%. These where obtained from:

a) $\hat{k}_n^0 + \hat{\rho}_1$ - (4.2.6) where the estimates of ρ are from the bootstrap procedure,

b) $\hat{k}_n^0 + \hat{\rho}_2$ (4.2.6) where the estimates of ρ are from Fraga Alves et al. (2001),

c) $\hat{k}_n^0 + \rho$ - (4.2.6) with true ρ and sign(c),

d) $\hat{k}_n^0 + (\lambda = 0)$ - from (4.2.1) with $\lambda = 0$,

$\hat{k}_n^0$ mean	st.dev.	$\hat{\gamma}(\hat{k}_n^0)$ mean	rmse	$\hat{\rho}_1$ mean	rmse	$\hat{\rho}_2$ mean	rmse	$\widehat{\text{sign}}$ % true	$\hat{x}(\hat{k}_n^0)$ mean	rmse
$t_1 : \gamma = 1, \rho = -2, \text{sign} = +, x_n = 636.6$										
200.	117.	1.00	.16	-1.31	.81	-1.01	1.00	100.0	706.1[1]	353.5[1]
$t_4 : \gamma = .25, \rho = -.5, \text{sign} = +, x_n = 8.6$										
33.	37.	.29	.08	- .55	.21	-.69	.19	100.0	9.1	2.9
$\text{F}_{0,1} : \gamma = 1, \rho = -1, \text{sign} = +, x_n = 1999.5$										
414.	231.	1.03	.11	-2.13	1.44	-1.26	.31	99.6	2470.6	1161.5
$\text{F}_{1,1} : \gamma = 1, \rho = -1, \text{sign} = +, x_n = 2000.5$										
708.	247.	.94	.08	-3.58	2.79	-[2]	-[2]	91.0	1634.8	714.6

[1] without extreme quantile estimate of 378782281
[2] not defined in most of the samples

Table 4.1: Bootstrap estimates and percentage $\widehat{\text{sign}}$ equals the true sign, 500 samples of size 2 000 (see text for details).

e) $(\hat{k}_n^0)^{.8} + (\lambda = 0)$ - from (4.2.1) with $k_n = (\hat{k}_n^0)^{.8}$ and $\lambda = 0$. Note that $(\hat{k}_n^0)^{.8}$ is a rather arbitrary choice, the only requirement that it should be smaller (of smaller order) than $\hat{k}_n^0$.

For each confidence interval $[\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n}]$, its coverage probability $P(\gamma \in [\underline{\gamma}_{n,k_n}, \overline{\gamma}_{n,k_n}]) \sim 1-\alpha$ was checked. Furthermore it is checked if coverage is equally weighted in each tail, where for the left-hand side it is desirable that $P(\gamma < \underline{\gamma}_{n,k_n}) \sim \alpha/2$ and for the right-hand side that $P(\gamma < \overline{\gamma}_{n,k_n}) \sim 1-\alpha/2$. In Table 4.2 these are shown in the order: total coverage, left-hand side coverage and right-hand side coverage.

Comparing the confidence intervals $(\hat{k}_n^0 + \hat{\rho}_1)$ and $(\hat{k}_n^0 + \rho)$, with $(\hat{k}_n^0 + (\lambda = 0))$, in general we consider the first two better (particularly for α small), since for these the mean length are smaller (except for Fréchet (1,1)) and the coverage probabilities are usually much closer to those expected. The fact that the mean lengths are larger in the cases $(\hat{k}_n^0 + \hat{\rho}_1)$ and $(\hat{k}_n^0 + \rho)$ for Fréchet (1,1), is due to the negative sign of the bias.

Sometimes the confidence intervals with $\lambda = 0$ seem to give close coverage probabilities in the right tail but, note that a coverage probability of 100% is totally non-informative, a situation so often obtained for this cases. Indeed what we get are biased confidence intervals. For all distributions associated with positive bias the upper limit of the confidence intervals are so large that they are too often larger than the true value. On the other hand, the lower limits are so large that again, they are too often larger than the true value. Systematically we see that the contribution to the coverage probability, considering both sides, for being less than 100% comes always from the wrong coverage of the lower limit. Similar considerations can be made for the distribution associated with negative bias.

From these simulations we see that when the bias of Hill's estimator is positive, in general the right-hand side confidence intervals are quite precise; the same when the bias is negative and for left-hand side confidence intervals.

Distribution	$\alpha = 2\%$ mean center	mean length	Cov. Prob. 98%	1%	99%	$\alpha = 4\%$ mean center	mean length	Cov. Prob. 96%	2%	98%	$\alpha = 10\%$ mean center	mean length	Cov. Prob. 90%	5%	95%
t_1															
$\hat{k}_n^0 + \hat{\rho}_1$	.98	.36	91.	3.	95.	.97	.31	85.	6.	91.	.96	.25	72.	10.	82.
$\hat{k}_n^0 + \hat{\rho}_2$	.99	.37	92.	3.	95.	.97	.32	87.	4.	91.	.96	.25	75.	9.	83.
$\hat{k}_n^0 + \rho$	1.01	.40	94.	4.	98.	1.00	.34	89.	6.	95.	.98	.27	79.	10.	89.
$\hat{k}_n^0 + (\lambda = 0)$	1.07	.48	92.	8.	100.	1.05	.40	88.	11.	99.	1.03	.30	81.	15.	96.
$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.13	.77	99.	1.	100.	1.08	.64	97.	2.	99.	1.04	.48	91.	5.	97.
t_4															
$\hat{k}_n^0 + \hat{\rho}_1$	.29	.23	80.	14.	94.	.27	.19	74.	15.	89.	.26	.14	63.	20.	83.
$\hat{k}_n^0 + \hat{\rho}_2$	.30	.26	85.	14.	99.	.29	.21	82.	15.	97.	.27	.16	70.	20.	90.
$\hat{k}_n^0 + \rho$	.29	.24	85.	13.	98.	.28	.20	81.	14.	95.	.26	.15	68.	18.	86.
$\hat{k}_n^0 + \lambda = 0$	.43	.48	79.	21.	100.	.38	.36	75.	25.	100.	.34	.25	67.	33.	100.
$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	.48	.62	90.	10.	100.	.41	.46	87.	13.	100.	.35	.31	82.	17.	100.
$F_{0,1}$															
$\hat{k}_n^0 + \hat{\rho}_1$	1.02	.27	80.	17.	97.	1.01	.23	75.	20.	95.	1.00	.18	67.	23.	90.
$\hat{k}_n^0 + \hat{\rho}_2$	1.01	.27	82.	15.	97.	1.01	.24	78.	18.	95.	1.00	.19	68.	22.	90.
$\hat{k}_n^0 + \rho$	1.01	.27	83.	14.	97.	1.00	.23	78.	17.	95.	1.00	.19	69.	21.	90.
$\hat{k}_n^0 + (\lambda = 0)$	1.06	.32	79.	21.	100.	1.05	.27	76.	24.	100.	1.04	.21	67.	31.	97.
$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.10	.56	98.	2.	100.	1.07	.47	96.	3.	99.	1.05	.36	90.	9.	99.
$F_{1,1}$															
$\hat{k}_n^0 + \hat{\rho}_1$	.97	.19	69.	0.	70.	.96	.17	59.	1.	60.	.96	.13	46.	3.	49.
$\hat{k}_n^0 + \rho$	.98	.20	80.	1.	81.	.98	.17	73.	2.	75.	.97	.14	57.	4.	61.
$\hat{k}_n^0 + (\lambda = 0)$	.95	.18	61.	0.	61.	.95	.16	52.	0.	52.	.95	.13	43.	1.	44.
$(\hat{k}_n^0)^{.8} + (\lambda = 0)$	1.02	.37	98.	0.	98.	1.01	.32	96.	1.	96.	1.00	.26	90.	2.	92.

Table 4.2: Means of the center points and the lengths of the confidence intervals and estimated coverage probabilities, 500 samples of size 2 000 (see text for details).

Distribution	$\alpha = 2\%$ mean upp lim	Cov. Prob. 98%	$\alpha = 4\%$ mean upp lim	Cov. Prob. 96%	$\alpha = 10\%$ mean upp lim	Cov. Prob. 90%
t_1						
$\hat{k}_n^0 + \hat{\rho}_1$	1631.7	93.	1268.5	85.	951.6	70.
$\hat{k}_n^0 + \hat{\rho}_2$	1584.8	93.	1215.9	86.	916.4	69.
$\hat{k}_n^0 + \rho$	2225.6	96.	1434.0	92.	1022.6	77.
$\hat{k}_n^0 + (\lambda = 0)$	6693.6	98.	2731.7	98.	1477.6	93.
t_4						
$\hat{k}_n^0 + \hat{\rho}_1$	11.4	80.	10.6	73.	9.6	62.
$\hat{k}_n^0 + \hat{\rho}_2$	11.7	85.	10.9	80.	9.9	67.
$\hat{k}_n^0 + \rho$	11.3	82.	10.6	77.	9.6	64.
$\hat{k}_n^0 + (\lambda = 0)$	14.8	95.	13.5	93.	11.9	86.
$\mathrm{F}_{0,1}$						
$\hat{k}_n^0 + \hat{\rho}_1$	5012.8	94.	4141.0	91.	3299.2	84.
$\hat{k}_n^0 + \hat{\rho}_2$	4769.2	95.	3919.2	91.	3143.9	82.
$\hat{k}_n^0 + \rho$	4478.5	95.	3739.8	91.	3028.7	81.
$\hat{k}_n^0 + (\lambda = 0)$	9482.0	99.	6566.3	99.	4283.6	95.
$\mathrm{F}_{1,1}$						
$\hat{k}_n^0 + \hat{\rho}_1$	7045.5	87.	5099.8	81.	3274.5	66.
$\hat{k}_n^0 + \rho$	11181.1	97.	5779.6	93.	3753.8	80.
$\hat{k}_n^0 + (\lambda = 0)$	4209.4	83.	3801.4	75.	2697.5	53.

Table 4.3: Means of upper limits of the quantile confidence intervals and estimated coverage probabilities, 500 samples of size 2 000 (see text for details).

Our simulations indicate that the inclusion of second order information in the construction of the confidence intervals gives significant improvement.

4.5.2 Simulations for high quantiles

In Table 4.3 are the results on one-sided confidence intervals of size 98%, 96% and 90%. They are based on the same samples used in tail index estimation.

Similar considerations can be made on the confidence intervals for quantile. For instance note the insensitivity of the confidence intervals with null bias for t_3 and $F_{0,1}$, where the coverages remain the same whether $\alpha = 2\%$ or 4%. Note the very large upper confidence limits when compared with the others considering the bias information.

4.5.3 Additional considerations on the sign estimation

The sign estimator depends on the chosen values of a_n, b_n and c_n. We found that for many common distributions the choice of $a_n = \log n$ and $b_n = c_n = n/\log\log n$ is quite reasonable. Mainly we have just chosen simple sequences verifying the conditions of Theorem 4.3.1.

4.6 Data analysis

We started the paper by noting that it is well known that estimators which balance the asymptotic bias squared and variance yield the lower asymptotic mean square error. Nevertheless, in practice confidence bands are commonly based on the estimators evaluated at the asymptotically suboptimal number of order statistics (taking $\lambda = 0$), such that the factor with sign is omitted. Here we demonstrate the relevance of using the confidence bands for the quantiles using the optimal number of order statistics on actual data. It is shown that these can yield a considerable reduction in capital loss estimates.

We used daily price quotes over the period 1-1-1980 to 14-5-2002 on three quite different financial series, each of them comprising 5835 observations. The first contract is the US dollar per UK pound spot foreign exchange rate contract, abbreviated as the forex contract. The second series is the S&P500 total return index, and the third contract is the Dutch Nedlloyd share price. The latter contract is known to be very volatile due to the cyclical business of sea transport, while the US index is naturally better diversified and hence less volatile, compare e.g. the S&P and Nedlloyd quantile estimates at $p = 1/n$ given in Table 4.4. The forex contract is also of interest since forex risk is an important risk driver in international portfolios of pension funds. The daily price quotes p_t are used to compute daily continuously compounded returns r_t by taking the logarithmic first differences of the price series, i.e. $r_t = \ln(p_t/p_{t-1})$. Since forex data for currencies from countries with similar monetary policies are known to be symmetrically distributed, we used the absolute returns for the forex series (except for the few zero quotes). Stock returns generally exhibit a positive mean due to positive growth of the economy. Therefore for the stock return data we focussed on the loss returns only. The loss returns comprised approximately 50% of the data.

In Table 4.4 the tail parameter estimates are displayed. The gamma point estimates indicate that the number of bounded moments are between 3 and 5. We record both the bootstrap based rho estimate from Danielsson et al. (2001) as $\hat{\rho}_1$, and the one based on Fraga Alves et al. (2001) recorded as $\hat{\rho}_2$. It can be seen that these differ quite considerably, but as we will see later, this difference is not so important for the construction of the confidence bands as is the inclusion of the sign correction factor. Nonetheless it is worth mentioning the closeness of all the estimates obtained from $\hat{\rho}_2$. The subsample bootstrap estimates of the optimal number of order statistics $\hat{k}_n^0$ is on the low side for the first and third series. The procedure sometimes runs into boundary problems due to insufficient data. In case of the forex contract, the plot of the bootstrap constructed mean square error reveals the surface is very flat over the range between $k = 8$ and 20 approximately, so that the global minimum is difficult to locate. The mean square error plot for the S&P series reveals a unique and clearly identifiable minimum, while the forex and Nedlloyd mean square error plots have multiple local minima for small values of k.

The confidence bands for the tail index gamma are displayed in Table 4.5. We give three different bands at three different confidence levels (at the 2%, 4% and 10%

Series	$\hat{k}_n^0$	$\gamma(\hat{k}_n^0)$	ρ_1	ρ_2	$\widehat{sign}$	$x_{n-\hat{k}_n^0}$	$\hat{x}(\hat{k}_n^0)$ at $p = 1/n$
forex	8	0.201	-0.310	-0.639	+	0.028	0.043
S&P	225	0.308	-1.263	-0.711	+	-0.016	-0.090
Nedlloyd	12	0.255	-0.430	-0.725	+	-0.106	-0.201

Table 4.4: Parameter estimates

series	$\alpha = 2\%$		$\alpha = 4\%$		$\alpha = 10\%$	
	LL	UL	LL	UL	LL	UL
forex						
signcorr.	.09	.32	.09	.28	.10	.23
nocorr.	.11	1.14	.12	.74	.13	.48
signcorr.(FA)	.09	.41	.10	.34	.11	.28
S&P						
signcorr.	.26	.35	.26	.34	.26	.33
nocorr.	.27	.36	.27	.36	.28	.35
signcorr.(FA)	.25	.34	.26	.34	.27	.33
Nedlloyd						
signcorr.	.13	.40	.13	.36	.14	.31
nocorr.	.15	.78	.16	.63	.17	.49
signcorr.(FA)	.13	.45	.14	.40	.15	.33

Table 4.5: Tail index confidence bands

level respectively). The first band is the sign factor corrected (optimal asymptotic mean square error) band, the second is the zero λ based band used in most studies.[1] The third band is also sign factor corrected, but where $\hat{\rho}_2$ is used in constructing this band rather than $\hat{\rho}_1$ which is used in the first band. There are some differences between the first and last band, but the most glaring differences are in comparison with the second band. It appears that if one does not correct for the sign factor the confidence bands are considerably larger. This is basically due to a larger upper limit UL, the lower limits more or less all coincide. But the exception is the S&P series, where all are quite close. The latter is due to the larger ρ values, see (4.4.4) for the influence of the second order parameter ρ. On the other hand, the larger is $\hat{k}_n^0$ the lower is the influence of the second order components.

A confidence band for the quantile estimates hinges of the choice of the quantile. We decided to report the quantiles located at the border of the sample, i.e. we took $p = 1/n$. Results are in Table 4.6. As in the previous table we report three different type of bands. Since these are about the possible loss, we report the left one-sided confidence interval. To indicate that we worked with the absolute returns in case of the forex series, the loss quantiles are reported positively in this case. Again the band based on the zero λ presumption yields much higher loss limits at the desired confidence level. What does this mean economically speaking? Consider

[1] Some studies may on purpose prefer the estiamtes evaluated such that $\lambda = 0$, since the criterion function gives more (negative) weight to the asymptotic bias term. For these cases it is difficult to pick a specific number of order statistics, since such studies usually do not provide an automatic procedure for picking the number of order statistics. Hence, even if the sign factor is ignored in the construction of the confidence band, we still use the same number of order statistics as for the case when the sign factor is included.

series	$\hat{x}(\hat{k}_n^0)$ at $p = 1/n$	$x_{\alpha=2\%}$	$x_{\alpha=4\%}$	$x_{\alpha=10\%}$
forex	.043			
signcorr.		.048	.046	.043
nocorr.		.062	.053	.053
signcorr.(FA)		.052	.049	.045
S&P	-.090			
signcorr.		-.107	-.103	-.097
nocorr.		-.116	-.112	-.105
signcorr.(FA)		-.104	-.100	-.094
Nedlloyd	-.201			
signcorr.		-.245	-.229	-.209
nocorr.		-.323	-.297	-.263
signcorr.(FA)		-.259	-.242	-.219

Table 4.6: Quantile confidence bands

the case of Nedlloyd, where an investment bank has taken a stake of 10 million in the company. From the first column labeled "$\widehat{q}(\hat{k}_n^0)$ at $p = 1/n$" in Table 4.6 one sees that once per 22 years there is a day on which the investment bank loses two or more million of the ten million investment. But taking into account the uncertainty pertaining to this estimate, one has to add another half million at the 2% level if one uses the bias corrected band. The band without the correction term requires quite a bid more, i.e. at least 1.2 million extra! For the case of an index investor with 10 million invested in the S&P composite, the extra loss stemming from the use of the confidence band without correction factor is a more moderate (an extra hundred thousand).

Appendix A. Tail index and quantile bootstrap estimation

The adaptive bootstrap method proposed by Danielsson et al. (2001) is a two-step sub-sample bootstrap method. From a sample of size n, in a first step take r independent bootstrap sub-samples of size n_1, where n_1 must be of the order $n^{1-\varepsilon}$, $0 < \varepsilon < 1/2$. For the simulations we took ε of about .05 in all cases. For r we used 500. Let ${}^1\hat{\gamma}(k_n)$ and ${}^2\hat{\gamma}(k_n)$ be two consistent estimators of γ. Then, let

$$k_1^* = \text{argmin}_k \frac{1}{r} \sum_{i=1}^{r} \left({}^1\hat{\gamma}_i^*(k) - {}^2\hat{\gamma}_i^*(k) \right)^2$$

where ${}^1\hat{\gamma}_i^*$ and ${}^2\hat{\gamma}_i^*$ are the estimates based on the i-th bootstrap sub-sample of size n_1. In a second step, repeat step 1 but with n_1 replaced by $n_2 = n_1^2/n$, to get k_2^* say. Then, it is shown that

$$\hat{k_n^0} = \frac{(k_1^*)^2}{k_2^*} C(\hat{\gamma}, \hat{\rho})$$

${}^q\hat{k}_n^0$		$\hat{\gamma}({}^q\hat{k}_n^0)$		$\hat{\rho}_1$		$\hat{x}({}^q\hat{k}_n^0)$	
mean	st.dev.	mean	rmse	mean	rmse	mean	rmse
156.	130.	.94	.19	-1.08	1.08	621.0	332.4

Table 4.A: Bootstrap estimates, $t_1 : \gamma = 1, \rho = -2, \text{sign} = +, x_n = 636.6$.

is a consistent estimator of k_n^0, where $C(\gamma, \rho)$ is some known constant depending on γ and ρ. To estimate ρ it is shown that

$$\hat{\rho} = \frac{\log k_1^*}{-2\log n_1 + 2\log k_1^*}$$

is a consistent estimator.

When estimating quantiles a similar algorithm can be used, where ${}^1\hat{\gamma}(k_n)$ and ${}^2\hat{\gamma}(k_n)$ are replaced by quantile estimators. Still, since $k_n^0/{}^qk_n^0 \sim 1$ both procedures with gamma or quantile estimators provide a consistent estimator of ${}^qk_n^0$.

In Table 4.A are the bootstrap results from the algorithm using the quantile estimators, for t_1 distribution. Compare this results with the ones in Table 4.1. The quantile estimates in Table 4.A are more accurate, but the ρ (and γ) estimates are less accurate. It turns out that in terms of confidence intervals, where all this estimates are used, in general we find that the use of the algorithm with quantiles brings no significant improvement.

Appendix B. Penultimate approximation of H_{n,k_n}

We shall need to extend condition (4.1.2) up to one more order. That is, suppose there exists a function b_1, with constant sign near infinity and $\lim_{t\to\infty} b_1(t) = 0$, such that

$$\lim_{t\to\infty} \frac{\frac{\log U(tx) - \log U(t) - \gamma\log x}{a(t)} - \frac{x^\rho - 1}{\rho}}{b_1(t)} = L_{\rho,\eta}(x), \quad \text{for all } x > 0,\ \eta < 0 \qquad (4.B.1)$$

where

$$L_{\rho,\eta}(x) = \frac{1}{\eta}\left(\frac{x^{\rho+\eta} - 1}{\rho+\eta} - \frac{x^\rho - 1}{\rho}\right).$$

The following lemma is an immediate consequence of Lemma 2.1 in Drees (1998a); see also Appendix A in de Haan and Pereira (1999).

Lemma 4.B.1. *Suppose* (4.B.1). *Then there exists a function $b(t)$ with $b(t)/b_1(t) \to 1$, as $t \to \infty$, with the property that for every $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 > 0$ there exists a t_0 such that for $t \ge t_0$, $tx \ge t_0$*

$$\lim_{t\to\infty} \sup_{x\ge 1} x^{-\gamma-\rho-\varepsilon_1} \left| \frac{\frac{\log U(tx) - \log U(t) - \gamma\log x}{a(t)} - \frac{x^\rho - 1}{\rho}}{b(t)} - L_{\rho,\eta}(x) \right| < \varepsilon_2 \qquad (4.B.2)$$

and

$$\lim_{t\to\infty}\sup_{x\geq 1} x^{-\rho-\eta-\varepsilon_3}\left|\frac{\frac{a(tx)}{a(t)}-x^\rho}{b(t)}-x^\rho\frac{x^\eta-1}{\eta}\right|<\varepsilon_4. \tag{4.B.3}$$

We are interested in results for the optimal sequence k_n^0 in (4.2.3). For simplicity we will give results only when the equality in (4.2.3) holds and for auxiliary functions $b(t)\sim\ c_b t^\eta$, as $t\to\infty$. Still, the main lines of the proofs can be adapted to the more general cases. Therefore, in this appendix we define

$$k_n^0=[\gamma^2(1-\rho)^2/(-2\rho\, c^2)]^{1/(1-2\rho)}\ n^{-2\rho/(1-2\rho)}. \tag{4.B.4}$$

Then

$$a(n/k_n^0)\sqrt{k_n^0}\to\frac{\operatorname{sign}(c)\gamma(1-\rho)}{\sqrt{-2\rho}},\qquad \text{as } n\to\infty. \tag{4.B.5}$$

Let f_n be defined by

$$f_n^{-1}\left[a(n/k_n^0)\sqrt{k_n^0}-\frac{\operatorname{sign}(c)\gamma(1-\rho)}{\sqrt{-2\rho}}\right]\to\Delta_1(\neq 0). \tag{4.B.6}$$

Then $b(n/k_n^0)/f_n\to\Delta_2\ (\neq 0)$.

Lemma 4.B.2. *If k_n^0 satisfies* (4.B.4) *and f_n satisfies* (4.B.6) *then, for $\eta>\rho$, there exists a positive sequence $t_n\to 0$ such that*

$$\lim_{n\to\infty}\frac{P\left(\left|k_n^0 Y_{n-k_n^0,n}/n-1\right|\leq t_n\right)}{f_n}=0, \tag{4.B.7}$$

where $\{Y_{i,n}\}_{i=1}^n$ are the n-th order statistics from the distribution function $1-x^{-1}$, $x>1$.

Proof. Take $t_n=n^{(\eta-\delta/2)/(1-2\rho)}$, $0<\delta<2(\eta-\rho)$. Then the result follows easily from Lemma 2 in Cheng and de Haan (2001), □

Theorem 4.B.1 (Gamma approximation). *Assume* (4.B.1) *and let k_n^0 and f_n be as in* (4.B.4) *and* (4.B.6)*, respectively. Then, for $\eta>\rho$,*

$$\lim_{n\to\infty}\frac{P\left(H_{n,k_n^0}\leq x\right)-\Gamma_{k_n^0}\left(k_n^0+\left(x-\frac{sign(c)}{\sqrt{-2\rho}}\right)\sqrt{k_n^0}\right)}{f_n}$$
$$=\left(\frac{\Delta_1}{\gamma(\rho-1)}+\frac{\Delta_2\, sign(c)}{\sqrt{-2\rho}(\rho+\eta-1)}\right)\phi\left(x-\frac{sign(c)}{\sqrt{-2\rho}}\right) \tag{4.B.8}$$

uniformly for all x, where Γ_k is the gamma distribution function with k degrees of freedom and ϕ is the standard normal density function.

Proof. Although we need more refined assumptions, many of the ingredients for the proof were taken from Cheng and de Haan (2001). As in this paper we use the representation $\{X_i\}_{i=1}^{\infty} \stackrel{d}{=} \{U(Y_i)\}_{i=1}^{\infty}$ and assume $\hat{\gamma}(k) = (1/k)\sum_{i=0}^{k-1} \log U(Y_{n-i,n}) - \log U(Y_{n-k,n})$. For simplicity of notation, we will wright k for k_n^0 throughout.

Using the same arguments as in Cheng and de Haan (2001), if we condition $P(H_{n,k} \leq x)$ on the event $|kY_{n-k,n}/n - 1| \leq t_n$, $t_n \to 0$ positive, then to evaluate

$$\lim_{n\to\infty} f_n^{-1} \left\{ P(H_{n,k} \leq x) - \Gamma_k \left(k + \sqrt{k}x - \operatorname{sign}(c)\sqrt{-2\rho} \right) \right\}$$

is enough to consider

$$\lim_{n\to\infty} \frac{P\left(H_{n,k} \leq x \,\middle|\, |kY_{n-k,n}/n - 1| \leq t_n\right) - \Gamma_k\left[k + \sqrt{k}(x - \operatorname{sign}(c)/\sqrt{-2\rho})\right]}{f_n}.$$

Replace t by $Y_{n-k,n}$ and x by $Y_{n-i,n}/Y_{n-k,n}$. Then from Lemma 4.B.1, eventually for $0 \leq i \leq k-1$,

$$\left| \frac{\frac{\log U(Y_{n-i,n}) - \log U(Y_{n-k,n}) - \gamma \log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)}{a(Y_{n-k,n})} - \frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\rho} - 1}{\rho}}{b(Y_{n-k,n})} - L_{\rho,\eta}\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right) \right|$$

$$< \varepsilon_2 \left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\gamma+\rho+\varepsilon_1} \tag{4.B.9}$$

If $|kY_{n-k,n}/n - 1| \leq t_n$ then

$$1 + \rho t_n(1 + o(1)) \leq \left(\frac{kY_{n-k,n}}{n}\right)^{\rho} \leq 1 - \rho t_n(1 + o(1))$$

Moreover as in Cheng and de Haan (2001), if n is so large that $(1 - t_n)^{\eta-\varepsilon} \leq 1 + \varepsilon$ and $(1 + t_n)^{\eta-\varepsilon} \geq 1 - \varepsilon$, since $b(t) \in \mathrm{RV}_\eta$,

$$(1-\varepsilon)^2 b(n/k) \leq b(Y_{n-k,n}) \leq (1+\varepsilon)^2 b(n/k).$$

Also from (4.B.3), with t replaced by n/k and x by $kY_{n-k,n}/n$, we get the inequalities

$$a(\frac{n}{k}) \left[b(\frac{n}{k}) \left\{ -\varepsilon_4(1 - (\rho + \eta + \varepsilon_3)t_n) + \frac{\varepsilon}{\eta}(1 + \rho t_n) \right\} + 1 + \rho t_n + o(t_n) \right]$$

$$\leq a(Y_{n-k,n}) \leq$$

$$a(\frac{n}{k}) \left[b(\frac{n}{k}) \left\{ \varepsilon_4(1 - (\rho + \eta + \varepsilon_3)t_n) - \frac{\varepsilon}{\eta}(1 - \rho t_n) \right\} + 1 - \rho t_n + o(t_n) \right].$$

Hence for $|kY_{n-k,n}/n-1| \le t_n$, we get by summing over k in (4.B.9)

$$\frac{a(n/k)}{\sqrt{k}}\left[b(\frac{n}{k})\left\{-\varepsilon_4(1-(\rho+\eta+\varepsilon_3)t_n)+\frac{\varepsilon}{\eta}(1+\rho t_n)\right\}+1+\rho t_n+o(t_n)\right]$$

$$\left[\sum_{i=0}^{k-1}\frac{\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\rho}-1}{\gamma\rho}+\frac{(1-\varepsilon)^2}{\gamma}b(\frac{n}{k})\sum_{i=0}^{k-1}\left\{L_{\rho,\eta}\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)-\varepsilon_2\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)^{\gamma+\rho+\varepsilon_1}\right\}\right]$$

$$\le \quad H_{n,k}-\frac{1}{\sqrt{k}}\sum_{i=0}^{k-1}\left\{\log\left(\frac{Y_{n-i,n}}{Y_{n-k,n}}\right)-1\right\}.$$

Similarly for the upper inequality. Define

$$\begin{aligned}
\hat{\Gamma}_k \quad = \quad & \sqrt{k}\left(\frac{1}{k}\sum_{i=0}^{k-1}E_i-1\right)+\frac{\text{sign}(c)}{\sqrt{-2\rho}}+\left[a(\frac{n}{k})\sqrt{k}-\frac{\text{sign}(c)\gamma(1-\rho)}{\sqrt{-2\rho}}\right]\frac{1}{\gamma(1-\rho)} \\
&+a(\frac{n}{k})\sqrt{k}\,t_n\frac{\rho}{\gamma(1-\rho)}(1+o(1)) \\
&+a(\frac{n}{k})\sqrt{k}\,b(\frac{n}{k})\left\{\frac{-\varepsilon_4+\varepsilon/\eta}{\gamma(1-\rho)}+\frac{(1-\varepsilon)^2}{\gamma}\right. \\
&\qquad\left.\left[-\frac{\varepsilon_2}{1-\gamma-\rho-\varepsilon_1}+\frac{1}{(1-\rho-\eta)(1-\rho)}\right]\right\}(1+o(1))
\end{aligned}$$

and

$$\begin{aligned}
Q_k \quad = \quad & \frac{1}{\sqrt{k}}\sum_{i=0}^{k-1}\left[\frac{\exp(\rho E_i)-1}{\gamma\rho}-\frac{1}{\gamma(1-\rho)}\right] \\
&+\frac{b(n/k)}{\sqrt{k}}\sum_{i=0}^{k-1}\left\{\frac{(1-\varepsilon)^2}{\gamma}\left[-\varepsilon_2\exp((\gamma+\rho+\varepsilon_1)E_i)+\frac{\varepsilon_2}{1-\gamma-\rho-\varepsilon_1}\right.\right. \\
&\left.+\frac{\exp((\rho+\eta)E_i)-1}{\eta(\rho+\eta)}-\frac{1}{\eta(1-\rho-\eta)}-\frac{\exp(\rho E_i)-1}{\eta\rho}+\frac{1}{\eta(1-\rho)}\right] \\
&\left.+(-\varepsilon_4+\frac{\varepsilon}{\eta})\left[\frac{\exp(\rho E_i)-1}{\gamma\rho}-\frac{1}{\gamma(1-\rho)}\right]\right\}(1+o_p(1)) \\
&+\frac{\rho\,t_n}{\sqrt{k}}\sum_{i=0}^{k-1}\left[\frac{\exp(\rho E_i)-1}{\gamma\rho}-\frac{1}{\gamma(1-\rho)}\right](1+o_p(1)),
\end{aligned}$$

where $E_0, E_1, \ldots, E_{k-1}$ are i.i.d. standard exponential. Hence

$$\begin{aligned}
&P\left\{H_{n,k}\le x \,\middle|\, |kY_{n-k,n}/n-1|\le t_n\right\} \\
&\quad\le P\left(\hat{\Gamma}_k\le x+\varepsilon\sqrt{k}a(n/k)f_n\right)+P\left(Q_k\le-\varepsilon\sqrt{k}f_n\right),
\end{aligned}$$

and the rest of the proof follows by the same arguments as in Cheng and de Haan (2001). □

Corollary 4.B.1 (Normal approximation). *Theorem 4.B.1 is still valid if one replaces in* (4.B.8), $\Gamma_{k_n^0}(k_n^0 + \cdot\sqrt{k_n^0})$ *by* $\Phi(\cdot)$, *where* Φ *is the standard normal distribution function.*

Proof. The result follows from Cramér's theorem, which reads in our case as, $\Gamma_k(k+x\sqrt{k}) = \Phi(x) + (1-x^2)\phi(x)/(3\sqrt{k} + o(1/\sqrt{k})$, uniformly for all x, as $k \to \infty$, and the fact that $f_n\sqrt{k_n^0} \to \infty$. □

Cheng and Pan (1998) and Cheng and de Haan (2001) gave one-term Edgeworth expansions for the distribution of Hill's estimator, the first with a normal approximation and the second with a gamma approximation. Both papers only consider the situation $\sqrt{k_n}a(n/k_n) \to 0$. Although their conditions on the growth of the intermediate sequence k_n are not exactly the same, sometimes they overlap. For instance when $k_n a(n/k_n) \to 0$ and $\log n = o(k_n)$ the gamma approximation turns out to have a better rate (for more details and other cases see these papers).

More recently Cuntz and Haeusler (2001) gave an Edgeworth expansion with normal approximation and with no further conditions on k_n than just $\sqrt{k_n}a(n/k_n) \to \lambda \in [0, \infty)$. Comparing their results with ours, it turns out that when $\sqrt{k_n}a(n/k_n) \to \lambda \in (0, \infty)$ the gamma and normal approximations have the same rate. In Theorem 4.B.1 we give the gamma approximation when the second and third order parameters verify $\rho < \eta$. To check that indeed Corollary 4.B.1 coincides with their Edgeworth expansion with k_n^0, take their constants C, C_1 and C_2 equal to, respectively, 1, $1/\eta$ and $-1/\eta$. For $\eta \le \rho$ the proof of Theorem 4.B.1 indicates that we would obtain a normal instead of gamma approximation, which leads to no improvement towards the Cuntz and Haeusler results.

Chapter 5

On maximum likelihood estimation of the extreme value index

Co-authors: Holger Drees and Laurens de Haan

Abstract. We prove asymptotic normality of the so-called maximum likelihood estimator of the extreme value index.

5.1 Introduction

Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) random variables (r.v.'s), from some unknown distribution function (d.f.) F. Denote the upper endpoint of F by x^*, where $x^* = \sup\{x : F(x) < 1\} \leq \infty$, and let

$$F_t(x) = P(X < t + x | X > t) = \frac{F(t+x) - F(t)}{1 - F(t)}, \tag{5.1.1}$$

with $1 - F(t) > 0$, $t < x^*$ and $x > 0$, be the conditional d.f. of $X - t$ given $X > t$. Then it is well known (Balkema and de Haan, 1974; Pickands, 1975) that up to scale and location transformations the generalised Pareto d.f., given by,

$$H_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma}, \tag{5.1.2}$$

$x > 0$ if $\gamma \geq 0$ and $0 < x < -1/\gamma$ if $\gamma < 0$ (for $\gamma = 0$ read $(1+\gamma x)^{-1/\gamma}$ as $\exp(-x)$), can provide a good approximation of conditional probabilities like (5.1.1). More precisely, it has been proved that there exists a normalising function $\sigma(t) > 0$, such that

$$\lim_{t \to x^*} F_t(x\sigma(t)) \to H_\gamma(x)$$

for all x, or equivalently

$$\lim_{t \to x^*} \sup_{0<x<x^*-t} |F_t(x) - H_\gamma(x/\sigma(t))| = 0 \tag{5.1.3}$$

if and only if F is in the maximum domain of attraction of an Extreme Value d.f. (Gnedenko, 1943), commonly denoted by $F \in D(G_\gamma)$.

Under this set-up it comes out that a major issue for estimating extreme events is the estimation of the extreme value index γ. A variety of procedures to estimate it are now available in the literature (e.g. Hill, 1975; Dekkers et al., 1989; Smith, 1987), although there are still open problems. Quite often the accuracy of these estimators rely heavily on the choice of some threshold but, it is not our aim here to address this type of optimality questions.

The maximum likelihood estimator (m.l.e.) of γ is one of the most common ones. For a sample of size n let $X_{1,n} \leq X_{2,n} \leq \dots X_{n,n}$ be the ascending order statistics. From the GP approximation, it becomes clear that we shall base our inferences on some higher order statistics, say $(X_{n-k,n}, X_{n-k+1,n}, \dots, X_{n,n})$. Consider the one to one mapping

$$\begin{cases} Y_0 & = X_{n-k,n} \\ Y_1 & = X_{n-k+1,n} - X_{n-k,n} \\ \dots & \\ Y_k & = X_{n,n} - X_{n-k,n}. \end{cases}$$

Any inference based on maximizing the likelihood of $(X_{n-k,n}, X_{n-k+1,n}, \dots, X_{n,n})$ is equivalent to considering the likelihood of $(Y_0, Y_1, \dots, Y_k)$. Now the distribution of $(Y_1, \dots, Y_k)$ given Y_0 equals the distribution of $(Y^*_{1,k}, \dots, Y^*_{k,k})$, where these are the order statistics of an i.i.d. sample $(Y^*_1, \dots, Y^*_k)$ with common distribution $F_{y_0}(x) = P(X < y_0 + x | X > y_0)$ - see e.g. Theorem 2.4.1 in Arnold et al. (1992). Then the common approach is: first to factorise the distribution of $(Y^*_1, \dots, Y^*_k)$ and to use the GP approximation from (5.1.3), second to ignore the factor related to the marginal distribution of Y_0 in the likelihood of $(Y_0, Y_1, \dots, Y_k)$. Summing up, given some threshold y_0, the so-called m.l.e. of γ (and σ) is obtained by maximizing in γ (and σ), $\prod_{i=1}^k h(y_1, \dots, y_k)$ where $h_{\gamma,\sigma}(y) = \partial H_{\gamma,\sigma}(y)/\partial y$. The m.l. equations are then based on

$$\begin{cases} \frac{\partial \log h_{\gamma,\sigma}(y)}{\partial \gamma} = \frac{1}{\gamma^2} \log\left(1 + \frac{\gamma}{\sigma} y\right) - \left(\frac{1}{\gamma} + 1\right) \frac{\frac{y}{\sigma}}{1+\frac{\gamma}{\sigma} y} = 0 \\ \frac{\partial \log h_{\gamma,\sigma}(y)}{\partial \sigma} = -\frac{1}{\sigma} - \left(\frac{1}{\gamma} + 1\right) \frac{-\frac{\gamma}{\sigma^2} y}{1+\frac{\gamma}{\sigma} y} = 0, \end{cases}$$

where for $\gamma = 0$ these equations should be interpreted as

$$\begin{cases} \frac{1}{2}\left(\frac{y}{\sigma}\right)^2 - \frac{y}{\sigma} = 0 \\ -\frac{1}{\sigma} + \frac{y}{\sigma^2} = 0. \end{cases}$$

The resulting m.l. equations are as follows (with a similar interpretation when

$\gamma = 0$),

$$\begin{cases} \sum_{i=1}^{k} \frac{1}{\gamma^2} \log\left(1 + \frac{\gamma}{\sigma}(X_{n-i+1,n} - X_{n-k,n})\right) - \left(\frac{1}{\gamma} + 1\right) \frac{\frac{1}{\sigma}(X_{n-i+1,n} - X_{n-k,n})}{1+\frac{\gamma}{\sigma}(X_{n-i+1,n} - X_{n-k,n})} = 0 \\ \sum_{i=1}^{k} \left(\frac{1}{\gamma} + 1\right) \frac{\frac{\gamma}{\sigma}(X_{n-i+1,n} - X_{n-k,n})}{1+\frac{\gamma}{\sigma}(X_{n-i+1,n} - X_{n-k,n})} = k \end{cases} \tag{5.1.4}$$

which can be simplified to, if $\gamma \neq 0$,

$$\begin{cases} \frac{1}{k} \sum_{i=1}^{k} \log\left(1 + \frac{\gamma}{\sigma}(X_{n-i+1,n} - X_{n-k,n})\right) = \gamma \\ \frac{1}{k} \sum_{i=1}^{k} \frac{1}{1+\frac{\gamma}{\sigma}(X_{n-i+1,n} - X_{n-k,n})} = \frac{1}{\gamma+1}, \end{cases}$$

and the maximization is over $(\gamma, \sigma) \in (-1/2, \infty) \times (0, \infty)$.

From the above reasoning it follows that the m.l.e. of γ is shift and scale invariant, and the m.l.e. of σ is shift invariant.

Next we sketch the proof of the asymptotic normality. Under the given - usual - conditions (see (5.2.1), $n \to \infty$, k_n intermediate sequence) we have

$$\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)_{t\in[0,1]} = \left(\frac{t^{-\gamma_0} - 1}{\gamma_0} + k_n^{-1/2} Y_n(t)\right)_{t\in[0,1]} \tag{5.1.5}$$

where $(Q_n(t))_{t\in[0,1]}$ is a distributionally equivalent version of the process $(X_{n-[k_n t],n})_{t\in[0,1]}$, $\{Y_n(t)\}$ is an asymptotically Gaussian process of known structure (Lemma 5.3.1), γ_0 is the true parameter and $\tilde{a}$ is a suitably chosen positive function. Hence for all $t \in [0,1]$

$$\begin{aligned} & t^{\gamma_0}\left(1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right) \\ = \; & 1 + \left(\frac{\gamma}{\tilde{\sigma}} - \gamma_0\right) \frac{1 - t^{\gamma_0}}{\gamma_0} + t^{\gamma_0} \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} Y_n(t) \end{aligned} \tag{5.1.6}$$

where $\tilde{\sigma} = \sigma/\tilde{a}(\frac{k_n}{n})$. Now if the sequence of solutions $(\gamma, \tilde{\sigma})$ satisfies

$$\gamma - \gamma_0 = O_p(k_n^{-1/2}) \quad \text{and} \quad \tilde{\sigma} - 1 = O_p(k_n^{-1/2}),$$

one can prove by using a construction similar to (5.1.5) that

$$C = \inf_{1/(2k_n) \le t \le 1} t^{\gamma_0} \left(1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right) > 0$$

with probability tending to 1 (Lemma 5.3.2). This implies (use (5.1.6) for the first equality)

$$\begin{aligned}
&\log\left(t^{\gamma_0}\left(1+\frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)\right)\\
&= \log\left(1+\left(\frac{\gamma}{\tilde{\sigma}}-\gamma_0\right)\frac{1-t^{\gamma_0}}{\gamma_0}+t^{\gamma_0}\frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}Y_n(t)\right)\\
&= \left(\frac{\gamma}{\tilde{\sigma}}-\gamma_0\right)\frac{1-t^{\gamma_0}}{\gamma_0}+t^{\gamma_0}\frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}Y_n(t)+o_p(k_n^{-1/2})
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{1+\frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}\\
&= t^{\gamma_0}\left(1-\left(\frac{\gamma}{\tilde{\sigma}}-\gamma_0\right)\frac{1-t^{\gamma_0}}{\gamma_0}-t^{\gamma_0}\frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}Y_n(t)+o_p(k_n^{-1/2})\right)
\end{aligned}$$

where the o_p-term is uniform for $1/(2k_n) \le t \le 1$ (proof of Proposition 5.3.1).

Hence, up to a $o_p(k_n^{-1/2})$-term we are able to turn the equations (5.1.4) into linear equations which can be solved readily. The proof in case $\gamma_0 = 0$ requires longer expansions but is similar. The statement is in Theorem 5.2.1. In Theorem 5.2.2 an equivalent explicit estimator is constructed for the case $\gamma_0 = 0$.

Proofs of the asymptotic normality of the m.l.e.'s of γ and σ are given in Smith (1987) and also in Drees (1998a) in the case $\gamma > 0$. Nonetheless we consider that some proofs are not easily understandable or that some of the conditions are somewhat restrictive. In this paper we present a relatively simple direct approach to prove asymptotic normality of the m.l.e.'s of γ and σ. They are based on some recent results on strong approximation on the empirical tail quantile function (Drees, 1998a).

Throughout we shall denote $F^{\leftarrow}$ the generalised inverse of F, $\to_d$ convergence in distribution and $\to_p$ convergence in probability.

5.2 Asymptotic normality of the maximum likelihood estimators

Assume that there exist measurable, locally bounded functions a, $\Phi : (0,1) \to (0,\infty)$ and $\Psi : (0,\infty) \to \mathbb{R}$ such that

$$\lim_{t\downarrow 0}\frac{\frac{F^{\leftarrow}(1-tx)-F^{\leftarrow}(1-t)}{a(t)}-\frac{x^{-\gamma_0}-1}{\gamma_0}}{\Phi(t)} = \Psi(x), \tag{5.2.1}$$

for some $\gamma_0 > -1/2$, for all $t \in (0,1)$ and $x > 0$, where $x \mapsto \Psi(x)/(x^{-\gamma_0}-1)$ is not constant, $\Phi(t)$ not changing sign eventually and $\Phi(t) \to 0$ as $t \downarrow 0$. Then, according to de Haan and Stadtmüller (1996), $|\Phi|$ is $-\rho$-varying at 0 for some $\rho \le 0$, i.e. $\lim_{t\downarrow 0} \Phi(tx)/\Phi(t) = x^{-\rho}$ for all $x > 0$, and

$$\Psi(x) = \begin{cases} (x^{-(\gamma_0+\rho)} - 1)/(\gamma_0 + \rho) & , \rho < 0 \\ -x^{-\gamma_0} \log(x)/\gamma_0 & , \gamma_0 \neq \rho = 0 \\ \log^2(x) & , \gamma_0 = \rho = 0, \end{cases} \tag{5.2.2}$$

provided that the normalising function a and the function Φ are chosen suitably. Condition (5.2.1) is a second order refinement of $F \in D(G_{\gamma_0})$. Still, it is a quite general condition, satisfied for all usual distributions satisfying the max-domain of attraction condition.

We assume througout that k_n is an intermediate sequence, i.e. $k_n \to \infty$ and $k_n/n \to 0$, as $n \to \infty$.

Theorem 5.2.1. *Assume condition* (5.2.1) *for some* $\gamma_0 > -1/2$ *and that the intermediate sequence* k_n *satisfies*

$$\Phi(k_n/n) = O(k_n^{-1/2}). \tag{5.2.3}$$

Then, the system of m.l. equations (5.1.4) *has a sequence of solutions* $(\hat{\gamma}_n, \hat{\sigma}_n)$ *that verifies, as* $n \to \infty$,

$$\begin{aligned} & k_n^{1/2}(\hat{\gamma}_n - \gamma_0) - \frac{(\gamma_0+1)^2}{\gamma_0} k_n^{1/2} \Phi(\frac{k_n}{n}) \int_0^1 \left(t^{\gamma_0} - (2\gamma_0+1)t^{2\gamma_0}\right) \Psi(t)dt \\ \to_d \quad & \frac{(\gamma_0+1)^2}{\gamma_0} \int_0^1 \left(t^{\gamma_0} - (2\gamma_0+1)t^{2\gamma_0}\right) \left(W(1) - t^{-(\gamma_0+1)}W(t)\right) dt \end{aligned} \tag{5.2.4}$$

$$\begin{aligned} & k_n^{1/2} \left(\frac{\hat{\sigma}_n}{a(\frac{k_n}{n})} - 1 \right) - \frac{\gamma_0+1}{\gamma_0} k_n^{1/2} \Phi(\frac{k_n}{n}) \int_0^1 \left((\gamma_0+1)(2\gamma_0+1)t^{2\gamma_0} - t^{\gamma_0}\right) \Psi(t)dt \\ \to_d \quad & \frac{\gamma_0+1}{\gamma_0} \int_0^1 \left((\gamma_0+1)(2\gamma_0+1)t^{2\gamma_0} - t^{\gamma_0}\right) \left(W(1) - t^{-(\gamma_0+1)}W(t)\right) dt, \end{aligned} \tag{5.2.5}$$

where for $\gamma_0 = 0$ *these equations should be interpreted as their limits when* $\gamma_0 \to 0$, *i.e. for* $\gamma_0 = 0$,

$$\begin{aligned} & k_n^{1/2}\hat{\gamma}_n + k_n^{1/2}\Phi(\frac{k_n}{n}) \int_0^1 (2 + \log t)\Psi(t)dt \\ \to_d \quad & -\int_0^1 (2 + \log t) \left(W(1) - t^{-1}W(t)\right) dt \end{aligned} \tag{5.2.6}$$

$$\begin{aligned} & k_n^{1/2} \left(\frac{\hat{\sigma}_n}{a(\frac{k_n}{n})} - 1 \right) - k_n^{1/2}\Phi(\frac{k_n}{n}) \int_0^1 (3 + \log t)\Psi(t)dt \\ \to_d \quad & \int_0^1 (3 + \log t) \left(W(1) - t^{-1}W(t)\right) dt, \end{aligned} \tag{5.2.7}$$

where W is a standard Brownian motion. Moreover, any other sequence of solutions $(\hat{\gamma}_n^*, \hat{\sigma}_n^*)$ satisfies $k_n^{1/2}|\hat{\gamma}_n^* - \gamma_0| \to_p \infty$ or $k_n^{1/2}|\hat{\sigma}_n^*/a(k_n/n) - 1| \to_p \infty$.

Remark 5.2.1. The term $\frac{(\gamma_0+1)^2}{\gamma_0} k_n^{1/2}\Phi(\frac{k_n}{n}) \int_0^1 \left(t^{\gamma_0} - (2\gamma_0+1)t^{2\gamma_0}\right)\Psi(t)dt$ in (5.2.4) is a bias term. It vanishes if $k_n^{1/2}\Phi(\frac{k_n}{n}) \to 0$. Similarly for (5.2.5)-(5.2.7).

Remark 5.2.2. Note that the m.l. equations for $\gamma = 0$ lead to $\sum_{i=1}^k X_i^2/(2k) = (\sum_{i=1}^k X_i/k)^2$.

Corollary 5.2.1. *Under the conditions of Theorem 5.2.1 and if*

$$k_n^{1/2}\Phi(\frac{k_n}{n}) \to \lambda \in \mathbb{R}, \tag{5.2.8}$$

the solutions (5.2.4)-(5.2.7) *verify*

$$k_n^{1/2}\begin{bmatrix} \hat{\gamma}_n - \gamma_0 \\ \hat{\sigma}_n/a(k_n/n) - 1 \end{bmatrix} \to_d N(\lambda\mu, \Sigma), \tag{5.2.9}$$

where N denotes the bivariate normal distribution, μ equals

$$\begin{array}{ll} \left[\frac{\rho(\gamma_0+1)}{(1-\rho)(\gamma_0-\rho+1)}, \frac{1-2\rho+\gamma_0-\rho\gamma_0}{(1-\rho)(\gamma_0-\rho+1)}\right]^T & , \text{if } \rho < 0, \\ [1, \gamma_0^{-1}]^T & , \text{if } \gamma_0 \neq \rho = 0, \\ [2, 0]^T & , \text{if } \gamma_0 = \rho = 0 \end{array}$$

and

$$\Sigma = \begin{bmatrix} (1+\gamma_0)^2 & -(1+\gamma_0) \\ -(1+\gamma_0) & 2+2\gamma_0+\gamma_0^2 \end{bmatrix}.$$

Remark 5.2.3. When comparing our results to those of Smith (1987) we see that the covariance matrix is the same except for the variance of the scale estimator. It is more difficult to compare the bias in both papers since the set-up is somewhat different.

We now show that if $\gamma_0 = 0$ the m.l.e.'s are equivalent in some sense to explicit estimators. Define

$$m_n^{(j)} = \frac{1}{k_n}\sum_{i=0}^{k_n-1}(X_{n-i,n} - X_{n-k_n,n})^j, \quad j = 1, 2.$$

Define

$$\hat{\gamma}_* = 1 - \frac{1}{2}(1 - \frac{(m_n^{(1)})^2}{m_n^{(2)}})^{-1}$$

and

$$\hat{a}_*(\frac{k_n}{n}) = \frac{2(m_n^{(1)})^3}{m_n^{(2)}}.$$

Let $(\hat{\gamma}_{MLE}, \hat{\sigma}_{MLE})$ be a sequence of solutions of (5.1.4) satisfying Theorem 5.2.1.

Theorem 5.2.2. *If F is in the class of distributions that satisfy* (5.2.1) *with* $\gamma_0 = 0$ *and if* (5.2.3) *holds, then*

$$k_n^{1/2}\left(\hat{\gamma}_* - \hat{\gamma}_{MLE}\right) \to_p 0$$

and

$$k_n^{1/2}\left(\frac{\hat{a}_*(\frac{k_n}{n}) - \hat{\sigma}_{MLE}}{a(\frac{k_n}{n})}\right) \to_p 0.$$

Remark 5.2.4. If, in addition, (5.2.1) holds with $\rho < 0$ and if $\sup\{x|F(x) < 1\} > 0$ then, provided

$$\frac{a(\frac{k_n}{n})}{F^{\leftarrow}(1-\frac{k_n}{n})} = o(k_n^{-1/2}), \tag{5.2.10}$$

we also have

$$k_n^{1/2}\left(\hat{\gamma}_{MOM} - \hat{\gamma}_{MLE}\right) \to_p 0$$

where (Dekkers et al., 1989)

$$\hat{\gamma}_{MOM} = M_n^{(1)} + 1 - \frac{1}{2}(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}})^{-1}$$

with $M_n^{(j)} = \frac{1}{k_n}\sum_{i=0}^{k_n-1}(\log X_{n-i,n} - \log X_{n-k_n,n})^j$, $j = 1, 2$. A similar statement also holds for the scale estimator. Condition (5.2.10) is more restrictive than condition (5.2.3); moreover no bias appears. We prove this remark in Section 5.3.

5.3 Proofs

Given (5.2.1) and (5.2.3), from Theorem 2.1 in Drees (1998a) one can find a probability space and define on that space a Brownian motion W and a sequence of stochastic processes Q_n such that: (i) for each n, $Q_n(t) =_d X_{n-[k_n t],n}$, $t \in [0,1]$, (ii) there exist functions $\tilde{a}(k_n/n) \sim a(k_n/n)$ and $\tilde{\Phi}(\frac{k_n}{n}) \sim \Phi(\frac{k_n}{n})$ such that for all $\varepsilon > 0$

$$\begin{aligned}\sup_{t\in[0,1]} t^{\gamma_0+1/2+\varepsilon}\Bigg| &\frac{Q_n(t) - F^{\leftarrow}(1-\frac{k_n}{n})}{\tilde{a}(\frac{k_n}{n})} \\ &- \left(\frac{t^{-\gamma_0}-1}{\gamma_0} - t^{-(\gamma_0+1)}\frac{W(k_n t)}{k_n} + \tilde{\Phi}(\frac{k_n}{n})\Psi(t)\right)\Bigg| \\ = o_p(k_n^{-1/2}) &+ o_p(\tilde{\Phi}(\frac{k_n}{n})), \quad \text{as} \quad n \to \infty.\end{aligned}$$

A similar expansion is also valid for $\gamma_0 \leq -1/2$. Define

$$Y_n(t) = k_n^{1/2}\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} - \frac{t^{-\gamma_0}-1}{\gamma_0}\right) \tag{5.3.1}$$

(read $(t^{-\gamma_0} - 1)/\gamma_0$ as $-\log t$, when $\gamma_0 = 0$). Hence we have the following lemma.

Lemma 5.3.1. *Suppose* (5.2.1) *and that the intermediate sequence k_n satisfies* (5.2.3). *Then, for all $\varepsilon > 0$ as $n \to \infty$,*

$$\begin{aligned} Y_n(t) &= W_n(1) - t^{-(\gamma_0+1)} W_n(t) + k_n^{1/2} \tilde{\Phi}(\frac{k_n}{n}) \Psi(t) \\ &\quad + o_p(1) t^{-(\gamma_0+1/2+\varepsilon)} \end{aligned} \tag{5.3.2}$$

where $W_n(t) = k_n^{-1/2} W(k_n t)$ is a standard Brownian motion and the o_p-term is uniform for $t \in [0,1]$.

From this lemma the following corollary follows easily.

Corollary 5.3.1. *Under the conditions of Lemma 5.3.1, for all $\varepsilon > 0$ as $n \to \infty$,*

$$Y_n(t) = O_p(1)\, t^{-(\gamma_0+1/2+\varepsilon)} \tag{5.3.3}$$

where the O_p-term is uniform for $t \in [0,1]$.

Remark 5.3.1. Lemma 5.3.1 and Corollary 5.3.1 both with $(1 \vee t^{-(\gamma_0+1/2+\varepsilon)})$ instead of only $t^{-(\gamma_0+1/2+\varepsilon)}$ are also valid for $\gamma_0 \le -1/2$.

Given the previous results, to prove Theorem 5.2.1 it is sufficient to consider the m.l. equations with $(X_{n-[k_n t],n} - X_{n-k_n,n})$ replaced by $Q_n(t) - Q_n(1)$, $t \in [0,1]$. It is convenient to reparametrize the equations in terms of $(\gamma, \tilde{\sigma}) = (\gamma, \sigma/\tilde{a}(k_n/n))$. Then we have the equations

$$\begin{cases} \int_0^1 \left\{ \frac{1}{\gamma^2} \log\left(1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right) - \left(\frac{1}{\gamma}+1\right) \frac{\frac{1}{\tilde{\sigma}} \frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}{1+\frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}} \right\} dt = 0 \\ \int_0^1 \left(\frac{1}{\gamma}+1\right) \frac{\frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}{1+\frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}} dt = 1. \end{cases} \tag{5.3.4}$$

Lemma 5.3.2. *Assume conditions* (5.2.1) *and* (5.2.3). *Let $(\gamma, \tilde{\sigma}) = (\gamma_n, \tilde{\sigma}_n)$ be such that*

$$|\gamma/\tilde{\sigma} - \gamma_0| = O_p(k_n^{-1/2}). \tag{5.3.5}$$

Then, if $-1/2 < \gamma_0 < 0$ or $\gamma_0 > 0$

$$P\left(1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} \ge C t^{-\gamma_0}, \quad t \in [\frac{1}{2k_n}, 1]\right) \to 1, \qquad n \to \infty, \tag{5.3.6}$$

for some r.v. $C > 0$, and if $\gamma_0 = 0$,

$$P\left(1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} \ge \frac{1}{2}, \quad t \in [\frac{1}{2k_n}, 1]\right) \to 1, \qquad n \to \infty. \tag{5.3.7}$$

Moreover, if $\gamma_0 = 0$,

$$\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} = O_p(1) \log k_n, \qquad n \to \infty, \tag{5.3.8}$$

where the O_p*-term is uniform for* $t \in [0, 1]$.

Proof. We shall prove the lemma for $F^{\leftarrow}(1 - U_{[k_n t]+1,n})$, where $U_{i,n}$, $1 \le i \le n$, are the i-th uniform order statistics. Then it follows that the statements in the lemma are also valid for $X_{n-[k_n t],n}$, and consequently for $Q_n(t)$.

Note that by Shorack and Wellner (1986 - Chapter 10, Section 3, p. 416, inequality 2) the process

$$\left\{ \frac{n}{k_n t} U_{[k_n t]+1,n} \right\}_{1/(2k_n) \le t \le 1} \tag{5.3.9}$$

is stochastically bounded away from zero and infinity, as $n \to \infty$. Also note that (5.2.1) implies (Drees, 1998a - Lemma 2.1), for some functions $\tilde{a}(s) \sim a(s)$ and $\tilde{\Phi}(s) \sim \Phi(s)$, $s \downarrow 0$, for all $x_0 > 0$ and $\varepsilon > 0$,

$$\lim_{s \downarrow 0} \sup_{0 < x \le x_0} x^{\gamma_0 + \varepsilon} \left| \frac{\frac{F^{\leftarrow}(1-sx) - F^{\leftarrow}(1-s)}{\tilde{a}(s)} - \frac{x^{-\gamma_0} - 1}{\gamma_0}}{\tilde{\Phi}(s)} - \Psi(x) \right| = 0.$$

Combining the two we get, as $n \to \infty$, $k_n \to \infty$, $k_n/n \to 0$,

$$\begin{aligned} \sup_{t \in [1/(2k_n), 1]} t^{\gamma_0 + \varepsilon} \Bigg| & \frac{\frac{F^{\leftarrow}\left(1 - U_{[k_n t]+1,n}\right) - F^{\leftarrow}\left(1 - \frac{k_n}{n}\right)}{\tilde{a}(\frac{k_n}{n})} - \frac{\left(\frac{n}{k_n} U_{[k_n t]+1,n}\right)^{-\gamma_0} - 1}{\gamma_0}}{\tilde{\Phi}(\frac{k_n}{n})} \\ & - \Psi\left(\frac{n}{k_n} U_{[k_n t]+1,n}\right) \Bigg| = o_p(1). \end{aligned} \tag{5.3.10}$$

Then we have, for $-1/2 < \gamma_0 < 0$ or $\gamma_0 > 0$,

$$\begin{aligned} \frac{X_{n-[k_n t],n} - X_{n-k_n,n}}{\tilde{a}(\frac{k_n}{n})} &= \frac{F^{\leftarrow}\left(1 - U_{[k_n t]+1,n}\right) - F^{\leftarrow}\left(1 - U_{k_n+1,n}\right)}{\tilde{a}(\frac{k_n}{n})} \\ = \; & \frac{1}{\gamma_0} \left(\frac{n}{k_n} U_{[k_n t]+1,n}\right)^{-\gamma_0} - \frac{1}{\gamma_0} \left(\frac{n}{k_n} U_{k_n+1,n}\right)^{-\gamma_0} + \tilde{\Phi}(\frac{k_n}{n}) \Psi\left(\frac{n}{k_n} U_{[k_n t]+1,n}\right) \\ & - \tilde{\Phi}(\frac{k_n}{n}) \Psi\left(\frac{n}{k_n} U_{k_n+1,n}\right) + o_p(1 \vee t^{-(\gamma_0 + \varepsilon)}) \tilde{\Phi}(\frac{k_n}{n}). \end{aligned}$$

Hence

$$
\begin{aligned}
1+\frac{\gamma}{\tilde\sigma}&\frac{X_{n-[k_nt],n}-X_{n-k_n,n}}{\tilde a(\frac{k_n}{n})}\\
=&\left\{1-\left(\frac{n}{k_n}U_{k_n+1,n}\right)^{-\gamma_0}\right\}-(\frac{\gamma}{\tilde\sigma}-\gamma_0)\frac{1}{\gamma_0}\left(\frac{n}{k_n}U_{k_n+1,n}\right)^{-\gamma_0}\\
&+\frac{\gamma}{\tilde\sigma}\frac{1}{\gamma_0}\left(\frac{n}{k_n}U_{[k_nt]+1,n}\right)^{-\gamma_0}+\frac{\gamma}{\tilde\sigma}\tilde\Phi(\frac{k_n}{n})\Psi\left(\frac{n}{k_n}U_{[k_nt]+1,n}\right)\\
&-\frac{\gamma}{\tilde\sigma}\tilde\Phi(\frac{k_n}{n})\Psi\left(\frac{n}{k_n}U_{k_n+1,n}\right)+o_p((1\vee t^{-(\gamma_0+\varepsilon)})k_n^{-1/2})\\
=&\ I+II+III+IV+V+VI.
\end{aligned}
$$

By (5.3.9) the product $t^{\gamma_0}III$ is bounded away from zero. The other terms are of lower order:

By the asymptotic normality of intermediate order statistics, part I is $O_p(k_n^{-1/2})$. Hence $t^{\gamma_0}I=o_p(1)$; this is trivial if $\gamma_0>0$ and for $-1/2<\gamma_0<0$ note that $t^{\gamma_0}k_n^{-1/2}\le 2^{-\gamma_0}k_n^{-\gamma_0-1/2}\to 0$, $k_n\to\infty$. By assumption and (5.3.9) part II is $O_p(k_n^{-1/2})$. For IV and V note that

$$t^{\gamma_0}\Psi(t)=o(t^{-1/2}),\quad t\downarrow 0.$$

This combined with (5.2.3) and (5.3.9) gives that IV and V are $o_p(t^{-\gamma_0})$. Part VI is now also obviously $o_p(t^{-\gamma_0})$, provided ε is small enough.

Now consider the case $\gamma_0=0$. Since (5.3.10) is still valid when $\gamma_0=0$, with the obvious changes, we get

$$
\begin{aligned}
&\frac{X_{n-[k_nt],n}-X_{n-k_n,n}}{\tilde a(\frac{k_n}{n})}\\
=&-\log\left(\frac{n}{k_n}U_{[k_nt]+1,n}\right)+\log\left(\frac{n}{k_n}U_{k_n+1,n}\right)+\tilde\Phi(\frac{k_n}{n})\Psi\left(\frac{n}{k_n}U_{[k_nt]+1,n}\right)\\
&-\tilde\Phi(\frac{k_n}{n})\Psi\left(\frac{n}{k_n}U_{k_n+1,n}\right)+o_p(t^{-\varepsilon})\tilde\Phi(\frac{k_n}{n})
\end{aligned}
\tag{5.3.11}
$$

hence,

$$
\begin{aligned}
1+\frac{\gamma}{\tilde\sigma}&\frac{X_{n-[k_nt],n}-X_{n-k_n,n}}{\tilde a(\frac{k_n}{n})}\\
=&\ 1-\frac{\gamma}{\tilde\sigma}\log t-\frac{\gamma}{\tilde\sigma}\log\left(\frac{n}{k_n}U_{[k_nt]+1,n}\right)+\frac{\gamma}{\tilde\sigma}\log\left(\frac{n}{k_n}U_{k_n+1,n}\right)\\
&+\frac{\gamma}{\tilde\sigma}\tilde\Phi(\frac{k_n}{n})\Psi\left(\frac{n}{k_n}U_{[k_nt]+1,n}\right)-\frac{\gamma}{\tilde\sigma}\tilde\Phi(\frac{k_n}{n})\Psi\left(\frac{n}{k_n}U_{k_n+1,n}\right)+o_p(\frac{\gamma}{\tilde\sigma}t^{-\varepsilon}k_n^{-1/2}).
\end{aligned}
$$

Hence by assumptions (5.2.3) and (5.3.5), (5.3.9) and since $t\ge 1/(2k_n)$, all the terms but the 1 in the last equality are negligible.

Finally, to verify (5.3.8) just note that (5.3.11) when $t = 1/(2k_n)$ is $O_p(\log k_n)$, provided $0 < \varepsilon < 1/2$. Hence given the monotonicity of $X_{n-[k_n\cdot],n} - X_{n-k_n,n}$ the result follows. □

Proposition 5.3.1. *Assume conditions* (5.2.1) *and* (5.2.3). *Any solution* $(\gamma, \tilde{\sigma})$ *of* (5.3.4) *for which* (5.3.5) *holds and for which* $\log\tilde{\sigma}$ *is bounded satisfies, as* $n \to \infty$,

$$\begin{aligned} &k_n^{1/2}(\gamma - \gamma_0) - \tfrac{(\gamma_0+1)^2}{\gamma_0}\int_0^1 \left(t^{\gamma_0} - (2\gamma_0+1)t^{2\gamma_0}\right) Y_n(t)dt = o_p(1) \\ &k_n^{1/2}(\tilde{\sigma} - 1) - \tfrac{\gamma_0+1}{\gamma_0}\int_0^1 ((\gamma_0+1)(2\gamma_0+1)t^{2\gamma_0} - t^{\gamma_0})Y_n(t)dt = o_p(1), \end{aligned} \tag{5.3.12}$$

where for $\gamma_0 = 0$ *these equations should be interpreted as its limit for* $\gamma_0 \to 0$, *i.e. for* $\gamma_0 = 0$,

$$\begin{aligned} &k_n^{1/2}\gamma + \int_0^1 (2 + \log t)Y_n(t)dt = o_p(1) \\ &k_n^{1/2}(\tilde{\sigma} - 1) - \int_0^1 (3 + \log t)Y_n(t)dt = o_p(1). \end{aligned} \tag{5.3.13}$$

Remark 5.3.2. For $\gamma_0 \neq 0$ the condition on $\log\tilde{\sigma}$ is not needed.

Proof. We consider the cases $\gamma_0 > 0$, $-1/2 < \gamma_0 < 0$ and $\gamma_0 = 0$ separately.

Case $\gamma_0 > 0$. In this case system (5.3.4) can be simplified to

$$\begin{cases} \int_0^1 \log\left(1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right) dt = \gamma \\ \int_0^1 \frac{1}{1+\frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}} dt = \frac{1}{\gamma+1}. \end{cases} \tag{5.3.14}$$

Next we will find expansions for the left-hand side of both equations.

Rewrite the first one as

$$\int_0^{(2k_n)^{-1}} \log\left(1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right) dt + \int_{(2k_n)^{-1}}^1 \log t^{-\gamma_0} dt$$

$$+ \int_{(2k_n)^{-1}}^1 \log\left\{t^{\gamma_0}\left(1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)\right\} dt$$

$$= I_1 + \gamma_0(1 - \frac{1}{2k_n}) + I_2.$$

First we prove that I_1 is negligible. Since $t \mapsto Q_n(t)$ is constant when $t \in [0, (2k_n)^{-1}]$, from Lemma 5.3.2 we have

$$1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} = 1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(\frac{1}{2k_n}) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} \geq (2k_n)^{\gamma_0}C, \tag{5.3.15}$$

for all $t \in [0, (2k_n)^{-1}]$. On the other hand from (5.3.1), (5.3.3) and (5.3.5),

$$1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(\frac{1}{2k_n}) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} = k_n^{\gamma_0+\varepsilon} O_p(1).$$

Hence it follows that $I_1 = o_p(k_n^{-1/2})$.

Next we turn to the main term I_2. We will use the inequality $0 \le x - \log(1+x) \le x^2/(2(1 \wedge (1+x))$, valid for all $x > -1$, with

$$x = t^{\gamma_0} \left(1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} \right) - 1.$$

Then, from Lemma 5.3.2 it follows that $0 < 1/(1 \wedge (1+x)) \le 1 \vee 1/C < \infty$ with probability tending to one. Moreover note that relation (5.3.3) implies

$$\int_0^{(2k_n)^{-1}} t^{\gamma_0} Y_n(t) dt = O\left(\int_0^{(2k_n)^{-1}} t^{-1/2-\varepsilon} dt \right) = O_p((2k_n)^{-1/2+\varepsilon}) = o_p(1),$$

for $\varepsilon \in (0, 1/2)$. Hence from (5.3.1), as $n \to \infty$,

$$\begin{aligned}
I_2 &= \int_{(2k_n)^{-1}}^1 \left((\frac{\gamma}{\tilde{\sigma}} - \gamma_0) \frac{1 - t^{\gamma_0}}{\gamma_0} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} t^{\gamma_0} Y_n(t) \right) dt \\
&\quad + O_p\left(\int_{(2k_n)^{-1}}^1 \left((\frac{\gamma}{\tilde{\sigma}} - \gamma_0) \frac{1 - t^{\gamma_0}}{\gamma_0} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} t^{\gamma_0} Y_n(t) \right)^2 dt \right) \\
&= \left\{ (\frac{\gamma}{\tilde{\sigma}} - \gamma_0) \frac{1}{(\gamma_0 + 1)} + O_p(k_n^{-1/2} (2k_n)^{-1}) \right\} \\
&\quad + \left\{ \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} \int_0^1 t^{\gamma_0} Y_n(t) dt + o_p(k_n^{-1/2}) \right\} \\
&\quad + O_p(k_n^{-1} + k_n^{-1} (2k_n)^{2\varepsilon} + k_n^{-1} (2k_n)^{-1/2+\varepsilon}) \\
&= (\frac{\gamma}{\tilde{\sigma}} - \gamma_0) \frac{1}{(\gamma_0 + 1)} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} \int_0^1 t^{\gamma_0} Y_n(t) dt + o_p(k_n^{-1/2}),
\end{aligned}$$

where for the last equality we took $\varepsilon < 1/4$. Hence we proved

$$\begin{aligned}
&\int_0^1 \log \left(1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} \right) dt \\
&= \gamma_0 + (\frac{\gamma}{\tilde{\sigma}} - \gamma_0) \frac{1}{(\gamma_0 + 1)} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} \int_0^1 t^{\gamma_0} Y_n(t) dt + o_p(k_n^{-1/2}).
\end{aligned}$$

This means that the first equation of (5.3.14) implies

$$\gamma = \gamma_0 + (\frac{\gamma}{\tilde{\sigma}} - \gamma_0) \frac{1}{(\gamma_0 + 1)} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} \int_0^1 t^{\gamma_0} Y_n(t) dt + o_p(k_n^{-1/2}).$$

Now we deal with the left-hand side of the second equation in (5.3.14). Using the equality

$$\frac{1}{1+x} = 1 - x + \frac{x^2}{1+x}$$

valid for $x > -1$, with

$$x = t^{\gamma_0}\left(1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right) - 1,$$

we get, uniformly for $1/(2k_n) \le t \le 1$,

$$\begin{aligned}\frac{1}{1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}} &= t^{\gamma_0}\Big\{1 - \Big(\frac{\gamma}{\tilde{\sigma}} - \gamma_0\Big)\frac{1-t^{\gamma_0}}{\gamma_0} - t^{\gamma_0}\frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}Y_n(t) \\ &\quad + \frac{\Big(\big(\frac{\gamma}{\tilde{\sigma}} - \gamma_0\big)\frac{1-t^{\gamma_0}}{\gamma_0} + t^{\gamma_0}\frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}Y_n(t)\Big)^2}{1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}\Big\}.\end{aligned}$$

Hence the left-hand side of the second equation in (5.3.14) equals

$$\begin{aligned}&\int_0^{(2k_n)^{-1}} \frac{1}{1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}dt \\ &+ \int_{(2k_n)^{-1}}^1 \Big\{t^{\gamma_0} - (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{t^{\gamma_0} - t^{2\gamma_0}}{\gamma_0} - \frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}t^{2\gamma_0}Y_n(t)\Big\}dt \qquad (5.3.16) \\ &+ \int_{(2k_n)^{-1}}^1 \frac{\Big((\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1-t^{\gamma_0}}{\gamma_0} + \frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}t^{\gamma_0}Y_n(t)\Big)^2}{1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}dt.\end{aligned}$$

From (5.3.15) it follows easily that the first integral is $o_p(k_n^{-1/2})$. For the second integral, direct calculations and (5.3.3) yield

$$\begin{aligned}&\frac{1}{\gamma_0 + 1} - (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{(\gamma_0+1)(2\gamma_0+1)} - \frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}\int_0^1 t^{2\gamma_0}Y_n(t)dt \\ &+ O_p((2k_n)^{-\gamma_0-1} + k_n^{-1/2}(2k_n)^{-\gamma_0-1} + k_n^{-1/2}(2k_n)^{-\gamma_0-1/2+\varepsilon})\end{aligned}$$

and, provided $\varepsilon < 1/2$ the O_p-term is $o_p(k_n^{-1/2})$. For the last integral of (5.3.16), by Lemma 5.3.2 it is bounded by

$$\begin{aligned}&O_p\left(\int_{(2k_n)^{-1}}^1 t^{\gamma_0}\left((\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1-t^{\gamma_0}}{\gamma_0} + \frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}t^{\gamma_0}Y_n(t)\right)^2 dt\right) \\ &= O_p(k_n^{-1} + k_n^{-1}(1 + (2k_n)^{-\gamma_0+2\varepsilon}) + k_n^{-1}(2k_n)^{-\gamma_0-1/2+\varepsilon}) = o_p(k_n^{-1/2}),\end{aligned}$$

if $\varepsilon < 1/4 + \gamma_0/2$. Therefore we proved

$$\int_0^1 \frac{1}{1+\frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}dt$$
$$= \frac{1}{\gamma_0+1} - (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{(\gamma_0+1)(2\gamma_0+1)} - \frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}\int_0^1 t^{2\gamma_0}Y_n(t)dt + o_p(k_n^{-1/2}).$$

Hence, under the given conditions system (5.3.14) implies

$$\begin{cases} \gamma_0 + (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{\gamma_0+1} + \frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}\int_0^1 t^{\gamma_0}Y_n(t)dt + o_p(k_n^{-1/2}) = \gamma \\ \frac{1}{\gamma_0+1} - (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{(\gamma_0+1)(2\gamma_0+1)} - \frac{\gamma}{\tilde{\sigma}}k_n^{-1/2}\int_0^1 t^{2\gamma_0}Y_n(t)dt + o_p(k_n^{-1/2}) = \frac{1}{\gamma+1}. \end{cases} \tag{5.3.17}$$

Next we prove that (5.3.17) implies (5.3.12). First note that by (5.3.3) and (5.3.5) we have that (5.3.17) implies

$$\begin{cases} \gamma_0 + (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{\gamma_0+1} + \gamma_0 k_n^{-1/2}\int_0^1 t^{\gamma_0}Y_n(t)dt + o_p(k_n^{-1/2}) = \gamma \\ \frac{1}{\gamma_0+1} - (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{(\gamma_0+1)(2\gamma_0+1)} - \gamma_0 k_n^{-1/2}\int_0^1 t^{2\gamma_0}Y_n(t)dt + o_p(k_n^{-1/2}) = \frac{1}{\gamma+1}. \end{cases} \tag{5.3.18}$$

Then note that from the first equation and (5.3.5) we have $|\gamma - \gamma_0| = O_p(k_n^{-1/2})$, hence $|\gamma - \gamma_0|^2 = o_p(k_n^{-1/2})$. Therefore $1/(\gamma_0+1) - 1/(\gamma+1) = (\gamma-\gamma_0)/(\gamma_0+1)^2 + o(k_n^{-1/2})$ and so (5.3.18) implies

$$\begin{cases} \gamma - \gamma_0 - (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{\gamma_0+1} - k_n^{-1/2}\gamma_0\int_0^1 t^{\gamma_0}Y_n(t)dt + o_p(k_n^{-1/2}) = 0 \\ \frac{\gamma-\gamma_0}{(\gamma_0+1)^2} - (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{(\gamma_0+1)(2\gamma_0+1)} - k_n^{-1/2}\gamma_0\int_0^1 t^{2\gamma_0}Y_n(t)dt + o_p(k_n^{-1/2}) = 0. \end{cases}$$

Then, solving this system in $\gamma - \gamma_0$ and $(\frac{\gamma}{\tilde{\sigma}} - \gamma_0)$ one easily gets (5.3.12).

Case $-1/2 < \gamma_0 < 0$. Again, in this case system (5.3.4) simplifies to (5.3.14). Rewrite the left-hand side of the first equation as

$$\int_0^{s_n} \log\left(1+\frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)dt + \int_{s_n}^1 \log t^{-\gamma_0}\,dt$$
$$+\int_{s_n}^1 \log\left\{t^{\gamma_0}\left(1+\frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)\right\}dt$$
$$= J_1 + \{\gamma_0 + O_p(s_n|\log s_n|)\} + J_2$$

and choose $s_n = k_n^{-\delta}$, with $\delta \in (1/2, (4\varepsilon)^{-1})$ for some $\varepsilon \in (0, 1/2)$.

Now we prove that J_1 is negligible. Note that since $t \mapsto Q_n(t)$ is constant when $t \in [0, (2k_n)^{-1}]$, (5.3.6) is trivially extended to $t \in [0,1]$ when $\gamma_0 < 0$. By definition $Q_n(t) - Q_n(1) \geq 0$, for all $t \in [0,1]$ and $\tilde{a}(k_n/n) > 0$. Hence from Lemma 5.3.2 with $\gamma < 0$,

$$P\left(\left|\log\left(1+\frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)\right| \leq |\log(Ct^{-\gamma_0})|, \quad t \in [0,1]\right) \to 1,$$

as $n \to \infty$, and so $\int_0^{s_n} |\log(Ct^{-\gamma_0})|\, dt = o_p(k_n^{-1/2})$, which gives $J_1 = o_p(k_n^{-1/2})$.

Next we see that the main contribution comes from J_2. From (5.3.1), Lemma 5.3.2, and since $0 \le x - \log(1+x) \le x^2/[2(1 \wedge (1+x))]$, valid for all $x > -1$, and taking $x = t^{\gamma_0}[1 + (\gamma/\tilde{\sigma})(Q_n(t) - Q_n(1))/a(k_n/n)]$, we have

$$\begin{aligned}
J_2 &= \int_{s_n}^1 \left((\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1-t^{\gamma_0}}{\gamma_0} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} t^{\gamma_0} Y_n(t) \right) dt \\
&\quad + O_p\left(\int_{s_n}^1 \left[(\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1-t^{\gamma_0}}{\gamma_0} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} t^{\gamma_0} Y_n(t) \right]^2 dt \right) \\
&= \left\{ (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{\gamma_0+1} + O_p(k_n^{-1/2} s_n^{\gamma_0+1}) \right\} + \left\{ \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} \int_0^1 t^{\gamma_0} Y_n(t) dt \right. \\
&\quad \left. + O_p(k_n^{-1/2} \int_0^{s_n} t^{-1/2-\varepsilon} dt) \right\} \\
&\quad + O_p\left(\int_{s_n}^1 \left[(\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1-t^{\gamma_0}}{\gamma_0} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} t^{\gamma_0} Y_n(t) \right]^2 dt \right)
\end{aligned}$$

and from the choice of s_n it follows that the O_p-terms are $o_p(k_n^{-1/2})$. Hence we proved that

$$\begin{aligned}
&\int_0^1 \log\left(1 + \frac{\gamma}{\tilde{\sigma}} \frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})} \right) dt \\
&= \gamma_0 + (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{(\gamma_0+1)} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} \int_0^1 t^{\gamma_0} Y_n(t) dt + o_p(k_n^{-1/2}).
\end{aligned}$$

Now we turn to the second equation in (5.3.14). Using the same decomposition as in the $\gamma > 0$ case, that is (5.3.16), we get say, $K_1 + K_2 + K_3$, and take this time $s_n = k_n^{-\delta}$ for some $\delta \in ((2\gamma_0+2)^{-1}, (-6\gamma_0-2)^{-1} \wedge (4\varepsilon - 2\gamma_0)^{-1} \wedge (-4\gamma_0 - 1 + 2\varepsilon)^{-1})$ and $\varepsilon \in (0, \gamma_0 + 1/2)$. From Lemma 5.3.2, $K_1 = O_p(s_n^{\gamma_0+1}) = o_p(k_n^{-1/2})$. Moreover

$$\begin{aligned}
K_2 &= \frac{1}{\gamma_0+1} - (\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1}{(\gamma_0+1)(2\gamma_0+1)} - \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} \int_0^1 t^{2\gamma_0} Y_n(t) dt \\
&\quad + O_p(s_n^{\gamma_0+1} + k_n^{-1/2} s_n^{2\gamma_0+1} + k_n^{-1/2} s_n^{\gamma_0+1/2-\varepsilon})
\end{aligned}$$

and given s_n we have that the O_p-term is $o_p(k_n^{-1/2})$. Finally from Lemma 5.3.2 and the definition of s_n,

$$\begin{aligned}
K_3 &= O_p\left(\int_{s_n}^1 t^{\gamma_0} \left[(\frac{\gamma}{\tilde{\sigma}} - \gamma_0)\frac{1-t^{\gamma_0}}{\gamma_0} + \frac{\gamma}{\tilde{\sigma}} k_n^{-1/2} t^{\gamma_0} Y_n(t) \right]^2 dt \right) \\
&= O_p\left(k_n^{-1}(s_n^{3\gamma_0+1} \vee (-\log s_n)) + k_n^{-1}(s_n^{\gamma_0-2\varepsilon} + 1) + k_n^{-1}(s_n^{2\gamma_0+1/2-\varepsilon} + 1) \right) \\
&= o_p(k_n^{-1/2}).
\end{aligned}$$

Hence the conclusion is the same as in the $\gamma_0 > 0$ case.

Case $\gamma_0 = 0$. In this case we use equations (5.3.4). Using (twice) the equality $1/(1+x) = 1 - x + x^2/(1+x)$, the inequality $|x - \log(1+x) - x^2/2 + x^3/3| \leq x^4/[4(1 \wedge (1+x))]$, valid for all $x > -1$, both with $x = (\gamma/\tilde{\sigma})(Q_n(t) - Q_n(1))/\tilde{a}(k_n/n)$, and (5.3.7) in Lemma 5.3.2, we have for the left-hand side of the first equation

$$\frac{1}{\gamma^2}\int_0^1 \log\left(1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right) - (1+\gamma)\frac{\frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}{1 + \frac{\gamma}{\tilde{\sigma}}\frac{Q_n(t)-Q_n(1)}{\tilde{a}(\frac{k_n}{n})}}dt$$

$$= \int_0^1 -\frac{1}{\tilde{\sigma}}\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}dt + \int_0^1 \left(\frac{1}{2} + \gamma\right)\frac{1}{\tilde{\sigma}^2}\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)^2 dt$$

$$- \int_0^1 \left(\frac{2}{3} + \gamma\right)\frac{\gamma}{\tilde{\sigma}^3}\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)^3 dt$$

$$+ O_p\left(\int_0^1 \frac{\gamma^2}{\tilde{\sigma}^4}\left[\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right]^4 + \frac{\gamma^3}{\tilde{\sigma}^5}\left[\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right]^5 dt\right). \quad (5.3.19)$$

From (5.3.1) the first integral in the right-hand side of the last equation equals $-\tilde{\sigma}^{-1} - \tilde{\sigma}^{-1}k_n^{-1/2}\int_0^1 Y_n(t)dt + o_p(k_n^{-1/2})$. For the second integral in the right-hand side of (5.3.19) consider,

$$\left(\frac{1}{2} + \gamma\right)\frac{1}{\tilde{\sigma}^2}\left(\int_0^{s_n}\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)^2 dt + \int_{s_n}^1\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(\frac{k_n}{n})}\right)^2 dt\right),$$

with $s_n = k_n^{-\delta}$, $\delta \in (1/2, (4\varepsilon)^{-1})$, $\varepsilon \in (0, 1/2)$. Then the first of these last two integrals is $o_p(k_n^{-1/2})$ from (5.3.8). Using (5.3.1) and (5.3.3), the second integral equals

$$\frac{1 + 2\gamma}{\tilde{\sigma}^2} - \frac{1}{\tilde{\sigma}^2}k_n^{-1/2}\int_0^1 (\log t)Y_n(t)dt + o_p(k_n^{-1/2}).$$

Recall that $|\gamma/\tilde{\sigma}| = O_p(k_n^{-1/2})$.

Using a similar reasoning, but with $\delta \in (1/2, 3(4\varepsilon)^{-1} \wedge 4(1+6\varepsilon)^{-1})$, $\varepsilon \in (0, 1/2)$, the third integral of (5.3.19) equals $-4\gamma\tilde{\sigma}^{-3} + o_p(k_n^{-1/2})$. Finally the O_p-term of (5.3.19) is clearly $o_p(k_n^{-1/2})$ from (5.3.8).

Hence we have that (5.3.19) equals

$$-\frac{1}{\tilde{\sigma}} + \frac{1 + 2\gamma}{\tilde{\sigma}^2} - \frac{4\gamma}{\tilde{\sigma}^3} - \frac{1}{\tilde{\sigma}}k_n^{-1/2}\int_0^1 Y_n(t)dt - \frac{1}{\tilde{\sigma}^2}k_n^{-1/2}\int_0^1 (\log t)Y_n(t)dt + o_p(k_n^{-1/2}).$$

To deal with the left-hand side of the second equation, use again the aforementioned equality for $1/(1+x)$ and (5.3.7) in Lemma 5.3.2 to get

$$\begin{aligned}
&\frac{1+\gamma}{\gamma}\left\{\int_0^1 \frac{\gamma}{\tilde\sigma}\frac{Q_n(t)-Q_n(1)}{\tilde a(\frac{k_n}{n})}dt - \int_0^1\left(\frac{\gamma}{\tilde\sigma}\frac{Q_n(t)-Q_n(1)}{\tilde a(\frac{k_n}{n})}\right)^2 dt\right.\\
&\qquad\left.+O_p\left(\int_0^1\left[\frac{\gamma}{\tilde\sigma}\frac{Q_n(t)-Q_n(1)}{\tilde a(\frac{k_n}{n})}\right]^3 dt\right)\right\}\\
&= (1+\gamma)\left[\frac{1}{\tilde\sigma}+\frac{1}{\tilde\sigma}k_n^{-1/2}\int_0^1 Y_n(t)dt - \int_0^{s_n}\gamma\left(\frac{1}{\tilde\sigma}\frac{Q_n(t)-Q_n(1)}{\tilde a(\frac{k_n}{n})}\right)^2 dt\right.\\
&\qquad\left.-\int_{s_n}^1\gamma\left(\frac{1}{\tilde\sigma}\frac{Q_n(t)-Q_n(1)}{\tilde a(\frac{k_n}{n})}\right)^2 dt + O_p\left(k_n^{-1}(\log k_n)^3\right)\right],
\end{aligned}$$

where for the O_p-term we used (5.3.8), and then it follows that this O_p-term is $(\gamma/\tilde\sigma)o_p(k_n^{-1/2})$. Next we consider the second and third integrals in the last equality, L_1 and L_2 say. As for L_1 it follows from (5.3.8) that

$$L_1 = O_p\left(s_n\gamma(\log k_n)^2\right) = o_p(k_n^{-1/2})$$

if $s_n = k_n^{-\delta}$, $\delta\in(0,1)$. As for L_2, from (5.3.3) with $\varepsilon\in(0,1/2)$, we get

$$\begin{aligned}
L_2 &= -2\frac{\gamma}{\tilde\sigma^2}+O_p\left(k_n^{-1/2}s_n(\log s_n)^2 + k_n^{-3/2}s_n^{-2\varepsilon}+k_n^{-1}\right)\\
&= -2\frac{\gamma}{\tilde\sigma^2}+o_p(k_n^{-1/2}).
\end{aligned}$$

Hence we proved that

$$\begin{aligned}
&\int_0^1\left(\frac{1}{\gamma}+1\right)\frac{\frac{\gamma}{\tilde\sigma}\frac{Q_n(t)-Q_n(1)}{\tilde a(\frac{k_n}{n})}}{1+\frac{\gamma}{\tilde\sigma}\frac{Q_n(t)-Q_n(1)}{\tilde a(\frac{k_n}{n})}}dt\\
&= \frac{1+\gamma}{\tilde\sigma}-\frac{2\gamma}{\tilde\sigma^2}+\frac{1}{\tilde\sigma}k_n^{-1/2}\int_0^1 Y_n(t)dt + o_p(k_n^{-1/2}).
\end{aligned}$$

Therefore under the given conditions, a solution of the m.l. equations must satisfy

$$\begin{cases}
(1-\tilde\sigma)+2\gamma-\frac{4\gamma}{\tilde\sigma}-\tilde\sigma k_n^{-1/2}\int_0^1 Y_n(t)dt - k_n^{-1/2}\int_0^1\log t Y_n(t)dt + o_p(k_n^{-1/2}) = 0\\
(1-\tilde\sigma)+\gamma-\frac{2\gamma}{\tilde\sigma}+k_n^{-1/2}\int_0^1 Y_n(t)dt + o_p(k_n^{-1/2}) = 0.
\end{cases}$$

Next note that the first equation implies $\tilde\sigma = 1+O_p(k_n^{-1/2})$, and so $\gamma/\tilde\sigma = \gamma + o_p(k_n^{-1/2})$. Simplifying the above equations, the result follows.

□

Proof of Theorem 5.2.1. From Proposition 5.3.1, under the conditions (5.2.1), (5.2.3), (5.3.5) and $\log\tilde\sigma$ bounded, one easily obtains the solutions (5.2.4)-(5.2.7). That is, starting from (5.3.12)-(5.3.13), use the expansion given in Lemma 5.3.1 for $Y_n(t)$. Note that the integral involving the $o_p(1)$-term in (5.3.2) vanishes, in probability, provided ε is taken small enough.

On the other hand one can easily check that given the solutions (5.3.12)-(5.3.13), one can go back through all the proof of Proposition 5.3.1, now without assuming condition (5.3.5) and that $\log\tilde\sigma$ is bounded. This shows that (5.2.4)-(5.2.7) are the only local possible solutions of the m.l. equations verifying $|\hat\gamma_n-\gamma_0| = O_p(k_n^{-1/2})$ and $|\hat\sigma_n/a(k_n/n)-1| = O_p(k_n^{-1/2})$, $n\to\infty$. □

Proof of Corollary 5.2.1. Since $k_n^{1/2}\Phi(k_n/n)\to\lambda$, the bias of $k_n^{1/2}(\hat\gamma_n-\gamma_0)$ equals $\lambda((\gamma_0+1)^2/\gamma_0)\int_0^1\left(t^{\gamma_0}-(2\gamma_0+1)t^{2\gamma_0}\right)\Psi(t)dt$. Using (5.2.2) and by simple calculations the result follows. Similarly for the other entries of μ.

To obtain the variance of $k_n^{1/2}(\hat\gamma_n-\gamma_0)$, let $X(t) = (\gamma_0+1)^2/\gamma_0 \left(t^{\gamma_0}-(2\gamma_0+1)t^{2\gamma_0}\right)\left(W(1)-t^{-(\gamma_0+1)}W(t)\right)$ and $R_X(s,t) = E[X(s)X(t)]$. Then, $\mathrm{var}(k_n^{1/2}(\hat\gamma_n-\gamma_0)) = \mathrm{var}\left(\int_0^1 X(t)dt\right) = \int_0^1\int_0^1 R_X(s,t)ds\,dt$. By simple calculations the result follows. To obtain the covariance of $k_n^{1/2}(\hat\gamma_n-\gamma_0)$ with $k_n^{1/2}(\hat\sigma_n/a(k_n/n)-1)$, let $Y(t) = (\gamma_0+1)\gamma_0\left((\gamma_0+1)(2\gamma_0+1)t^{2\gamma_0}-t^{\gamma_0}\right)\left(W(1)-t^{-(\gamma_0+1)}W(t)\right)$ and $R_{X,Y}(s,t) = E[X(s)Y(t)]$. Then, $\mathrm{cov}(k_n^{1/2}(\hat\gamma_n-\gamma_0), k_n^{1/2}(\hat\sigma_n/a(k_n/n)-1)) = \int_0^1\int_0^1 R_{X,Y}(s,t)ds\,dt$. The other entries of Σ follow similarly. □

Proof of Theorem 5.2.2. Integration of the various terms of (5.3.2) yields for $\gamma=0$

$$\begin{aligned} &k_n^{1/2}\left(\int_0^1 \frac{Q_n(t)-Q_n(1)}{\tilde a(\frac{k_n}{n})}dt + \int_0^1 \log t\,dt\right)\\ &= \int_0^1 \left(W_n(1)-t^{-1}W_n(t)\right)dt + k_n^{1/2}\tilde\Phi(\frac{k_n}{n})\int_0^1\Psi(t)dt + o_p(1), \end{aligned}$$

hence

$$k_n^{1/2}\left(\frac{m_n^{(1)}}{\tilde a(\frac{k_n}{n})}-1\right) - k_n^{1/2}\Phi(\frac{k_n}{n})\int_0^1\Psi(t)dt \to_d \int_0^1\left(W(1)-t^{-1}W(t)\right)dt$$

with W a Brownian motion. Similarly we obtain

$$\begin{aligned} &k_n^{1/2}\left(\frac{m_n^{(2)}}{(\tilde a(\frac{k_n}{n}))^2}-2\right) + k_n^{1/2}\Phi(\frac{k_n}{n})\int_0^1 2(\log t)\Psi(t)dt\\ &\to_d -\int_0^1 2(\log t)\left(W(1)-t^{-1}W(t)\right)dt. \end{aligned}$$

Application of Cramér's delta method then gives

$$k_n^{1/2}\hat\gamma_* + k_n^{1/2}\Phi(\frac{k_n}{n})\int_0^1(2+\log t)\Psi(t)dt \to_d -\int_0^1(2+\log t)\left(W(1)-t^{-1}W(t)\right)dt.$$

Hence by (5.2.6)

$$k_n^{1/2}\left(\hat{\gamma}_* - \hat{\gamma}_{MLE}\right) \to_p 0.$$

The proof of the second statement is similar. □

Proof of Remark 5.2.4. Under the stated conditions the following analogue of (5.3.2) holds:

$$\begin{aligned} & k_n^{1/2}\left(\frac{\log Q_n(t) - \log Q_n(1)}{\tilde{a}(\frac{k_n}{n})/F^{\leftarrow}(1-\frac{k_n}{n})} + \log t\right) \\ = \quad & W_n(1) - t^{-1}W_n(t) - k_n^{1/2}\Phi^*(\frac{k_n}{n})\frac{\log^2 t}{2} + o_p(1)t^{-1/2-\varepsilon}, \end{aligned}$$

provided

$$\Phi^*(\frac{k_n}{n}) \sim \frac{\tilde{a}(\frac{k_n}{n})}{F^{\leftarrow}(1-\frac{k_n}{n})} = O(k_n^{-1/2}), \tag{5.3.20}$$

see Draisma et al. (1999), Appendix. Following the same reasoning as in the proof of Theorem 5.2.2, we then get the result. □

Chapter 6

Tail dependence in independence

Co-authors: Gerrit Draisma, Holger Drees and Laurens de Haan

Submitted

Abstract. We propose a new estimator of the parameter η, introduced by Ledford and Tawn (1996), governing dependence in bivariate distributions with asymptotically independent componentwise maxima. We prove asymptotic normality of this estimator and two other estimators proposed in the quoted paper. For the latter we develop a weighted approximation result for a two-dimensional rank-process. We compare the estimators and a related test for asymptotic independence in a simulation study. Also we show consistency of the resulting estimator for failure probabilities in this set-up. Our estimator for η is inspired by the work of Peng (1999). Our less strict second order conditions are satisfied by the normal distribution.

6.1 Introduction

Suppose a region is protected by a river dam against flooding. The water level is regularly observed at two stations, yielding a sample $(X_i, Y_i), 1 \le i \le n$. If there is no other protection within the region, the whole area will be flooded if the water level exceeds the height of the dam at one of both points. Hence the probability of a flooding at a particular date is of the form

$$P\{X_i > u \text{ or } Y_i > v\}. \tag{6.1.1}$$

We assume that (if necessary, after a suitable declustering) the vectors (X_i, Y_i) are independent and identically distributed with distribution function F, say. If the heights u and v of the dam are large, then multivariate extreme value theory provides a framework which allows a systematic estimation of the probability (6.1.1). For this, assume that there exist normalising constants $a_n, c_n > 0$ and $b_n, d_n \in \mathbb{R}$ such that

$$\begin{aligned}\lim_{n\to\infty} F^n(a_n x + b_n, c_n y + d_n) &= \lim_{n\to\infty} P\Big\{\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \le x, \frac{\bigvee_{i=1}^n Y_i - d_n}{c_n} \le y\Big\} \\ &= G(x,y) \qquad (6.1.2)\end{aligned}$$

for all but denumerable many vectors (x, y). Here $\bigvee_{i=1}^n X_i$ denotes the maximum of n consecutive water levels at the first station and G is a distribution function with non-degenerate marginals (cf. Resnick, 1987; Chapter 5). Taking logarithms, one concludes from (6.1.2) that

$$\lim_{n\to\infty} nP\Big\{\frac{X - b_n}{a_n} > x \text{ or } \frac{Y - d_n}{c_n} > y\Big\} = -\log G(x,y) \qquad (6.1.3)$$

for a random vector (X, Y) with distribution function F.

For the sake of simplicity, in this introduction we concentrate on the case when both marginals are uniformly distributed; this can be achieved by transforming the random variables X and Y with their pertaining marginal distribution functions F_i (cf. (6.3.1)). Then (6.1.3) simplifies to

$$\lim_{n\to\infty} nP\{1 - X < x/n \text{ or } 1 - Y < y/n\} = -\log G(-x, -y). \qquad (6.1.4)$$

and, in fact, even

$$\lim_{s\to 0} s^{-1}P\{1 - X < sx \text{ or } 1 - Y < sy\} = -\log G(-x, -y) \qquad (6.1.5)$$

with s running through $\mathbb{R}$. Dividing the analogous equation where s is replaced with st by (6.1.5), one sees that

$$P\{1 - X < tx \text{ or } 1 - Y < ty\} \approx tP\{1 - X < x \text{ or } 1 - Y < y\} \qquad (6.1.6)$$

for small x and y, i.e., the function $t \mapsto P\{1 - X < tx \text{ or } 1 - Y < ty\}$ is regularly varying at 0 with index 1.

Recall that we want to estimate the probability (6.1.1) with u and v so close to 1 that no or only very few observations lie in the failure region $\{(r,s) \in [0,1]^2 \mid 1-r < 1-u \text{ or } 1-s < 1-v\}$. Now choose a sufficiently small t such that the set

$$\{(r,s) \in [0,1]^2 \mid 1 - r < (1-u)/t \text{ or } 1 - s < (1-v)/t\} \qquad (6.1.7)$$

does contain a considerable number of observations and hence the probability that (X, Y) lies in (6.1.7) can be estimated using the empirical distribution. Then we can use (6.1.6) with $x = (1-u)/t$ and $y = (1-v)/t$ to estimate the probability (6.1.1) we are actually interested in.

However, in many situations one may also be interested in the probability that both thresholds are exceeded, i.e., $P\{X > u \text{ and } Y > v\}$. This probability is of interest, e.g., if the levels of two different air pollutants, the losses suffered in two different investments or different variables relevant for the probability of a flooding (cf. Section 6.5) are observed. Convergence (6.1.3) implies

$$\lim_{n\to\infty} nP\Big\{\frac{X-b_n}{a_n} > x \text{ and } \frac{Y-d_n}{c_n} > y\Big\} = -\log G(x,y) + \log G_1(x) + \log G_2(y), \tag{6.1.8}$$

since the marginal distributions converge to the marginals G_1 and G_2 of the limit distribution. Note that if the marginals of the limit distributions are independent, that is, $G(x,y) = G_1(x)G_2(y)$, the limit in (6.1.8) is identically zero. In that case we say that the maxima of the X_i and those of the Y_i are asymptotically independent. This is a rather common situation; for instance, it holds for nondegenerate bivariate normal distributions.

Unfortunately, in this case the reasoning used above to derive estimators for the probability (6.1.1) does not lead to anything one can employ for the estimation of the probability of a joint exceedance, since the analog to (6.1.6) does not hold.

In order to overcome this problem, Ledford and Tawn (1996,1997,1998) (see also Coles et al., 1999) introduced a quite general submodel, where the tail dependence is characterized by a coefficient $\eta \in (0,1]$. More precisely, in the setting with uniform marginals, they assumed that the function $t \mapsto P\{1-X < t \text{ and } 1-Y < t\}$ is regularly varying at 0 with index $1/\eta$. Then $\eta = 1$ in case of asymptotic dependence, whereas $\eta < 1$ implies asymptotic independence. When η is less than 1, the value of η determines the amount of dependence in asymptotic independence (see (6.2.1) below and the comments thereafter). Thus the submodel can also be used to device a test for asymptotic independence in the basic relation (6.1.2).

Moreover Ledford and Tawn proposed an estimator for η. Peng (1999) presented a theoretical background for their model and proposed a non-parametric estimator for η. Peng proved asymptotic normality of his estimator under second order conditions. The present paper contains the following contributions:

(1) Peng's conditions are generalised so that, e.g., the normal distribution is included.

(2) Asymptotic normality of two modified versions of estimators introduced by Ledford and Tawn is shown under second order conditions (Section 6.2).

(3) A new estimator is introduced and its asymptotic normality is derived (Section 6.2).

(4) A procedure is set up to estimate the probability of a failure set that works under asymptotic dependence as well as under asymptotic independence. The estimator is proved to be consistent in our model (Section 6.3).

(5) A simulation study compares the behavior of the estimators and their use in testing for asymptotic independence. Also the behavior of the estimator for failure probabilities is studied in a simple situation (Section 6.4).

In Section 6.5 we examine the dependence between still water level, wave heights and wave periods at a particular point of the Dutch coastal protection. Sections 6.6 and 6.7 contain the proofs of the results of Section 6.2 and Section 6.3 respectively. An Appendix provides some helpful analytical results.

6.2 Estimating asymptotic dependence or independence

Let (X,Y) be a random vector whose distribution function F has continuous marginal distribution functions F_1 and F_2. Our basic assumption is that

$$\lim_{t\downarrow 0}\Big(\frac{P\{1-F_1(X)<tx \text{ and } 1-F_2(Y)<ty\}}{q(t)}-c(x,y)\Big)\Big/q_1(t)=:c_1(x,y) \quad (6.2.1)$$

exists, for $x,y\geq 0$ (but $x+y>0$), with q positive, $q_1\to 0$ as $t\to 0$ and c_1 non-constant and not a multiple of c. Moreover we assume that the convergence is uniform on

$$\{(x,y)\in[0,\infty)^2 \mid x^2+y^2=1\}.$$

It follows that the function q is regularly varying at zero of order $1/\eta$, $\eta\in(0,1]$; q_1 is also regularly varying at zero, but with order $\tau\geq 0$. Without loss of generality we may take $c(1,1)=1$, and we may assume that $q(t)=P\{1-F_1(X)<t \text{ and } 1-F_2(Y)<t\}$ (see Appendix). We also assume that $l:=\lim_{t\downarrow 0}q(t)/t$, exists. Since $F_1(X)$ and $F_2(Y)$ are uniformly distributed, obviously $\limsup q(t)/t\leq 1$, and $l=0$ when $\eta<1$. Our assumptions imply that (6.2.1) holds locally uniformly on $(0,\infty)^2$ (see Appendix). The bivariate normal distribution satisfies these conditions: see the example at the end of this section.

The function c is homogeneous of order $1/\eta$, i.e., $c(tx,ty)=t^{1/\eta}c(x,y)$. The measure ν defined by $\nu([0,x]\times[0,y])=c(x,y)$ inherits this homogeneity:

$$\nu(tA)=t^{1/\eta}\nu(A) \quad (6.2.2)$$

for $t>0$ and all bounded Borel sets $A\subset[0,\infty)^2$.

The parameter η is Ledford's and Tawn's coefficient of asymptotic dependence, cf. Ledford and Tawn (1996,1997). Now $l>0$ implies asymptotic dependence, and $l=0$ implies asymptotic independence. Hence $\eta<1$ implies asymptotic independence. Condition (6.2.1) is somewhat similar to condition (2.8) in Ledford and Tawn (1998).

Now we turn to estimators for η, given an i.i.d. sample $\{(X_1,Y_1),(X_2,Y_2),\dots(X_n,Y_n)\}$. We start with an informal introduction to the estimators of Ledford and Tawn (1996). They proposed first to standardize the marginals to the unit Fréchet distribution, using either the empirical marginal distributions (that is, using the ranks of the components) or extreme value estimators for the marginal tails, and then to estimate η as the shape parameter of the minimum of the components, e.g. by the maximum likelihood estimator or the Hill estimator. However, since these estimators have larger bias for Fréchet distributions than for Pareto distributions, we prefer to standardize to the unit Pareto distribution using the ranks of the components.

For this consider the random vector

$$T := \frac{1}{1 - F_1(X)} \wedge \frac{1}{1 - F_2(Y)}$$

which is in the domain of attraction of the extreme value distribution with parameter $1/\eta$. Since the marginal d.f.'s F_i are unknown, we replace them with their empirical counterparts. This leads to (with a small modification to prevent division by 0):

$$T_i^{(n)} := \frac{n+1}{n+1-R_i^X} \wedge \frac{n+1}{n+1-R_i^Y}, \quad i = 1, \dots, n,$$

with R_i^X denoting the rank of X_i among $(X_1, X_2, \dots, X_n)$ and R_i^Y that of Y_i. Now η can be estimated by the maximum likelihood estimator in a generalised Pareto model, based on the largest $m = m(n)$ order statistics of the $T_i^{(n)}$. This estimator will be denoted by $\hat{\eta}_1$. Alternatively the Hill estimator can be used:

$$\hat{\eta}_2 := \frac{1}{m} \sum_{i=1}^{m} \log \frac{T_{n,n-i+1}^{(n)}}{T_{n,n-m}^{(n)}}.$$

Note that one important advantage of the maximum likelihood estimator over the Hill estimator in the classical i.i.d. setting, namely its location invariance, is not relevant here: there is no shift after standardizing the marginals to unit Pareto (see Lemma 6.6.3). Since $\hat{\eta}_2$ has smaller variance, one might expect $\hat{\eta}_2$ to outperform $\hat{\eta}_1$ (however, see Section 6.4).

Next we introduce Peng's estimator and our new proposal. Equation (6.2.1) implies for $k/n \to 0$ and $s > 0$

$$\frac{P\{1 - F_1(X) < s\,k/n \text{ and } 1 - F_2(Y) < s\,k/n\}}{P\{1 - F_1(X) < k/n \text{ and } 1 - F_2(Y) < k/n\}} = s^{1/\eta}(1 + o(1)) \qquad (6.2.3)$$

locally uniformly. Denote by $X_{n,i}$ and $Y_{n,i}$ the ith order statistics of the X_j and Y_j, $j = 1, \dots, n$, respectively. To estimate η from the sample we may replace in (6.2.3) P, $1 - F_1$ and $1 - F_2$ by their empirical counterparts. Write

$$S_n(j,k) := \sum_{i=1}^{n} \mathbf{1}\{X_i > X_{n,n-j} \text{ and } Y_i > Y_{n,n-k}\}. \qquad (6.2.4)$$

Note that $S_n(j,k) \approx n\,P\{1 - F_1(X) < j/n \text{ and } 1 - F_2(Y) < k/n\}$.

Using $s = 2$ in (6.2.3) leads to Peng's (1999) estimator:

$$\hat{\eta}_3 = \log 2 / \log\Big(\frac{S_n(2k,2k)}{S_n(k,k)}\Big).$$

We propose the following estimator, based on integrating (6.2.3) with respect to s from 0 to 1:

$$\hat{\eta}_4 := \frac{\sum_{j=1}^{k} S_n(j,j)}{kS_n(k,k) - \sum_{j=1}^{k} S_n(j,j)} \tag{6.2.5}$$

with S_n as in equation (6.2.4).

Note that $\hat{\eta}_1$ and $\hat{\eta}_2$ are based on the empirical quantile function, and $\hat{\eta}_3$ and $\hat{\eta}_4$ on the empirical distribution function.

We first have to prove the consistency of the new estimator.

Theorem 6.2.1 (Consistency). *Suppose for* $x, y \geq 0$

$$\lim_{t\to 0} \frac{\Pr\{1 - F_1(X) \leq tx \text{ and } 1 - F_2(Y) \leq ty\}}{q(t)} = c(x,y) \tag{6.2.6}$$

where q *and* c *are positive functions. Let* $k = k(n)$, $r(n) := n\, q(k/n) \to \infty$ *(this implies* $k \to \infty$*) and* $k/n \to 0$ *for* $n \to \infty$*. Then*

$$\hat{\eta}_4 \to \eta$$

in probability, with η *the reciprocal of the index of regular variation of* q *at 0.*

Remark 6.2.1. Note (6.2.1) is not needed here. Moreover $\hat{\eta}_1$ and $\hat{\eta}_2$ are consistent too if $m = [r(n)]$.

The next theorem states the asymptotic normality of all estimators considered.

Theorem 6.2.2 (Asymptotic normality). *Assume* (6.2.1). *Additionally assume that* c *has first order partial derivatives* $c_x = \frac{\partial}{\partial x} c(x,y)$ *and* $c_y = \frac{\partial}{\partial y} c(x,y)$*. Suppose* $k = k(n)$, $r(n) = n\, q(k/n) \to \infty$ *(this implies* $k \to \infty$*),* $k/n \to 0$, $\sqrt{r(n)}q_1(k/n) \to 0$ *as* $n \to \infty$*, and* $m = m(n) = [r(n)]$.

Under these conditions the $\sqrt{r(n)}\big(\hat{\eta}_i - \eta\big)$ *are asymptotically normal with mean*

0 and variance σ_i^2*,* $i = 1, 2, 3, 4$*. The variances are*

$$\sigma_1^2 = (1+\eta)^2(1-l)(1-2lc_x(1,1)c_y(1,1)) \tag{6.2.7}$$

$$\sigma_2^2 = \eta^2(1-l)(1-2lc_x(1,1)c_y(1,1)) \tag{6.2.8}$$

$$\begin{aligned}\sigma_3^2 = 2\eta^4 (\log 2)^{-2} (1-2^{-1/\eta})\Big[\frac{1}{2}(1-3l)(1-2lc_x(1,1)c_y(1,1)) \\ + lc(1,2)c_x(1,1)(1-lc_y(1,1)) + lc(2,1)c_y(1,1)(1-lc_x(1,1))\Big]\end{aligned} \tag{6.2.9}$$

$$\begin{aligned}\sigma_4^2 = \frac{(1+\eta)^2\eta^2}{2\eta+1}\Big[(1-3l)(1-2lc_x(1,1)c_y(1,1)) \\ + 4lc_x(1,1)(1-lc_y(1,1))\int_0^1 c(u,1)du \\ + 4lc_y(1,1)(1-lc_x(1,1))\int_0^1 c(1,u)du\Big].\end{aligned} \tag{6.2.10}$$

Remark 6.2.2. The assertion for $\hat{\eta}_3$ is a generalisation of Peng's, since our conditions are weaker.

Remark 6.2.3. Note that instead of (6.2.1) the weaker condition $\lim_{t\to 0} P\{1 - F_1(X) < tx \text{ and } 1 - F_2(Y) < ty\}/q(t) - c(x,y) = O(q_1(t))$ is sufficient for Theorem 6.2.2. However, under (6.2.1) similar results can be easily deduced if the intermediate sequence k is such that $\sqrt{r(n)}q_1(k/n) \to c \geq 0$. In that case, usually a non-negligible bias occurs if $c > 0$ (and the present results correspond to the simpler case $c = 0$).

Theorem 6.2.2 may be stated without the unknown sequence $r(n)$ entering explicitly the formulation, as in the following corollary.

Corollary 6.2.1. *Assume the conditions of Theorem 6.2.2. For* $i = 3, 4$

$$\sqrt{S_n(k,k)}(\hat{\eta}_i - \eta)$$

has the limiting distribution of Theorem 6.2.2, with $S_n(j,k)$ *as in equation* (6.2.4)*.*

Remark 6.2.4. When using $\hat{\eta}_1$ or $\hat{\eta}_2$, the choice of the number $m = [r(n)]$ of largest order statistics from $T_{n,i}^{(n)}$ is up to the statistician, so there is no need to estimate $r(n)$.

Corollary 6.2.1, together with consistent estimators for the unknown quantities in the asymptotic variances in Theorem 6.2.2, can be used to construct a confidence interval for η or to test the hypothesis $\eta = 1$. The following theorem provides these estimators.

Theorem 6.2.3. *(i) Define*

$$\begin{aligned}
\hat{c}_x(1,1) &:= k^{1/4}\frac{S_n([k(1+k^{-1/4})],k)-S_n(k,k)}{S_n(k,k)},\\
\hat{c}_y(1,1) &:= k^{1/4}\frac{S_n(k,[k(1+k^{-1/4})])-S_n(k,k)}{S_n(k,k)},\\
\hat{d}_1 &:= \frac{\sum_{j=1}^k S_n(j,k)}{k\, S_n(k,k)},\\
\hat{d}_2 &:= \frac{\sum_{j=1}^k S_n(k,j)}{k\, S_n(k,k)},\\
\hat{l} &:= \frac{S_n(k,k)}{k}
\end{aligned}$$

with $S_n(i,j)$ as in equation (6.2.4). If the conditions of Theorem 6.2.2 hold then

$$\hat{l} \to_p l.$$

If, in addition, $\eta > 1/2$ then

$$\hat{c}_x(1,1) \to_p c_x(1,1), \qquad \hat{c}_y(1,1) \to_p c_y(1,1),$$
$$\hat{d}_1 \to_p \int_0^1 c(u,1)du, \qquad \hat{d}_2 \to_P \int_0^1 c(1,u)du.$$

Moreover, let

$$\hat{\sigma}_1^2 := (1+\hat{\eta})^2(1-\hat{l})(1-2\hat{l}\hat{c}_x(1,1)\hat{c}_y(1,1))$$

and define $\hat{\sigma}_i^2$, $i = 2,3,4$, likewise by (2.8)–(2.10) with $\eta, l, c_x(1,1), c_y(1,1)$, $\int_0^1 c(u,1)\,du$ and $\int_0^1 c(1,u)\,du$ replaced by their respective estimator. Then $\hat{\sigma}_i^2$, $i = 1,\dots,4$, are consistent estimators of σ_i^2 for all $\eta \in (0,1]$.

(ii) The analogous assertion to (i) holds for the estimators

$$\tilde{l} := \frac{m}{n} T_{n,n-m}^{(n)}$$
$$\tilde{c}_x(1,1) := \frac{\hat{k}^{5/4}}{n}\left(T_{n,m}^{(n,\hat{k}^{-1/4})} - T_{n,m}^{(n)}\right)$$

with $m := m(n) := [r(n)]$, $\hat{k} := m/\tilde{l}$, and $T_{n,i}^{(n,u)}, i = 1,\dots,n$, the order statistics of

$$T_i^{(n,u)} := \min\left(\frac{n+1}{n+1-R_i^X}(1+u), \frac{n+1}{n+1-R_i^Y}\right), \quad i = 1,\dots,n,$$

and $\tilde{c}_y(1,1)$ defined analogously to $\tilde{c}_x(1,1)$.

Remark 6.2.5. Note that $\tilde{c}_y(1,1)$ may also be estimated as $1/\hat{\eta}-\tilde{c}_x(1,1)$, provided $\eta > 1/2$.

Example 6.2.1. The bivariate normal distribution with mean 0, variance 1 and correlation coefficient $\rho \notin \{1,-1\}$, satisfies (6.2.1) with

$$\eta = (1+\rho)/2, \qquad c(x,y) = (xy)^{1/(1+\rho)},$$

$$q(t) = k_1(\rho) t^{2/(1+\rho)} (-\log t)^{-\rho/(1+\rho)} \left\{ 1 - k_2(\rho) \frac{\log(-\log t)}{2\log t} \right\},$$

$$c_1(x,y) = -k_3(\rho) - k_4(x,y,\rho), \qquad q_1(t) = \frac{1}{2\log t},$$

where

$$k_1(\rho) = \frac{(1-\rho^2)^{3/2}}{(1-\rho)^2}(4\pi)^{-\rho/(1+\rho)}, \qquad k_2(\rho) = \frac{\rho}{1+\rho},$$

$$k_3(\rho) = \frac{\rho\log(4\pi)+2}{1+\rho} - \frac{(1+\rho)(2-\rho)}{1-\rho},$$

$$\begin{aligned} k_4(x,y,\rho) &= \log x + \log y \\ &\quad + \frac{(\rho-1)(\log x + \log y) + \rho(\log x)(\log y) - \rho\left((\log x)^2 + (\log y)^2\right)/2}{(1-\rho^2)}. \end{aligned}$$

This can be checked using the tail expansion of the bivariate normal distribution by Ruben(1964) as given in Ledford and Tawn (1997), combined with a sufficiently precise expansion of the function f, the inverse function of $1/(1-\Phi)$ where Φ is the standard univariate normal distribution function:

$$\begin{aligned} f^2(t) &= 2\log t - \log(\log t) - \log(4\pi) + \frac{\log(\log t)}{2\log t} + \frac{\log(4\pi)-2}{2\log t} \\ &\quad + \frac{1}{2}\left(\frac{\log(\log t)}{2\log t}\right)^2 + o\left(\left(\frac{\log(\log t)}{\log t}\right)^2\right), \text{ as t} \to \infty. \end{aligned}$$

6.3 Estimation of failure probabilities

Throughout this section we assume that the marginal distribution functions F_i of F are continuous and belong to the domain of attraction of a univariate extreme value distribution. Moreover, condition (6.2.6) and further conditions ensuring $\hat{\eta} - \eta = O_P((r(n))^{-1/2})$ shall hold (cf. Section 6.2).

Recall from (6.1.6) that, if we want to estimate the probability of an extreme set of the form $\{X > x \text{ or } Y > y\}$ and we assume that F belongs to the domain of attraction of a bivariate extreme value distribution, then we can use the approximate equality

$$\begin{aligned} &P\{1 - F_1(X) < 1 - F_1(x) \text{ or } 1 - F_2(Y) < 1 - F_2(y)\} \\ &\qquad \approx tP\{1 - F_1(X) < (1 - F_1(x))/t \text{ or } 1 - F_2(Y) < (1 - F_2(y))/t\} \end{aligned} \tag{6.3.1}$$

since for small t the right hand side can be estimated using the empirical distribution function de Haan and Sinha (1999). However, if the marginals are asymptotically independent and the failure set is e.g. of the form $\{X > x \textit{ and } Y > y\}$ then a different approximate equality holds under condition (6.2.1) or (6.2.6):

$$\begin{aligned} &P\{1 - F_1(X) < 1 - F_1(x) \text{ and } 1 - F_2(Y) < 1 - F_2(y)\} \\ &\qquad \approx t^{1/\eta} P\{1 - F_1(X) < (1 - F_1(x))/t \text{ and } 1 - F_2(Y) < (1 - F_2(y))/t\}. \end{aligned} \tag{6.3.2}$$

We develop an estimation procedure which works in this situation.

More generally, we aim at the estimation of the failure probability $p_n = P\{(X, Y) \in C_n\}$ for failure regions $C_n \subset [x_n, \infty] \times [y_n, \infty]$ for some $x_n, y_n \in \mathbb{R}$ such that

$$(x, y) \in C_n \quad \Longrightarrow \quad [x, \infty] \times [y, \infty] \subset C_n. \tag{6.3.3}$$

The latter property means that if an observation (x, y) causes a failure (e.g., a flooding of a dike) then an event with both components larger will do so, too. Asymptotically we let both x_n and y_n converge to the right endpoint of the pertaining marginal distribution to ensure that $p_n \to 0$, i.e., that indeed we are estimating the probability of an extremal event.

The basic idea is to use a generalised version of the scaling property (6.3.2) to inflate the transformed failure set $(1-F_1, 1-F_2)(C_n) := \{(1-F_1(x), 1-F_2(y)) \mid (x, y) \in C_n\}$ such that it contains sufficiently many observations and hence the empirical probability gives an accurate estimate. Since the marginal distribution functions F_i are unknown, their tails are estimated by suitable generalised Pareto distributions.

To work out this program, first recall from univariate extreme value theory that there exist normalising constants $a_i(n/k) > 0$ and $b_i(n/k) \in \mathbb{R}$ such that the following generalised Pareto approximation is valid:

$$1 - F_i(x) \approx \frac{k}{n}\Big(1 + \gamma_i \frac{x - b_i(n/k)}{a_i(n/k)}\Big)^{-1/\gamma_i} =: \frac{k}{n}(1 - F_{a_i, b_i, \gamma_i}(x)), \quad i = 1, 2,$$

for x close to the right endpoint $F_i^{-1}(1)$. Here a_i and b_i are abbreviations for $a_i(n/k)$ and $b_i(n/k)$, respectively, and $(1 + \gamma x)^{-1/\gamma}$ is defined as ∞ if $\gamma > 0$ and $x \leq -1/\gamma$, and it is defined as 0 if $\gamma < 0$ and $x \geq -1/\gamma$. Dekkers et al. (1989) proposed and analyzed the following estimators of the parameters a_i, b_i and γ_i. Define

$$\begin{aligned}
M_r(X) &:= \frac{1}{k}\sum_{j=0}^{k-1}(\log X_{n,n-j} - \log X_{n,n-k})^r,\ r = 1,2, \\
\hat{\gamma}_1 &:= M_1(X) + 1 - \frac{1}{2}\left(1 - \frac{(M_1(X))^2}{M_2(X)}\right)^{-1}, \\
\hat{b}_1\left(\frac{n}{k}\right) &:= X_{n,n-k}, \\
\hat{a}_1\left(\frac{n}{k}\right) &:= \frac{X_{n,n-k}\sqrt{3M_1(X)^2 - M_2(X)}}{\sqrt{(1-4\hat{\gamma}_1^-)/((1-\hat{\gamma}_1^-)^2(1-2\hat{\gamma}_1^-))}} \quad \text{with} \quad \hat{\gamma}_1^- := \hat{\gamma}_1 \wedge 0,
\end{aligned}$$

for $\hat{\gamma}_2$, $\hat{a}_2$ and $\hat{b}_2$ replace X by Y in the previous formulas. The estimator $\hat{\gamma}_i$ for the extreme value index γ_i is often called moment estimator.

Using these definitions, $\frac{n}{k}(1 - F_i(x))$ may be estimated by

$$1 - F_{\hat{a}_i,\hat{b}_i,\hat{\gamma}_i}(x) = \left(1 + \hat{\gamma}_i \frac{x - \hat{b}_i(n/k)}{\hat{a}_i(n/k)}\right)^{-1/\hat{\gamma}_i}.$$

Write $\mathbf{1} - \boldsymbol{F}(x,y)$ as a short form for $(1 - F_1(x), 1 - F_2(y))$, and likewise $\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}} = (1 - F_{a_1,b_1,\gamma_1}, 1 - F_{a_2,b_2,\gamma_2})$ and $\mathbf{1} - \boldsymbol{F}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}} = (1 - F_{\hat{a}_1,\hat{b}_1,\hat{\gamma}_1}, 1 - F_{\hat{a}_2,\hat{b}_2,\hat{\gamma}_2})$ are functions from $\mathbb{R}^2$ to $[0,\infty]^2$. Then the transformed failure set $\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(C_n))$ can be approximated by

$$D_n := \mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(C_n)$$

which in turn is estimated by

$$\hat{D}_n := \mathbf{1} - \boldsymbol{F}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(C_n).$$

Now we may argue heuristically as follows, using a generalisation of the scaling property (6.3.2) to inflate the transformed failure set by the factor $1/c_n$ for some $c_n \to 0$ chosen in a suitable way:

$$\begin{aligned}
p_n &= P\{\mathbf{1} - \boldsymbol{F}(X,Y) \in \mathbf{1} - \boldsymbol{F}(C_n)\} \\
&\approx P\Big\{\frac{n}{k}\big(\mathbf{1} - \boldsymbol{F}(X,Y)\big) \in D_n\Big\} \\
&\approx c_n^{1/\eta} P\Big\{\frac{n}{k}\big(\mathbf{1} - \boldsymbol{F}(X,Y)\big) \in \frac{D_n}{c_n}\Big\} && (6.3.4) \\
&\approx c_n^{1/\hat{\eta}} P\{(X,Y) \in B\}_{|B = \boldsymbol{F}^{-\mathbf{1}}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1} - \frac{\hat{D}_n}{c_n})} \\
&\approx c_n^{1/\hat{\eta}} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\Big\{(X_i,Y_i) \in \boldsymbol{F}^{-\mathbf{1}}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\Big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\Big)\Big\} && (6.3.5) \\
&=: \hat{p}_n && (6.3.6)
\end{aligned}$$

where $\hat{\eta}$ denotes one of the estimators examined in Section 6.2.

In the sequel we state the exact conditions under that we will prove consistency of the estimator $\hat{p}_n$, that is, $\hat{p}_n/p_n \to 1$ in probability as $n \to \infty$. In order not to overload the paper, we will not determine the nondegenerate limit distribution of the standardized estimation error. However, employing the ideas of de Haan and Sinha (1999), one may establish asymptotic normality of $\hat{p}_n$ under more complex conditions.

To study the asymptotic behavior of $\hat{p}_n$, we have to impose a regularity condition on the sequence of failure sets C_n, or rather on the transformed sets D_n. Note that D_n shall shrink towards the origin because we are interested in extremal events. We assume that, after a suitable standardization, D_n converges in the following sense:

(D) There exist a sequence $d_n \to 0$ and a measurable bounded set $A \subset [0,\infty)^2$ with $\nu(A) > 0$ such that for all $\varepsilon > 0$ one has for sufficiently large n

$$A_{-\varepsilon} \subset \frac{D_n}{d_n} \subset A_{+\varepsilon}.$$

Here $A_{+\varepsilon} := \{\boldsymbol{x} \in [0,\infty)^2 \mid \inf_{\boldsymbol{y}\in A} \|\boldsymbol{x}-\boldsymbol{y}\| \le \varepsilon\}$ and $A_{-\varepsilon} := [0,\infty)^2 \setminus (([0,\infty)^2 \setminus A)_{+\varepsilon})$ denote the outer and inner ε–neighborhood of A with respect to the maximum norm $\|\boldsymbol{x}-\boldsymbol{y}\| = |x_1-y_1| \vee |x_2-y_2|$, and ν is the measure corresponding to the function c (cf. Section 6.2).

Note that d_n and A are not determined by this condition as the former may be multiplied by a fixed factor and the latter divided by the same number. Moreover, even for given d_n the set A is determined only up to its boundary.

Condition (6.3.3) on C_n implies

$$(x,y) \in D_n \quad \Longrightarrow \quad [0,x]\times[0,y] \subset D_n. \tag{6.3.7}$$

Example 6.3.1. For $C_n = [x_n,\infty]\times[y_n,\infty]$ we have $D_n = [0, 1-F_{a_1,b_1,\gamma_1}(x_n)] \times [0, 1-F_{a_2,b_2,\gamma_2}(y_n)]$. Hence (D) is satisfied with $d_n = 1-F_{a_1,b_1,\gamma_1}(x_n)$ if $(1-F_{a_2,b_2,\gamma_2}(y_n))/(1-F_{a_1,b_1,\gamma_1}(x_n))$ converges in $(0,\infty)$.

This example demonstrates that essentially (D) means that the convergence of the failure set in the x- and the y-direction is balanced.

Next we need a certain rate of convergence for the marginal estimators to ensure that the transformation of the failure set does not introduce too big an error. For that purpose recall that

$$R_i(t,x) := t\big(1 - F_i(a_i(t)x + b_i(t))\big) - (1+\gamma_i x)^{-1/\gamma_i} \;\to\; 0, \quad i = 1,2,$$

locally uniformly for $x \in (0,\infty]$ as $t \to \infty$, since F_i belongs to the domain of attraction of an extreme value distribution. Here we impose the following slightly stricter condition:

$$R_{x_1,x_2}(t) := \max_{i=1,2} \sup_{x_i < x < 1/((-\gamma_i)\vee 0)} \left| R_i(t,x)(1+\gamma_i x)^{1/\gamma_i} \right| \;\to\; 0 \tag{6.3.8}$$

for some $-1/(\gamma_i \vee 0) < x_i < 1/((-\gamma_i) \vee 0)$, $i = 1, 2$. Observe that then (6.3.8) even holds for all such x_i. For example, if F_i satisfies the second order condition

$$\frac{R_i(t,x)}{A_i(t)} \to \Psi(x)$$

for some ρ_i–varying function A_i with $\rho_i < 0$ $(i = 1, 2)$, then (6.3.8) holds true with $R_{x_1,x_2}(t) = O(A_1(t) \vee A_2(t))$. In addition, we require that not too many order statistics are used for estimation of the marginal parameters:

$$k^{1/2} R_{x,x}\Big(\frac{n}{k}\Big) = O(1) \tag{6.3.9}$$

for some $x < 0$. Then it follows that the estimators $\hat{a}_i$, $\hat{b}_i$ and $\hat{\gamma}_i$ are $\sqrt{k}$–consistent in the following sense:

$$\Big|\frac{\hat{a}_i}{a_i} - 1\Big| \vee \Big|\frac{\hat{b}_i - b_i}{a_i}\Big| \vee |\hat{\gamma}_i - \gamma_i| = O_P(k^{-1/2}), \quad i = 1, 2 \tag{6.3.10}$$

(cf. Dekkers et al., 1989; de Haan and Resnick, 1993).

We will see that using the estimated parameters instead of the unknown true ones for the transformation of the failure sets does not cause problems provided

$$w_{\gamma_1 \wedge \gamma_2}(d_n) = o(k^{1/2}) \quad \text{with} \quad w_\gamma(x) := -x^\gamma \int_x^1 u^{-\gamma-1} \log u \, du. \tag{6.3.11}$$

Check that

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

as $x \to 0$. Though, at first glance, (6.3.11) seems rather strict a condition if one of the extreme value indices is negative, it is indeed a natural one; for without it the difference between the transformed set D_n and its estimate $\hat{D}_n$ would be at least of the same order in probability as the typical elements of D_n, namely at least of the order d_n, which of course would render impossible any further statistical inference on the failure probability.

In addition, the scaling factor c_n chosen by the statistician when applying the estimator $\hat{p}_n$ must be related to the actual scaling factor d_n as follows:

$$d_n = O(c_n), \quad w_{\gamma_1 \wedge \gamma_2}\Big(\frac{c_n}{d_n}\Big) = o(k^{1/2}) \quad \text{and} \quad \Big(\frac{c_n}{d_n}\Big)^{1/\eta} = o\big((r(n))^{1/2}\big). \tag{6.3.12}$$

In particular, (6.3.12) is satisfied if c_n and d_n are of the same order. Below the choice of c_n is discussed more thoroughly.

Recall from Section 6.2 that the scaling property (6.3.2) is a consequence of approximation (6.2.6) and the homogeneity of the measure ν. In order to justify

(6.3.4) in the motivation for $\hat{p}_n$ given above, we need the following modification of (6.2.6), which is suitable for more general sets than just upper quadrants:

$$\sup_{B\in\bar{\mathcal{B}}_n}\left|\frac{P\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}}{q(k/n)\nu\big(\frac{n}{k}(\mathbf{1}-\boldsymbol{F}(B))\big)}-1\right|\to 0\quad\text{as } n\to\infty \tag{6.3.13}$$

where

$$\mathcal{B}_n := \Big\{\boldsymbol{F}^{-1}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}}}\big(\mathbf{1}-\frac{\mathbf{1}-\boldsymbol{F}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}}}(C_n)}{c_n}\big)\;\Big|\;\|\frac{\tilde{\boldsymbol{a}}}{\boldsymbol{a}}-1\|\vee\|\frac{\tilde{\boldsymbol{b}}-\boldsymbol{b}}{\boldsymbol{a}}\|\vee\|\tilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|\le\varepsilon_n\Big\}$$

for some $\varepsilon_n\to 0$ such that $k^{1/2}\varepsilon_n\to\infty$, and

$$\bar{\mathcal{B}}_n := \mathcal{B}_n\cup\Big\{C_n,\bigcup_{B\in\mathcal{B}_m,m\ge n}B\Big\}.$$

It will turn out (see (6.7.7)) that for sufficiently large n the denominator in (6.3.13) is strictly positive.

Notice that the convergence of the absolute value in (6.3.13) for sets of the type $\mathbf{1}-\boldsymbol{F}(B)=[0,xk/n]\times[0,yk/n]$ is equivalent to convergence (6.2.6) with $t=k/n$.

Finally, to make approximation (6.3.5) rigorous, we need a kind of uniform law of large numbers. This is provided by the theory of Vapnik-Cervonenkis (VC) classes of sets as outlined, e.g., in the monograph by Pollard (1984, Section II.4). For this we require

$$\mathcal{B}=\bigcup_{n\in\mathbb{N}}\mathcal{B}_n\ \text{ is a VC class.} \tag{6.3.14}$$

Theorem 6.3.1. *Suppose the conditions (D),* (6.3.3) *(or* (6.3.7)*),* (6.3.8), (6.3.9) *and* (6.3.11)–(6.3.14) *are satisfied. If* $\hat{\eta}-\eta=O_P((r(n))^{-1/2})$, $\log c_n=o((r(n))^{1/2})$, *and* $k(n)/n$ *is almost decreasing, which means* $\sup_{m\ge n}k(m)/m=O(k(n)/n)$, *then*

$$\frac{\hat{p}_n}{p_n}\to 1\quad\textit{in probability.}$$

Remark 6.3.1. (i) In the most important case that np_n is bounded, the conditions (6.3.11)–(6.3.13) can be jointly satisfied only if $\gamma_1\wedge\gamma_2>-1/2$.

(ii) The sequence $k(n)/n$ is almost decreasing, e.g., if $k(n)$ is regularly varying with exponent less than 1 or, more general, has an upper Matuszewska index $\alpha\le 1$ (see Bingham et al., 1987; Theorem 2.2.2).

The scaling factor $1/c_n$ by which the transformed failure set is inflated determines the number of large observations taken into account for the empirical probability (6.3.5). More precisely, according to (6.7.8) in the proof of Lemma 6.7.4, this number is of the order $r(n)(d_n/c_n)^{1/\eta}$. Hence if d_n and c_n are of the same order then one uses essentially the same number $S_n(k,k)$ of observations as for the estimation of η, which seems quite natural.

In practice, of course, d_n is not known. However, conversely one may choose c_n such that about $S_n(k,k)$ observations lie in the inflated set $\hat{D}_n/c_n$. To be more concrete, let

$$c_n(\lambda) := \sup\Big\{c > 0 \,\Big|\, \sum_{i=1}^{n} \mathbf{1}\big\{(X_i, Y_i) \in \boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c}\big)\big\} \geq \lambda S_n(k,k)\Big\} \qquad (6.3.15)$$

for some $\lambda > 0$. Following the lines of the proof of Theorem 6.3.1, one may show that $c_n(\lambda)$ and d_n are of the same order in probability, and that the resulting estimator $\hat{p}_n$ is consistent for p_n. Alternatively, one may employ a heuristic approach which is common in univariate extreme value statistics: one plots $\hat{p}_n$ as a function of c_n and choose a value c_n where this graph seems sufficiently stable.

Finally, it is worth mentioning that it is not necessary to use the same number k in the estimation of the marginal parameters and in the definition of $\hat{\eta}_3$ and $\hat{\eta}_4$. In fact, one may prove an analog to Theorem 6.3.1 in more general settings, provided it is guaranteed that the estimation error introduced when standardizing the marginals is asymptotically negligible, that is, one has (6.3.10), (6.3.11) and (6.3.12) for some k which may differ from the one used in the definition of the estimator for η. Likewise one may use other estimators for the marginal parameters, like e.g. the maximum likelihood estimator examined by Smith(1987), provided these estimators converge with the same rate.

6.4 Simulations

6.4.1 Methods

The estimators were tested on 4 different distribution functions:

(1) the bivariate Cauchy distribution ($\eta = 1$),

(2) the bivariate extreme value distribution (BEV) with a logistic dependence function, with $\alpha = 0.75$ ($\eta = 1$), Ledford and Tawn (1996,1997),

(3) the bivariate normal distribution with $\rho = 0.6$ ($\eta = 0.8$) and

(4) the Morgenstern distribution with $\alpha = 0.75$ ($\eta = 0.5$), Ledford and Tawn (1996,1997).

From each distribution we generated 250 samples of size 1000. A sample of each the distributions is shown in Figure 1. Dependence in these distributions ranges from clear asymptotic dependence (1) through weak asymptotic dependence (2) and non-asymptotic dependence (3) to clear asymptotic independence (4).

The ML-estimator $\hat{\eta}_1$ was estimated by the GAUSS `maxlik` procedure Shoenberg(1996).

For each estimator two estimates for the root variance are reported:

$\hat{\sigma}_{(i)}$ the root variance for the general case, calculated as $m^{-1/2}\hat{\sigma}_i$, $i = 1, 2$, resp. $S_n(k,k)^{-1/2}\ \hat{\sigma}_i, i = 3, 4$ (cf. Theorems 6.2.2 and 6.2.3), and

$\hat{\sigma}_{(d)}$ the root variance for the case of asymptotic dependence, calculated similarly, with $\eta = 1$.

For comparison the observed empirical standard deviation was calculated from the 250 simulated η estimates.

Correspondingly, one-sided 5% tests for dependence were carried out in two ways. Asymptotic dependence is not rejected when

$$\Phi((1-\hat{\eta})/\hat{\sigma}_{(i)}) \le 0.95 \quad \text{or alternatively} \quad \Phi((1-\hat{\eta})/\hat{\sigma}_{(d)}) \le 0.95$$

where Φ represents the standard normal df.

Furthermore, we studied the finite sample behavior of the proposed estimators of a failure probability. For this, failure sets of the type $[a,\infty)^2$ were considered, where a was chosen such that the failure probability equals $p_n = (100n)^{-1} = 10^{-5}$ for the sample size $n = 1000$. We used $\hat{\eta}_4$ as the estimator for the parameter of tail dependence and considered three different estimators of p_n:

$$\begin{aligned} \hat{p}_{\hat{\eta}} &= \hat{p}_n \quad \text{as defined in (6.3.6)}, \\ \hat{p}_1 &= c_n^{1-1/\hat{\eta}}\hat{p}_n, \quad \text{and} \\ \hat{p} &= \begin{cases} \hat{p}_1 & \text{if } \eta = 1 \text{ is not rejected}, \\ \hat{p}_{\hat{\eta}} & \text{if } \eta = 1 \text{ is rejected}. \end{cases} \end{aligned} \tag{6.4.1}$$

Here $c_n = c_n(1)$ is defined by (6.3.15) and the test for $\eta = 1$ is based on the variance estimate $\hat{\sigma}_{(d)}$. Note that $\hat{p}_1$ is a natural analog to $\hat{p}_{\hat{\eta}}$ if it is known in advance that $\eta = 1$. In particular, in that case it is a consistent estimator of p_n.

For the normal distribution this resulted in many zero estimates; as this effect was caused by the poor estimates of the marginal parameters, we also considered a distribution with the marginals transformed to standard exponential.

6.4.2 Estimating η and testing for asymptotic dependence

The results are presented in Figures 2 and 3 and in Tables 1 and 2.

To make the performance of the different estimators for η comparable, m and k were chosen in a range where the overall performance of the estimator under consideration is best. This led to a smaller m for the Hill than for the maximum likelihood estimator, because of the larger bias of former. Recall that Peng's estimator is constructed from $S_n(k,k)$ and $S_n(2k,2k)$, while $\hat{\eta}_4$ is based on $S_n(j,j)$ only up to $j = k$. For that reason we chose k for $\hat{\eta}_4$ double as large as for $\hat{\eta}_3$.

The general picture is that $\hat{\eta}_2$, $\hat{\eta}_3$ and $\hat{\eta}_4$ show a bias (negative for the Cauchy, BEV and normal distributions) that increases with m or k. The ML estimator $\hat{\eta}_1$ shows no clear trend, but it is biased for two of the distributions, though considerably less than the other estimators. Note that although $\eta \le 1$, the estimates $\hat{\eta}_i$ may be larger than 1.

A comparison of the observed standard deviation with the appropriate estimates (Tables 1-2: $\hat{\sigma}_{(d)}$ for the Cauchy and BEV distributions, $\hat{\sigma}_{(i)}$ for Normal and Morgenstern) shows the estimates are reasonable good. Note that the standard deviation of $\hat{\eta}_1$ and $\hat{\eta}_2$ on one hand and of $\hat{\eta}_3$ and $\hat{\eta}_4$ on the other hand are not fully comparable as m and k have a different meaning.

Some observations:

- Peng's estimator and ours are not stable at small k leading to missing values for either $\hat{\eta}$ or $\hat{\sigma}$ or both.

- The tests for asymptotic dependence tend to accept dependence for small k or m and to reject dependence for larger values. This effect is due to the increasing bias of the estimators for η, which is not taken into account by the tests. Consequently, the effect is weakest for the test based on $\hat{\eta}_1$.

- Hill's estimator has the smallest observed and estimated variances.

- Our estimator has a somewhat smaller observed variance than Peng's, but both have relatively large variance estimates for small k. Overall the ML estimator has a variance comparable to ours but for small k resp. m it is clearly smaller.

To conclude: the outcome of all tests for asymptotic dependence depend on k, the sample fraction used. The test based on the ML estimator $\hat{\eta}_1$ has the great advantage to be less dependent on k, but it is biased to rejecting dependence. Finally, due to the smaller variance of the Hill estimator the corresponding test detects even small deviations from the hypothesis, but on the other hand, due to its considerable bias, for the Cauchy and BEV distribution the hypothesis is much more often wrongly rejected than one would expect from the nominal level of the test. This disappointing behavior indicates that the approximation of the distribution of the Hill estimator by a centered normal distribution is rather inaccurate for moderate sample sizes.

All estimators and tests would benefit from a guideline for choosing k.

6.4.3 Failure probabilities

Table 3 summarizes the main results for the failure probability estimators. The empirical distribution of the estimators for three values of k is shown in Figure 4.

For the Cauchy distribution we have asymptotic dependence, so $\hat{p}_1$ is appropriate. Figure 4 shows that it is biased for small k, probably related with the negative bias of the γ estimates of the marginal distributions. As expected $\hat{p}_{\hat{\eta}}$ has larger variance; its smaller bias for small k is sort of a surprise.

For the normal distribution the main problem is estimating the marginals. The γ_1 and γ_2 estimates are negative. This implies upper bounds for the marginals and in quite a number of cases the failure area lies outside one or both of the bounds leading to a zero estimated failure probability.

In samples with the marginals transformed to exponential the estimator behaves much better. The marginals are estimated more accurately now with $\hat{\gamma}_i \approx 0$. Still when both $\hat{\gamma}_i$ estimates are negative a number of zero estimates result. The estimator $\hat{p}_1$ assuming asymptotic dependence over-estimates the probability, while $\hat{p}_\eta$ under-estimates it.

The Morgenstern distribution has asymptotically independent marginals. The $\hat{p}_{\hat{\eta}}$ estimate is nearly unbiased for $k = 80, 160$ whereas the $\hat{p}_1$ estimate is strongly biased. Estimating the marginals does not cause problems here as the Morgenstern distribution has extreme value (Fréchet) marginals.

6.5 An application: dependence of sea state parameters

In the course of the Neptune project, financed by the European Union (grant MAS2-CT94-0081), the joint distribution of sea state variables was studied and its consequences for the seawall at Petten. The data set, supplied by the Dutch National Institute for Marine and Coastal management, consists of date, time and sea characteristics recorded from 1979 till 1991, at three-hourly intervals at the Eierland station, 20 kilometers off the Dutch coast. After a declustering routine a set of independent observations of waveheight `Hm0`, wave-periode `Tpb` and still water level `SWL` was constructed and analysed. De Haan and de Ronde (1998) concluded that the variables were asymptotically dependent, and estimated the failure probability of the Pettemer zeewering assuming asymptotic dependence between the variables. Figure 5 shows the joint distribution of pairs of these variables and illustrates the estimation of asymptotic dependence. For none of the pairs asymptotic dependence can be rejected although for quite a number of values of k the variances can not be calculated.

6.6 Proofs for Section 6.2

The first results in this section closely follow Peng (1999). We first state slightly rephrased versions of his Lemmas 2.1 and 2.2 concerning empirical probability measures. Define uniformly distributed random variables $U_i := 1 - F_1(X_i)$, $V_i := 1 - F_2(Y_i)$ and denote the order statistics by $U_{n,i}$ and $V_{n,i}$, with the convention $U_{n,0} = V_{n,0} = 0$.

We will use the following notation:

$$\begin{aligned} S_1(x,y) &:= \sum_{i=1}^{n} \mathbf{1}\{U_i \le x \text{ and } V_i \le y\}, \\ P_1(x,y) &:= P\{U_1 \le x \text{ and } V_1 \le y\}, \\ S_2(x,y) &:= \sum_{i=1}^{n} \mathbf{1}\{U_i \le x \text{ or } V_i \le y\}, \\ P_2(x,y) &:= P\{U_1 \le x \text{ or } V_1 \le y\}. \end{aligned} \tag{6.6.1}$$

Note that $S_2(x,y) = S_2(x,0) + S_2(0,y) - S_1(x,y)$, $P_2(x,y) = x + y - P_1(x,y)$, and

$S_n(j,k)$ (equation (6.2.4)) equals $S_1(U_{n,j},V_{n,k})$ a.s.

Lemma 6.6.1. *Assume* (6.2.6)*. Let* $r(n) = n\, q(k/n) \to \infty$ *(which implies* $k \to \infty$*) and* $k/n \to 0$*. Then we have*

$$\sqrt{r(n)}\Big(\frac{S_1(\frac{k}{n}x,\frac{k}{n}y)}{r(n)} - \frac{P_1(\frac{k}{n}x,\frac{k}{n}y)}{q(k/n)}\Big) \xrightarrow{D} W_1(x,y).$$

Here, and below, $\xrightarrow{D}$ *denotes convergence in distribution in* $D([0,\infty)^2)$ *and* $W_1(x,y)$ *is a Gaussian process with mean zero and covariance structure*

$$\mathrm{E}\{W_1(x_1,y_1)W_1(x_2,y_2)\} = c(x_1 \wedge x_2, y_1 \wedge y_2).$$

Proof. See Peng (1999), Huang (1992) and Einmahl (1997, Theorem 3.1). □

Corollary 6.6.1. *Assume* (6.2.1)*. If additionally the sequence* $k(n)$ *is such that* $r(n) = n\, q(k/n) \to \infty$, $k/n \to 0$ *and* $\sqrt{r(n)}\, q_1(k/n) \to 0$ *then*

$$\sqrt{r(n)}\Big(\frac{S_1(\frac{k}{n}x,\frac{k}{n}y)}{r(n)} - c(x,y)\Big) \xrightarrow{D} W_1(x,y).$$

Proof. The extra condition on the sequence $k(n)$ ensures that

$$\sqrt{r(n)}\Big(\frac{P_1(\frac{k}{n}x,\frac{k}{n}y)}{q(k/n)} - c(x,y)\Big) \to 0$$

uniformly on $[0,A]^2$ for any $A > 0$. □

Lemma 6.6.2. *Assume* (6.2.1)*. Let* $k \to \infty$ *and* $k/n \to 0$*. Then we have*

$$\sqrt{k}\Big(\frac{S_2(\frac{k}{n}x,\frac{k}{n}y)}{k} - \frac{n}{k}P_2(\frac{k}{n}x,\frac{k}{n}y)\Big) \xrightarrow{D} W_2(x,y).$$

Here $W_2(x,y)$ *is a Gaussian process with mean zero and covariance structure*

$$\begin{aligned}\mathrm{E}\{W_2(x_1,y_1)W_2(x_2,y_2)\} &= x_1 \wedge x_2 + y_1 \wedge y_2 - lc(x_1,y_1) - lc(x_2,y_2) \\ &\quad + lc(x_1 \vee x_2, y_1 \vee y_2)\end{aligned}$$

Proof. See Peng (1999, proof of Lemma 2.2) and Einmahl (1997, Theorem 3.1). □

Corollary 6.6.2. *Assume* (6.2.1)*. Let* $k \to \infty$ *and* $k/n \to 0$*. Then*

$$\begin{aligned}\sqrt{k}\Big(\frac{n}{k}U_{n,[kx]} - x\Big) &\xrightarrow{D} -W_2(x,0)\\ \sqrt{k}\Big(\frac{n}{k}V_{n,[ky]} - y\Big) &\xrightarrow{D} -W_2(0,y).\end{aligned}$$

Proof. We will prove the first equation. Lemma 6.6.2 implies

$$\sqrt{k}\Big(\frac{1}{k}\sum_{i=1}^{n}\mathbf{1}\{U_i \le \frac{k}{n}x\} - x\Big) \xrightarrow{D} W_2(x,0).$$

Note that the generalised inverse of $x \mapsto 1/k\sum_{i=1}^{n}\mathbf{1}\{U_i \le k/nx\}$ equals $x \mapsto (n/k)U_{n,[kx]}$; applying Vervaat's (1997) lemma gives the result of the corollary. □

Corollary 6.6.3. *Assume the conditions of Theorem 6.2.2. Then*

$$\sqrt{r(n)}\Big(\frac{S_1(U_{n,[kx]},V_{n,[ky]})}{r(n)} - c(x,y)\Big) \xrightarrow{D} W(x,y).$$

$W(x,y)$ is a Gaussian process with mean zero and covariance structure depending on l:

in case $l = 0$

$$W(x,y) = W_1(x,y);$$

in case $l > 0$

$$\begin{aligned} W(x,y) = {} & \frac{1}{\sqrt{l}}\left(W_2(x,0) + W_2(0,y) - W_2(x,y)\right) \\ & - \sqrt{l}c_x(x,y)W_2(x,0) - \sqrt{l}c_y(x,y)W_2(0,y), \end{aligned}$$

where the term in the first line of the right hand side has the same distribution as $W_1(x,y)$.

Proof. For $l = 0$ the result follows from corollaries 6.6.2 and 6.6.1: we have $r(n) = o(k)$ and

$$\begin{aligned} \sqrt{r(n)}\Big(\frac{n}{k}U_{n,[kx]} - x\Big) &\to_p 0 \\ \sqrt{r(n)}\Big(\frac{n}{k}V_{n,[ky]} - y\Big) &\to_p 0. \end{aligned}$$

Otherwise, $r(n)/k \to l$ with $l > 0$. Write

$$\begin{aligned} S_1(U_{n,[kx]},V_{n,[ky]}) &= [kx] + [ky] - S_2(U_{n,[kx]},V_{n,[ky]}) \\ P_1(U_{n,[kx]},V_{n,[ky]}) &= U_{n,[kx]} + V_{n,[ky]} - P_2(U_{n,[kx]},V_{n,[ky]}) \end{aligned}$$

and the result follows from Lemma 6.6.2 and Corollary 6.6.2 (see Peng, 1999). □

Corollary 6.6.4. *Assume the conditions of Theorem 6.2.2. Then*

$$\sqrt{r(n)}\big(\frac{S_1(U_{n,[kx]},V_{n,[kx]})}{S_1(U_{n,k},V_{n,k})} - x^{1/\eta}\big) \xrightarrow{d} W(x,x) - x^{1/\eta}W(1,1) =: V(x).$$

Here $\xrightarrow{d}$ is convergence in $D[0,1]$. The process $V(x)$ in this equation is Gaussian with mean zero and covariance depending on η and l.

For $l = 0$

$$\mathrm{E}\{V(x)V(y)\} = (x \wedge y)^{1/\eta} - (xy)^{1/\eta}.$$

For $l > 0$

$$\begin{aligned}\mathrm{E}\{V(x)V(y)\} =& \big(1 - 2lc_x(1,1)c_y(1,1)\big)(x \wedge y - (1+l)xy) \\ &+ lc_x(1,1)(1 - lc_y(1,1))\big(yc(x,1) + xc(y,1) - c(x \wedge y, x \vee y)\big) \\ &+ lc_y(1,1)(1 - lc_x(1,1))\big(yc(1,x) + xc(1,y) - c(x \vee y, x \wedge y)\big).\end{aligned}$$

Proof. From Corollary 6.6.3 we have

$$\begin{aligned}\frac{S_1(U_{n,[kx]}, V_{n,[kx]})}{S_1(U_{n,k}, V_{n,k})} =& \frac{c(x,x) + r(n)^{-1/2}(W(x,x) + o_p(1))}{c(1,1) + r(n)^{-1/2}(W(1,1) + o_p(1))} \\ =& \frac{c(x,x)}{c(1,1)}\Big[1 + r(n)^{-1/2}\big(\frac{W(x,x)}{c(x,x)} - \frac{W(1,1)}{c(1,1)}\big) \\ &+ r(n)^{-1/2}\big(\frac{o_p(1)}{c(x,x)} - \frac{o_p(1)}{c(1,1)}\big)\Big] \\ =& x^{1/\eta} + r(n)^{-1/2}\frac{W(x,x) - x^{1/\eta}W(1,1)}{c(1,1)} + o_p(r(n)^{-1/2}).\end{aligned}$$

For the proof of the covariance formula in case of $l > 0$, note that then $c_x(1,1) + c_y(1,1) = 1$. □

Remark 6.6.1. For $l = 0$ the process $\{V(x^\eta)\}$ is just a Brownian bridge.

Proof of Theorem 6.2.1. From Corollary 6.6.1 we have

$$\lim_{n\to\infty} \frac{S_1(\frac{k}{n}x, \frac{k}{n}y)}{r(n)} = c(x,y)$$

uniformly on say $0 \le x, y \le 2$. Since

$$\frac{n}{k}U_{n,[kx]} \to x \text{ and } \frac{n}{k}V_{n,[ky]} \to y$$

uniformly on $0 \le x, y \le 2$ by Corollary 6.6.2,

$$\lim_{n\to\infty} \frac{S_1(U_{n,[kx]}, V_{n,[ky]})}{r(n)} = c(x,y)$$

uniformly. Hence $S_n(k,k)/r(n) = S_1(U_{n,k}, V_{n,k})/r(n) \to c(1,1)$ and

$$\frac{\frac{1}{k}\sum_{j=1}^{k} S_n(j,j)}{r(n)} = \frac{\int_0^1 S_1(U_{n,[kx]}, V_{n,[kx]})dx}{r(n)} \to \int_0^1 c(x,x)dx = \frac{1}{1 + 1/\eta}.$$

□

Proof of Theorem 6.2.2 (normality of $\hat{\eta}_4$). By convergence in $D[0,1]$ (Corollary 6.6.4)

$$\sqrt{r(n)}\Big(\int_0^1 \frac{S_1(U_{n,[kx]},V_{n,[kx]})}{S_1(U_{n,k},V_{n,k})}dx - \int_0^1 x^{1/\eta}dx\Big) \xrightarrow{d} \int_0^1 V(x)dx$$

or equivalently

$$\sqrt{r(n)}\Big(1/k\sum_{j=1}^{k}\big(\frac{S_1(U_{n,j},V_{n,j})}{S_1(U_{n,k},V_{n,k})}\big) - \frac{1}{1+1/\eta}\Big) \xrightarrow{d} \int_0^1 V(x)dx. \tag{6.6.2}$$

The distribution of $\int V(x)dx$ is normal with

$$E\left\{\int_0^1 V(x)dx\right\} = \int_0^1 \mathrm{E}\left\{V(x)\right\}dx = 0$$

and variance

$$\mathrm{E}\left\{\int_0^1 V(x)dx\int_0^1 V(y)dy\right\} = 2\int_0^1\int_0^y \mathrm{E}\left\{V(x)V(y)\right\}dxdy.$$

Using Corollary 6.6.4, this variance equals

$$\begin{aligned}&2\int_0^1\int_0^y x^{1/\eta}(1-y^{1/\eta})dxdy\\ &\quad= \frac{1/\eta}{(2+1/\eta)(1/\eta+1)^2}, \qquad \text{for } l=0,\end{aligned}$$

or

$$\begin{aligned}&\frac{1}{3}lc_x(1,1)\big(1-lc_y(1,1)\big)\int_0^1 c(u,1)du\\ &\quad+\frac{1}{3}lc_y(1,1)\big(1-lc_x(1,1)\big)\int_0^1 c(1,u)du\\ &\quad+\frac{1}{2}l^2c_x(1,1)c_y(1,1)-\frac{1}{3}lc_x(1,1)-\frac{1}{3}lc_y(1,1)+\frac{1}{12}\\ &\quad+\frac{1}{12}l(c_x(1,1)^2+c_y(1,1)^2), \qquad \text{for } l>0.\end{aligned}$$

Finally

$$\hat{\eta}_4-\eta=(1+\eta)^2\big(\frac{1}{1+1/\hat{\eta}_4}-\frac{1}{1+1/\eta}\big)(1+o(1)).$$

This proves Theorem 6.2.2 for $\hat{\eta}_4$. □

Proof of Corollary 6.2.1 (for $\hat{\eta}_4$). By Corollary 6.6.3 we have

$$r(n)^{-1} S_1(U_{n,k}, V_{n,k}) \to_p 1;$$

So $S_n(k,k) = S_1(U_{n,k}, V_{n,k})$ a.s. is a consistent estimator of $r(n)$ in the theorem. This proves Corollary 6.2.1. □

Remark 6.6.2. It is worth mentioning that the asymptotic normality of $\hat{\eta}_3$ can be derived from Corollary 6.6.3 in a similar way.

Now we turn to the Ledford and Tawn - type estimators $\hat{\eta}_1$ and $\hat{\eta}_2$.

Let $m_n = [r(n)]$ and denote by Q_n the tail empirical quantile function pertaining to $T_i^{(n)}$, $1 \le i \le n$, i.e.

$$Q_n(t) := T^{(n)}_{n,n-[m_n t]}, \qquad 0 < t < n/m_n.$$

The following lemma is central to the proof of the asymptotic normality of estimators for η based on largest order statistics of $T_i^{(n)}$.

Lemma 6.6.3. *Under the conditions of Theorem 2.2 there exist suitable versions of Q_n, a suitable process $\bar{W}$ equal in distribution to a standard Brownian motion if $l = 0$ and to $x \mapsto W(x,x)$ if $l > 0$ such that for all $t_0, \varepsilon > 0$*

$$\sup_{0<t\le t_0} t^{\eta+1/2+\varepsilon} \left| m_n^{1/2}\Big(\frac{k}{n} Q_n(t) - t^{-\eta}\Big) - \eta t^{-(\eta+1)} \bar{W}(t) \right| = o_P(1).$$

Proof. First check that

$$\begin{aligned}\sum_{i=1}^n 1\{T_i^{(n)} > x\} &= \sum_{i=1}^n 1\{R_i^X > (n+1)(1-1/x) \text{ and } R_i^Y > (n+1)(1-1/x)\} \\ &= \sum_{i=1}^n 1\{U_i < U_{n,\lceil (n+1)/x \rceil} \text{ and } V_i < V_{n,\lceil (n+1)/x\rceil}\} \qquad \text{a.s.}\end{aligned}$$

with the convention $U_{n,n+1} = V_{n,n+1} = 1$. Hence

$$\bar{F}_n(x) := \frac{1}{n}\sum_{i=1}^n 1\{\tfrac{k}{n+1} T_i^{(n)} > x\} = \frac{1}{n} S_1\Big(U_{n,\lceil k/x\rceil}-, V_{n,\lceil k/x \rceil}-\Big)$$

where $f(x-)$ denotes the left-hand limit of f at x. From Corollary 6.6.1 one readily obtains that

$$\begin{aligned} & m_n^{1/2}\Big(\frac{\bar{F}_n(x)}{q(k/n)} - x^{-1/\eta}\Big)_{0<x<\infty} \longrightarrow \Big(W(1/x,1/x)\Big)_{0<x<\infty} \\ \Longrightarrow\quad & m_n^{1/2}\Big(\frac{\bar{F}_n(x^{-\eta})}{q(k/n)} - x\Big)_{0<x<\infty} \longrightarrow \Big(W(x^\eta, x^\eta)\Big)_{0<x<\infty} =: \bar{W} \\ \Longrightarrow\quad & m_n^{1/2}\Big(\Big(\bar{F}_n^{-1}(q(k/n)t)\Big)^{-1/\eta} - t\Big)_{0<t<\infty} \longrightarrow -\bar{W}\end{aligned}$$

weakly in $D(0,\infty)$, where in the last step Vervaat's lemma has been used. For this, note that $\bar W$ has a.s. continuous sample paths, because by the definition of W it is a Brownian motion for $l=0$ and it can be represented as a sum of Brownian motions if $l>0$.

Consequently for suitable versions

$$\Big(\bar F_n^{-1}(q(k/n)t)\Big)^{-1/\eta} = t - m_n^{-1/2}\bar W(t) + o(m_n^{-1/2})$$

a.s. uniformly on compact intervals bounded away from 0. The δ–method yields

$$F_n^{-1}(q(k/n)t) = t^{-\eta}\Big(1 + m_n^{-1/2}\eta t^{-1}\bar W(t) + o(m_n^{-1/2})\Big)$$

uniformly in the same sense. Check that $F_n^{-1}(q(k/n)t) = k/(n+1)Q_n(r(n)t/m_n) = k/nQ_n(t) + O(1/m_n)$ uniformly and $\sup_{0<t\le\vartheta} t^{-1/2+\varepsilon}|\bar W(t)| = o_P(1)$ as $\vartheta\downarrow 0$ by the law of the iterated logarithm and the aforementioned representation of $\bar W$. Thus it remains to prove that for all $\delta>0$

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Big\{\sup_{0<t\le\vartheta} m_n^{1/2}t^{\eta+1/2+\varepsilon}\Big|\frac{k}{n+1}Q_n(t) - t^{-\eta}\Big| > \delta\Big\} = 0. \qquad (6.6.3)$$

For this, we restrict ourselves to considering

$$\begin{aligned} &P\Big\{\sup_{0<t\le\vartheta} m_n^{1/2}t^{\eta+1/2+\varepsilon}\Big(\frac{k}{n+1}Q_n(t) - t^{-\eta}\Big) > \delta\Big\}\\ \le\ & P\Big\{\exists 1\le i\le m_n\vartheta+1 : \frac{k}{n+1}T^{(n)}_{n,n-i+1} > x_{i,n}\Big\}\\ =\ & P\Big\{\exists 1\le i\le m_n\vartheta+1 : \frac{k}{n+1}T^{(n)}_{n,n-i+1} > x_{i,n} \text{ and } x_{i,n} < k\Big\} \qquad (6.6.4)\end{aligned}$$

with

$$x_{i,n} := \Big(\frac{i}{m_n}\Big)^{-\eta} + \delta m_n^{-1/2}\Big(\frac{i}{m_n}\Big)^{-(\eta+1/2+\varepsilon)}.$$

(The other inequality can be treated in a similar way.)

Let $A_i := 1/U_i$, $B_i := 1/V_i$ and

$$\tilde S_1(x,y) := \sum_{i=1}^n 1_{\{A_i>x \text{ and } B_i>y\}} = S_1(1/x-, 1/y-).$$

Then the right-hand side of (6.6.4) equals

$$P\Big\{\exists 1\le i\le m_n\vartheta+1 : \tilde S_1(A_{n,n-\lceil k/x_{i,n}\rceil+1}, B_{n,n-\lceil k/x_{i,n}\rceil+1}) \ge i \text{ and } x_{i,n} < k\Big\}.$$

Now we distinguish two different ranges of i-values.

Case 1: $i <= i_n := [(\delta m_n^{\varepsilon}/L)^{1/(1/2+\varepsilon)}]$, $x_{i,n} < k$

According to Shorack and Wellner (1986, Theorem 10.3.1), for all $\bar{\varepsilon} > 0$ there exists $\bar{\delta} > 0$ such that

$$\limsup_{n\to\infty} P\{\exists\, 2 \le j \le m_n + 1 : \frac{j-1}{n} A_{n,n-j+1} < \bar{\delta}\} \le \bar{\varepsilon},$$

and likewise for $B_{n,n-i+1}$. Thus

$$\begin{aligned} & P\Big\{\exists\, 1 \le i \le i_n : \tilde{S}_1(A_{n,n-\lceil k/x_{i,n}\rceil+1}, B_{n,n-\lceil k/x_{i,n}\rceil+1}) \ge i \text{ and } x_{i,n} < k\Big\} \\ \le\ & P\Big\{\exists\, 1 \le i \le i_n : \tilde{S}_1(x_{i,n}\bar{\delta}n/k, x_{i,n}\bar{\delta}n/k) \ge i\Big\} + 2\bar{\varepsilon}. \end{aligned}$$

Check that

$$\frac{n}{k} x_{i,n}\bar{\delta} \ \ge\ \delta\frac{n}{k} m_n^{-1/2}\Big(\frac{i}{m_n}\Big)^{-(\eta+1/2+\varepsilon)}\bar{\delta} \ \ge\ \bar{\delta}Lk^{-1}m_n^{\eta}n^{1-\eta}(n/i)^{\eta}. \tag{6.6.5}$$

Denote by F_T the d.f. of $T_i := \min(A_i, B_i)$, i.e. $1 - F_T(x) = P_1(1/x, 1/x)$, so that F_T^{-1} is $(-\eta)$–varying at 1.

In case of $\eta < 1$, we have $k = o(m_n^{\eta+\iota})$ and $F_T^{-1}(1-t) = o(t^{-(\eta+\iota)})$ as $t \downarrow 0$ for all $\iota > 0$, so that the right-hand side of (6.6.5) is of larger order than $F_T^{-1}(1-2i/(\bar{\delta}Ln))$, provided $\iota < (1-\eta)/2$.

If $\eta = 1$, then one can show that, in analogy to Lemma 2.1 of Drees (1998a),

$$\sup_{x\le 1} x^{\iota-1}\Big|\frac{P_1(tx,tx)}{P_1(t,t)} - x\Big| = o(q_1(t)).$$

Apply this bound with $t = k/n$ and $x = i/(\bar{\delta}Lm_n)$ to obtain $1 - F_T(x_{i,n}\bar{\delta}n/k) \le 2i/(\bar{\delta}Ln)$, since $P_1(k/n, k/n) \sim q(k/n) \sim m_n/n$ and $(i/m_n)^{1-\iota}q_1(k/n)$ $= o(m_n^{1/2}q_1(k/n)i/m_n) = o(i/m_n)$ uniformly for $1 \le i \le i_n$.

Hence it follows that

$$\begin{aligned} & \limsup_{n\to\infty} P\Big\{\exists\, 1 \le i \le i_n : \tilde{S}_1(x_{i,n}\bar{\delta}n/k, x_{i,n}\bar{\delta}n/k) \ge i\Big\} \\ \le\ & \limsup_{n\to\infty} P\Big\{\exists\, 1 \le i \le i_n : T_{n,n-i+1} > \frac{n}{k}x_{i,n}\bar{\delta}\Big\} \\ \le\ & \limsup_{n\to\infty} P\Big\{\max_{1\le i\le m_n+1} \frac{T_{n,n-i+1}}{F_T^{-1}(1-2i/(\bar{\delta}Ln))} > 1\} \\ <\ & \bar{\varepsilon} \end{aligned}$$

for sufficiently large L, where for the last step again Theorem 10.3.1 of Shorack and Wellner (1986) has been used.

Case 2: $i_n < i \le m_n\vartheta + 1$

In this case we use the convergence

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Big\{\sup_{0<t\le\vartheta} k^{1/2}t^{3/2+\iota}\Big|\frac{k}{n}A_{n,n-\lceil kt\rceil+1} - t^{-1}\Big| > \tilde{\delta}\Big\} = 0$$

for all $\tilde{\delta}, \iota > 0$, which is immediate from Theorem 2.1 of Drees (1998a). By arguments similar to the ones given above, it suffices to consider $P\{\exists\, i_n < i \le m_n\vartheta + 1 : \tilde{S}_1(y_{i,n}, y_{i,n}) \ge i\}$ with

$$\begin{aligned} y_{i,n} &:= \frac{n}{k} x_{i,n} - \tilde{\delta} n k^{-3/2} x_{i,n}^{3/2+\iota} \\ &= \frac{n}{k}\Big(\frac{i}{m_n}\Big)^{-\eta}\Big[1 + \delta m_n^{-1/2}\Big(\frac{i}{m_n}\Big)^{-(1/2+\varepsilon)} - \tilde{\delta} k^{-1/2}\Big(\frac{i}{m_n}\Big)^{-\eta(1/2+\iota)} \times \\ &\qquad \times \Big(1 + \delta m_n^{-1/2}\Big(\frac{i}{m_n}\Big)^{-(1/2+\varepsilon)}\Big)^{3/2+\iota}\Big] \\ &\ge \frac{n}{k}\Big(\frac{i}{m_n}\Big)^{-\eta}\Big[1 + \frac{\delta}{2} m_n^{-1/2}\Big(\frac{i}{m_n}\Big)^{-(1/2+\varepsilon)}\Big] \end{aligned}$$

for $\iota < \varepsilon$ and $\tilde{\delta} \le \delta(1+L)^{-(3/2+\iota)}/2$, since $k \ge m_n$ and $\eta \le 1$. Therefore

$$\begin{aligned} &\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\{\exists\, i_n < i \le m_n\vartheta + 1 : \tilde{S}_1(y_{i,n}, y_{i,n}) \ge i\} \\ &\le \lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\Big\{\exists\, i_n < i \le m_n\vartheta + 1 : \\ &\qquad m_n^{1/2}\Big(\frac{i}{m_n}\Big)^{\eta+1/2+\varepsilon}\Big(\frac{k}{n}T_{n,n-i+1} - \Big(\frac{i}{m_n}\Big)^{-\eta}\Big) > \delta/2\Big\} \\ &= 0, \end{aligned}$$

again by Theorem 2.1 of Drees (1998a), where (2.1) implies Condition 1 of that paper and $m_n^{1/2} q_1(k/n) \to 0$ ensures that the bias is asymptotically negligible.

Combining both cases one arrives at (6.6.3). □

Proof of Theorem 6.2.2 (asymptotic normality of $\hat{\eta}_1$ and $\hat{\eta}_2$). Note that this approximation is analogous to the approximation of the tail empirical quantile function established in Drees (1998a) in the classical situation of i.i.d. random variables. Hence the asymptotic normality of $\hat{\eta}_1$ and $\hat{\eta}_2$ follows from Lemma 6.6.3 exactly as in Drees (1998a, Example 4.1) and Drees (1998b, Example 3.1) using the δ-method. The asymptotic variance is given by

$$\int_0^1\int_0^1 \operatorname{Cov}(\bar{W}(s), \bar{W}(t))(st)^{-(\eta+1)}\,\nu_\eta(ds)\,\nu_\eta(dt)$$

with $\nu_\eta(dt) := (\eta+1)^2(t^\eta - (2\eta+1)t^{2\eta})/\eta\,dt + (\eta+1)\varepsilon_1(dt)$ for the maximum likelihood estimator $\hat{\eta}_1$ and $\nu_\eta(dt) := \eta(t^\eta dt - \varepsilon_1(dt))$ in case of the Hill estimator. (Here ε_1 denotes the Dirac measure at 1.) Now using the homogeneity of order 1 of the covariance function which implies $\int_0^t \operatorname{Cov}(\bar{W}(s), \bar{W}(t))(st)^{-1}\,ds = \int_0^1 \operatorname{Cov}(\bar{W}(u), \bar{W}(1))u^{-1}\,du$, one obtains $(\eta+1)^2\operatorname{Var}(\bar{W}(1))$ and $\eta^2\operatorname{Var}(\bar{W}(1))$, respectively, as asymptotic variance and thus the assertion, using $c_x(1,1) + c_y(1,1) = 1/\eta$. □

Proof of Theorem 6.2.3. Note that according to Corollary 6.6.3

$$\frac{S_n(i,j)}{r(n)} = c\Big(\frac{i}{k}, \frac{j}{k}\Big) + O_P\Big((r(n))^{-1/2}\Big)$$

uniformly for $1 \le i, j \le 2k$. Hence $\hat{l} \to l$ in probability by the definition of $r(n)$. Moreover,

$$\begin{aligned}
\hat{c}_x(1,1) &= k^{1/4}\frac{c([k(1+k^{-1/4})]/k,1) - c(1,1) + O_P((r(n)^{-1/2}))}{1+O_P((r(n)^{-1/2}))} \\
&= c_x(1,1) + O_P\Big(k^{1/4}(r(n))^{-1/2}\Big) \\
&\overset{P}{\to} c_x(1,1)
\end{aligned}$$

if $\eta > 1/2$. The consistency of $\hat{c}_y(1,1), \hat{d}_1$ and $\hat{d}_2$ can be proved in a similar way, so that the consistency of $\hat{\sigma}_i^2$ follows readily in that case.

In case of $\eta \le 1/2$, we have

$$\hat{l}^{1/2}\hat{c}_x(1,1) = O_P\Big((r(n)/k)^{1/2}(1 + k^{1/4}(r(n))^{-1/2})\Big) = o_P(1)$$

and likewise $\hat{l}^{1/2}(\hat{c}_y(1,1) + \hat{d}_1 + \hat{d}_2) \to 0$ in probability. Thus the consistency of $\hat{\sigma}_i$ is obvious because of $l = 0$.

Assertion (ii) follows similarly from

$$\frac{k}{n}T^{(n,u)}_{n,n-[m_n t]} = \left(\frac{t}{c(1+u,1)}\right)^{-\eta} + O_P(m_n^{-1/2})$$

which in turn can be verified using the same arguments as in the proof of Lemma 6.6.3. □

6.7 Proof of Theorem 6.3.1

The proof of Theorem 6.3.1 will be established in several steps. The following sequence of equalities and asymptotic in probability equivalences provides an overview over the reasoning:

$$\begin{aligned}
p_n &= P\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(C_n)\}\\
&\overset{(6.3.13)}{\sim} q(\frac{k}{n})\nu(\frac{n}{k}(\mathbf{1}-\boldsymbol{F}(C_n)))\\
&\overset{\text{Lemma }6.7.2}{\sim} q(\frac{k}{n})\nu(D_n)\\
&\overset{(6.2.2)}{=} c_n^{1/\eta}q(\frac{k}{n})\nu(\frac{D_n}{c_n})\\
&\overset{\text{cor. }6.7.3}{\sim} c_n^{1/\eta}q(\frac{k}{n})\nu\Big(\mathbf{1}-\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(\boldsymbol{F}^{-\mathbf{1}}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1}-\frac{\hat{D}_n}{c_n}))\Big)\\
&\overset{\text{Lemma }6.7.3}{\sim} c_n^{1/\eta}q(\frac{k}{n})\nu\Big(\frac{n}{k}(\mathbf{1}-\boldsymbol{F}(\boldsymbol{F}^{-\mathbf{1}}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1}-\frac{\hat{D}_n}{c_n}))\Big)\\
&\overset{(6.3.13)}{\sim} c_n^{1/\eta}P\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}_{|B=\boldsymbol{F}^{-\mathbf{1}}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1}-\frac{\hat{D}_n}{c_n})}\\
&\overset{\text{Lemma }6.7.4}{\sim} c_n^{1/\eta}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\Big\{(X_i,Y_i)\in\boldsymbol{F}^{-\mathbf{1}}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1}-\frac{\hat{D}_n}{c_n})\Big\}\\
&\sim \hat{p}_n. \qquad (6.7.1)
\end{aligned}$$

Lemma 6.7.1. *Let $a=a(n),\tilde{a}>0$, $b,\tilde{b},\gamma,\tilde{\gamma}\in\mathbb{R}$ denote sequences such that*

$$\big|\frac{\tilde{a}}{a}-1\big|\vee\big|\frac{\tilde{b}-b}{a}\big|\vee|\tilde{\gamma}-\gamma|=O(\varepsilon_n)$$

for some $\varepsilon_n\downarrow 0$. Suppose that the sequence $\lambda_n>0$ is bounded and satisfies $\varepsilon_n\log\lambda_n\to 0$ and $\varepsilon_n w_\gamma(\lambda_n)\to 0$ with w_γ defined in (6.3.11)*. Then*

$$1-F_{\tilde{a},\tilde{b},\tilde{\gamma}}(F^{-1}_{a,b,\gamma}(1-x))=x+o(\lambda_n) \qquad (6.7.2)$$

uniformly for $0\le x\le\lambda_n$.

Proof. First check that

$$T(x):=1-F_{\tilde{a},\tilde{b},\tilde{\gamma}}(F^{-1}_{a,b,\gamma}(1-x))=\Big[1+\tilde{\gamma}\frac{a}{\tilde{a}}\Big(\frac{x^{-\gamma}-1}{\gamma}+\frac{b-\tilde{b}}{a}\Big)\Big]^{-1/\tilde{\gamma}},$$

where, as usual, $(x^{-\gamma}-1)/\gamma:=-\log x$ if $\gamma=0$. Now we distinguish three cases.

$\boldsymbol{\gamma>0}$ **:** Then

$$\begin{aligned}
T(x) &= \big(1+(1+O(\varepsilon_n))(x^{-\gamma}-1+O(\varepsilon_n))\big)^{-(1+O(\varepsilon_n))/\gamma}\\
&= \big(x^{-\gamma}(1+O(\varepsilon_n))+O(\varepsilon_n)\big)^{-(1+O(\varepsilon_n))/\gamma}\\
&= x\exp(O(\varepsilon_n)\log x)(1+o(1)).
\end{aligned}$$

For $\lambda_n\varepsilon_n\le x\le\lambda_n$

$$|\log x|\varepsilon_n\le\big(|\log\lambda_n|+|\log\varepsilon_n|\big)\varepsilon_n\ \to\ 0,$$

so that $T(x) = x(1+o(1)) = x + o(\lambda_n)$ uniformly.

Otherwise, i.e. for $0 < x < \lambda_n \varepsilon_n$,

$$T(x) \le T(\lambda_n \varepsilon_n) = \lambda_n \varepsilon_n (1+o(1)) = o(\lambda_n) = x + o(\lambda_n)$$

by the monotonicity of T.

$\boldsymbol{\gamma < 0}$ **:** Choose $\delta_n \to 0$ such that $\varepsilon_n (\lambda_n \delta_n)^\gamma \to 0$ and hence also $\varepsilon_n \log \delta_n \to 0$ (e.g. $\delta_n = (\varepsilon_n \lambda_n^\gamma)^{-1/(2\gamma)}$). Then uniformly for $\lambda_n \delta_n \le x \le \lambda_n$

$$T(x) = x^{1+O(\varepsilon_n)} \big(1 + O(\varepsilon_n) + O(\varepsilon_n (\lambda_n \delta_n)^\gamma)\big)^{-(1+O(\varepsilon_n))/\gamma} = x(1+o(1))$$

and again (6.7.2) follows from the monotonicity of T.

$\boldsymbol{\gamma = 0}$ **:** Note that $\tilde{\gamma} |\log x| \to 0$ uniformly for $\lambda_n \varepsilon_n \le x \le \lambda_n$. Hence a Taylor expansion of log yields

$$\begin{aligned} T(x) &= \exp\Big(-\frac{1}{\tilde{\gamma}} \log\big(1 + \tilde{\gamma}(1+O(\varepsilon_n))(-\log x + O(\varepsilon_n))\big)\Big) \\ &= \exp\Big(-\frac{1}{\tilde{\gamma}} \big[\tilde{\gamma}(1+O(\varepsilon_n)(-\log x + O(\varepsilon_n)) + O(\tilde{\gamma}^2 (\log x + O(\varepsilon_n))^2)\big]\Big) \\ &= x \exp\big(O(\varepsilon_n) \log x + O(\varepsilon_n) + O(\varepsilon_n \log^2 x)\big) \\ &= x(1+o(1)) \end{aligned}$$

and thus the assertion by the aforementioned arguments. □

Remark 6.7.1. For fixed sequences a, b and γ, assertion (6.7.2) even holds true uniformly for

$$(\tilde{a}, \tilde{b}, \tilde{\gamma}) \in M(\varepsilon_n) := \Big\{(\bar{a}, \bar{b}, \bar{\gamma}) \in (0,\infty) \times \mathbb{R}^2 \;\Big|\; \Big|\frac{\bar{a}}{a} - 1\Big| \vee \Big|\frac{\bar{b} - b}{a}\Big| \vee |\bar{\gamma} - \gamma| \le \varepsilon_n\Big\}. \tag{6.7.3}$$

Corollary 6.7.1. *If condition (D),* (6.3.7) *and* (6.3.10)–(6.3.11) *are satisfied then, for all* $\delta > 0$,

$$P\Big\{A_{-\delta} \subset \frac{\hat{D}_n}{d_n} \subset A_{+\delta}\Big\} \;\to\; 1.$$

Proof. Since the set A is bounded, there exists $L > 0$ such that $D_n \subset [0, d_n L]^2$ for all sufficiently large n. Because of (6.3.11), one can find a sequence $\varepsilon_n \to 0$ such that $k^{-1/2} = o(\varepsilon_n)$ and the conditions of Lemma 6.7.1 hold for $\lambda_n = d_n L$. Then $P\{(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{\gamma}}) \in (M(\varepsilon_n))^2\} \to 1$ with $M(\varepsilon_n)$ defined in (6.7.3) and Lemma 6.7.1 yields

$$\sup_{(x,y)\in D_n} \|\mathbf{1} - \boldsymbol{F}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\boldsymbol{F}^{-1}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(\mathbf{1} - (x,y))) - (x,y)\| \le \frac{\delta}{2} d_n \tag{6.7.4}$$

with probability tending to 1. Thus, in view of $\hat{D}_n = \mathbf{1} - \boldsymbol{F}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\boldsymbol{F}^{-1}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(\mathbf{1} - D_n))$ and condition (D),

$$P\Big\{\frac{\hat{D}_n}{d_n} \subset \Big(\frac{D_n}{d_n}\Big)_{+\delta/2} \subset A_{+\delta}\Big\} \to 1.$$

On the other hand, by the definition of the inner neighborhood of a set, $(x,y) \in (D_n/d_n)_{-\delta/2}$ implies $(x+\delta/2, y+\delta/2) \in D_n/d_n$. Since, in view of (6.7.4),

$$d_n(x,y) \le \mathbf{1} - \boldsymbol{F}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\Big(\boldsymbol{F}^{-1}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\Big(\mathbf{1} - d_n\big(x+\frac{\delta}{2}, y+\frac{\delta}{2}\big)\Big)\Big)$$

componentwise, (6.3.7) shows that indeed $d_n(x,y) \in \hat{D}_n$. Hence, again by condition (D),

$$P\{A_{-\delta} \subset \big(\frac{D_n}{d_n}\big)_{-\delta/2} \subset \frac{\hat{D}_n}{d_n}\} \to 1.$$

□

Corollary 6.7.2. *If the conditions of Corollary 6.7.1 hold and, in addition,* (6.3.12) *then, for all* $\delta > 0$,

$$P\Big\{A_{-\delta} \subset \frac{c_n}{d_n}\big(\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\big)\big)\big) \subset A_{+\delta}\Big\} \to 1.$$

Proof. According to Corollary 6.7.1, there exists $L > 0$ such that $P\{\hat{D}_n/c_n \subset [0,\lambda_n]^2\} \to 1$ for $\lambda_n := Ld_n/c_n$. It follows from (6.3.10) and (6.3.12) that $\lambda_n^{\hat{\gamma}_i} = \lambda_n^{\gamma_i}(1+o_P(1))$, $i = 1,2$. Hence one may apply Lemma 6.7.1 with $(a,b,\gamma) = (\hat{a}_i, \hat{b}_i, \hat{\gamma}_i)$ and $(\tilde{a}, \tilde{b}, \tilde{\gamma}) = (a_i, b_i, \gamma_i)$ to obtain

$$\sup_{(x,y)\in\hat{D}_n/c_n} \big\|\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1} - (x,y))\big) - (x,y)\big\| \le \frac{\delta}{2}\frac{d_n}{c_n}$$

with probability tending to 1 for all $\delta > 0$. Now one may conclude the proof following the lines of the preceding proof. □

Corollary 6.7.3. *Under the conditions of Corollary 6.7.2*

$$\nu\Big(\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\big)\big)\Big) = \nu\big(\frac{D_n}{c_n}\big)(1+o_P(1)).$$

Proof. Denote the boundary of the set A by ∂A. Condition (6.3.7) implies a slightly weaker version for A, namely $(x,y) \in A \Rightarrow [0,x) \times [0,y) \subset A$. Hence $\lambda \cdot \partial A \subset \partial A$ for all $\lambda \in (0,1)$ and these sets are pairwise disjoint. Since ν is homogeneous in the sense of (6.2.2) and $\nu(A) < \infty$ by the boundedness of A, it follows that $\nu(\partial A) = 0$. Moreover, $A_{+\delta} \setminus A_{-\delta} \downarrow \partial A$ as $\delta \downarrow 0$, so that $\nu(A_{+\delta} \setminus A_{-\delta}) \to 0$. Thus Corollary 6.7.2 and condition (D) yield

$$\nu\Big(\frac{c_n}{d_n}\big(\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\big)\big)\big)\Big) \to \nu(A)$$

and $\nu(D_n/d_n) \to \nu(A)$. Now the assertion is an obvious consequence of the homogeneity (6.2.2). □

Lemma 6.7.2. *If condition (D),* (6.3.7) *and* (6.3.8) *hold, then*

$$\nu(D_n) = \nu\big(\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(C_n))\big)(1 + o(1)).$$

Proof. There exists $L > 0$ such that $D_n \subset [0, d_n L]^2$ for all sufficiently large n. Choose arbitrary $-1/(\gamma_i \vee 0) < x_i < 1/((-\gamma_i) \vee 0)$, $i = 1, 2$. Then, by (6.3.8), for all $(x, y) \in D_n$

$$\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(\boldsymbol{F}^{-1}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(\mathbf{1} - (x, y)))) = (x(1 + \delta_x), y(1 + \delta_y)) \tag{6.7.5}$$

with $|\delta_x| \vee |\delta_y| \le R_{x_1,x_2}(n/k)$ for sufficiently large n. According to (6.3.7), the left-hand side of (6.7.5) is an element of $D_n(1 + R_{x_1,x_2}(n/k))$. Thus, by the definition of D_n,

$$\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(C_n)) \subset D_n\big(1 + R_{x_1,x_2}\big(\frac{n}{k}\big)\big).$$

Likewise, (6.7.5) together with (6.3.7) implies

$$D_n\big(1 - R_{x_1,x_2}\big(\frac{n}{k}\big)\big) \subset \frac{n}{k}(\mathbf{1} - \boldsymbol{F}(C_n))$$

eventually. Now the assertion is obvious from the homogeneity property (6.2.2). □

Lemma 6.7.3. *Under the conditions (D),* (6.3.7), (6.3.8) *and* (6.3.10)–(6.3.12) *one has*

$$\nu\Big(\big(\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\big)\big)\big)\Big) = \nu\Big(\frac{n}{k}\big(\mathbf{1} - \boldsymbol{F}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\big)\big)\big)\Big)(1 + o_P(1)).$$

Proof. The proof is very much the same as that for Lemma 6.7.2 with D_n replaced by $\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1} - \hat{D}_n/c_n))$. For this note that, by the boundedness of d_n/c_n and the assertion of Corollary 6.7.2, this set is eventually bounded. Hence (6.3.8) is applicable for sufficiently small x_1 and x_2. □

Lemma 6.7.4. *If the conditions of Theorem 6.3.1 are satisfied, then*

$$\sup_{B \in \mathcal{B}_n} \left| \frac{\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\mathbf{1} - \boldsymbol{F}(X_i, Y_i) \in \mathbf{1} - \boldsymbol{F}(B)\}}{P\{\mathbf{1} - \boldsymbol{F}(X, Y) \in \mathbf{1} - \boldsymbol{F}(B)\}} - 1 \right| \to 0 \quad \textit{in probability.}$$

Proof. We will apply Theorem 5.1 of Alexander (1987). To check the conditions of this uniform law of large numbers, first note that every set $B \in \mathcal{B}_n$ can be represented as

$$B = \boldsymbol{F}^{-1}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\mathbf{1} - \boldsymbol{F}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}}}(C_n)}{c_n}\big) \tag{6.7.6}$$

with $(\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{\gamma}}) \in (M(\varepsilon_n))^2$ (cf. (6.7.3)). Therefore the arguments of the proofs for Lemma 6.7.3 and Corollary 6.7.3 show that

$$\begin{aligned}\nu\big(\frac{n}{k}(\mathbf{1}-\boldsymbol{F}(B))\big) &= \nu\big(\mathbf{1}-\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(B))(1+o(1)) = \nu\big(\frac{D_n}{c_n}\big)(1+o(1)) \\ &= \big(\frac{d_n}{c_n}\big)^{1/\eta}\nu(A)(1+o(1)) \qquad (6.7.7)\end{aligned}$$

uniformly for $B \in \mathcal{B}_n$ (cf. Remark 6.7.1). Now (6.3.13) leads to

$$P\{\mathbf{1}-\boldsymbol{F}(X,Y) \in \mathbf{1}-\boldsymbol{F}(B)\} = q\big(\frac{k}{n}\big)\big(\frac{d_n}{c_n}\big)^{1/\eta}\nu(A)(1+o(1)) \qquad (6.7.8)$$

uniformly. In particular, there exists n_0 such that $P\{\mathbf{1}-\boldsymbol{F}(X,Y) \in \mathbf{1}-\boldsymbol{F}(B)\} < 1/2$ for all $n \geq n_0$ and all $B \in \mathcal{B}_n$.

Next note that

$$\begin{aligned}\bar{B}_t &:= \bigcup_{B\in\mathcal{B}_n, n\geq n_0, P\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}(1-P\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\})\leq t} B \\ &\subset \bigcup_{B\in\mathcal{B}_n, n\geq n_0, P\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}\leq 2t} B. \qquad (6.7.9)\end{aligned}$$

In view of (6.7.6), one may prove as in Corollary 6.7.2 that, for all $\delta > 0$, eventually $\mathbf{1}-\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(B) \subset A_{+\delta}d_n/c_n$ for all $B \in \mathcal{B}_n$. Hence it follows as in the proof of Lemma 6.7.2 that

$$\frac{n}{k}(\mathbf{1}-\boldsymbol{F}(B)) \subset \frac{d_n}{c_n}A_{+\delta}(1+o(1)) \qquad (6.7.10)$$

uniformly for $B \in \mathcal{B}_n$.

Let $n(t) := \min\{n \geq n_0 \mid q(k/n)(d_n/c_n)^{1/\eta}\nu(A) \leq 3t\}$, which tends to ∞ as t tends to 0. Combining (6.7.8)–(6.7.10), we arrive at

$$\begin{aligned}\mathbf{1}-\boldsymbol{F}(\bar{B}_t) &\subset \bigcup_{n\geq n(t)} \frac{k(n)d_n}{nc_n}A_{+\delta}(1+o(1)) \\ &\subset 2\sup_{n\geq n(t)} \frac{k(n)d_n}{nc_n}A_{+\delta}\end{aligned}$$

for sufficiently small t. By (6.3.13), the regularity condition on $k(n)$ and the definition of $n(t)$ it follows that

$$\begin{aligned}P\{\mathbf{1}-\boldsymbol{F}(X,Y) \in \mathbf{1}-\boldsymbol{F}(\bar{B}_t)\} &= O\Big(q\big(\frac{k(n(t))}{n(t)}\big)\big(\frac{n(t)}{k(n(t))}\sup_{n\geq n(t)}\frac{k(n)d_n}{nc_n}\big)^{1/\eta}\Big) \\ &= O\Big(q\big(\frac{k(n(t))}{n(t)}\big)\big(\frac{d_n}{c_n}\big)^{1/\eta}\Big) \\ &= O(t).\end{aligned}$$

Since $\mathcal{B}_n$ is a VC class, Theorem 5.1 of Alexander (1987) yields

$$\sup\left\{\left|\frac{\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\mathbf{1}-\boldsymbol{F}(X_i,Y_i)\in\mathbf{1}-\boldsymbol{F}(B)\}}{P\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}}-1\right|\;\middle|\; B\in\mathcal{B}_n, P\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}\geq\varepsilon_n\right\}\to 0,$$

provided $n\varepsilon_n\to\infty$. Because of (6.7.8) and the last assumption of (6.3.12), the choice $\varepsilon_n = q(k/n)(d_n/c_n)^{1/\eta}\nu(A)/2$ leads to the assertion. □

Proof of Theorem 6.3.1. Now the consistency of $\hat{p}_n$ can be proven as shown in (6.7.1). For this note that, because of (6.3.10), $\boldsymbol{F}^{-\mathbf{1}}_{\hat{\mathbf{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1}-\hat{D}_n/c_n)$ belongs to $\mathcal{B}_n$ with probability tending to 1 and that $\log c_n = o((r(n))^{1/2})$ implies $c_n^{1/\hat{\eta}} = c_n^{1/\eta}(1+o_P(1))$ since $\hat{\eta}$ was assumed $\sqrt{r(n)}$–consistent for η. □

Appendix. Some analytical results

Write $Q(x,y) = P\{1-F_1(X) < x \text{ and } 1-F_2(Y) < y\}$. As in (6.2.1) suppose

$$\lim_{t\downarrow 0}\frac{\frac{Q(tx,ty)}{q(t)}-c(x,y)}{q_1(t)} =: \tilde{c}_1(x,y) \tag{6.A.1}$$

exists for $x,y\geq 0$ (but $x+y>0$) with q positive, $q_1(t)\to 0, (t\downarrow 0)$, $\tilde{c}_1$ non constant and not a multiple of c and w.l.o.g. $c(1,1)=1$. Moreover assume that (6.A.1) holds uniformly on

$$\{(x,y)\mid x^2+y^2=1, x\geq 0, y\geq 0\}.$$

It is easy to see that this implies the same for the limit relation

$$\lim_{t\downarrow 0}\frac{\frac{Q(tx,ty)}{Q(t,t)}-c(x,y)}{q_1(t)} = \tilde{c}_1(x,y)-\tilde{c}_1(1,1)c(x,y) =: c_1(x,y) \tag{6.A.2}$$

with $c_1(x,y)\not\equiv 0$. Clearly q_1 is a regularly varying function with non-negative index.

Proposition 6.A.1. *Under the stated conditions, relations* (6.A.1) *and* (6.A.2) *hold locally uniformly on* $(0,\infty)^2$. *If the index of the regularly varying function* q_1 *is strictly positive, the relation holds locally uniformly on* $[0,\infty)^2$.

Proof. Relation (6.A.2) implies that the function $Q(x,x)$ is regularly varying of second order (cf. de Haan and Stadtmuller, 1996), hence we can assume that $c(x,x) = x^{1/\eta}$ and $c_1(x,x) = x^{1/\eta}\frac{x^\tau-1}{\tau}$ with $1/\eta$ the index of regular variation of q and $\tau\geq 0$ the index of regular variation of q_1.

Let $(x(t),y(t))$ converge to $(x,y)\in(0,\infty)^2$ as $t\downarrow 0$. Write $(x(t),y(t)) = a(t)(u(t),v(t))$ with $u^2(t)+v^2(t)=1$. Then, as $t\downarrow 0$, $(u(t),v(t))\to(u,v)$ and

$a(t) \to a > 0$, say and

$$\begin{aligned}
\frac{Q(tx(t),ty(t))}{Q(t,t)} &= \frac{Q(ta(t)u(t),ta(t)v(t))}{Q(ta(t),ta(t))} \cdot \frac{Q(ta(t),ta(t))}{Q(t,t)} \\
&= \Big(c(u(t),v(t)) + q_1(ta(t))\big[c_1(u(t),v(t)) + o(1)\big]\Big) \\
&\quad \cdot \Big(a(t)^{1/\eta} + q_1(t)\big[c_1(a(t),a(t)) + o(1)\big]\Big) \\
&= c(u(t),v(t))a(t)^{1/\eta} \\
&\quad \cdot \Big(1 + q_1(t)a(t)^{\tau}(1+o(1))\big[\frac{c_1(u(t),v(t))}{c(u(t),v(t))} + o(1)\big]\Big) \\
&\quad \cdot \Big(1 + q_1(t)a(t)^{-1/\eta}\big[c_1(a(t),a(t)) + o(1)\big]\Big) \\
&= c(x(t),y(t))\Big(1 + q_1(t)\big[a(t)^{\tau}\frac{c_1(u(t),v(t))}{c(u(t),v(t))} \\
&\quad + a(t)^{-1/\eta}(c_1(a(t),a(t)) + o(1))\big]\Big).
\end{aligned}$$

It follows that

$$\lim_{t\downarrow 0} \frac{\frac{Q(tx(t),ty(t))}{Q(t,t)} - c(x(t),y(t))}{q_1(t)} = \Big(a^{\tau}\frac{c_1(u,v)}{c(u,v)} + \frac{c_1(a,a)}{c(a,a)}\Big)c(x,y).$$

□

The proof shows that the following relation holds.

Corollary 6.A.1. *For $a, u, v > 0$*

$$c_1(au,av) = a^{1/\eta+\tau}c_1(u,v) + a^{1/\eta}\frac{a^{\tau}-1}{\tau}c(u,v) \tag{6.A.3}$$

(remember we have chosen q_1 in such a way that $c_1(a,a) = a^{1/\eta}\frac{a^{\tau}-1}{\tau}$).

Remark 6.A.1. Write

$$R(s,t) := \frac{c_1(e^s,e^t)}{c(e^s,e^t)}.$$

Then for all h, s and t

$$R(s+h,t+h) - R(s,t) = R(s,t)(e^{h\tau}-1) + R(h,h).$$

Hence

$$R_1(s,t) = \lim_{h\to 0}\frac{R(s+h,t+h) - R(s,t)}{h} = \tau R(s,t) + R_1(0,0) = \tau R(s,t) + 1.$$

This means that for $\tau = 0$,

$$R_1(s,t) = 1 \text{ for all } s,t$$

and $\tau > 0$

$$R_1(s,t) = \tau \tilde{R}(s,t)$$

with $\tilde{R}(s,t) := R(s,t) + 1/\tau$. Hence τ and the values of c_1/c on the unit circle determine the values of $R(s,t)$ everywhere.

Acknowledgments. Holger Drees was partly supported by a Heisenberg fellowship, granted by the DFG. Ana Ferreira's research was partly supported by PRAXIS XXI/FEDER.

Tables and figures

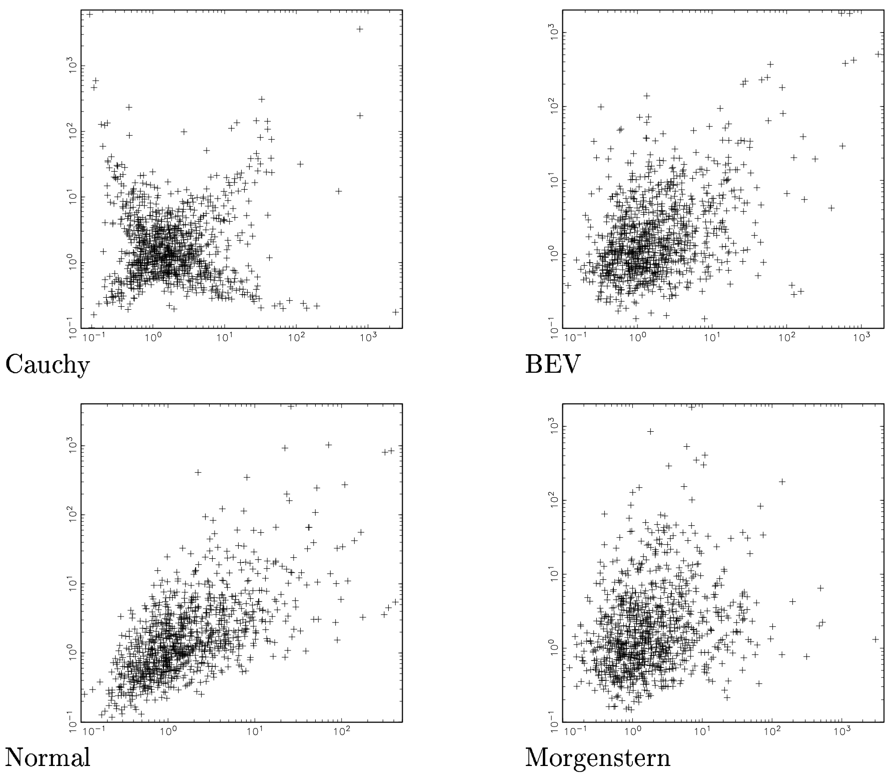


Figure 1: Scatterplot of a sample of size $n = 1000$ of each distribution. The BEV and Morgenstern distributions have Fréchet marginal distributions; for easy comparison a marginal transformation to the same distribution was applied to the Cauchy and normal samples.

Table 1: The ML-estimator, $\hat{\eta}_1$, and the Hill estimator, $\hat{\eta}_2$ (sample size $n = 1000$). Tabulated are mean and observed standard deviation of the estimator, and mean of estimates $\hat{\sigma}_{(i)}$ and $\hat{\sigma}_{(d)}$. The last column indicates the proportion of samples in which asymptotic dependence hypothesis is accepted in size 5% tests, based on $\hat{\sigma}_{(i)}$ resp. $\hat{\sigma}_{(d)}$.

	m_n	$\hat{\eta}$ avg.	Standard deviation observed	$\hat{\sigma}_{(i)}$	$\hat{\sigma}_{(d)}$	$\eta = 1$ accepted; test with $\hat{\sigma}_{(i)}$	with $\hat{\sigma}_{(d)}$
ML, $\hat{\eta}_1$							
Cauchy	80	0.96	0.171	0.167	0.171	0.888	0.916
	160	1.01	0.125	0.112	0.111	0.932	0.952
	240	1.03	0.099	0.083	0.082	0.956	0.964
BEV	80	0.91	0.159	0.146	0.153	0.812	0.860
	160	0.91	0.112	0.094	0.098	0.676	0.720
	240	0.90	0.093	0.070	0.073	0.552	0.584
Normal	80	0.72	0.166	0.125	0.146	0.360	0.384
	160	0.74	0.120	0.080	0.092	0.160	0.184
	240	0.74	0.090	0.059	0.067	0.044	0.052
Morgenstern	80	0.47	0.156	0.123	0.167	0.052	0.060
	160	0.49	0.105	0.077	0.104	0.000	0.000
	240	0.50	0.082	0.057	0.076	0.000	0.000
Hill, $\hat{\eta}_2$							
Cauchy	40	0.93	0.119	0.115	0.124	0.808	0.88
	80	0.89	0.083	0.076	0.085	0.572	0.63
	120	0.84	0.064	0.056	0.067	0.148	0.22
BEV	40	0.87	0.112	0.100	0.114	0.60	0.71
	80	0.84	0.075	0.064	0.076	0.29	0.34
	120	0.82	0.058	0.049	0.059	0.05	0.08
Normal	40	0.73	0.099	0.0817	0.112	0.120	0.184
	80	0.74	0.067	0.0535	0.073	0.008	0.008
	120	0.73	0.052	0.0409	0.056	0.000	0.000
Morgenstern	40	0.51	0.072	0.0663	0.129	0.0	0.0
	80	0.53	0.050	0.0443	0.084	0.0	0.0
	120	0.54	0.042	0.0344	0.064	0.0	0.0

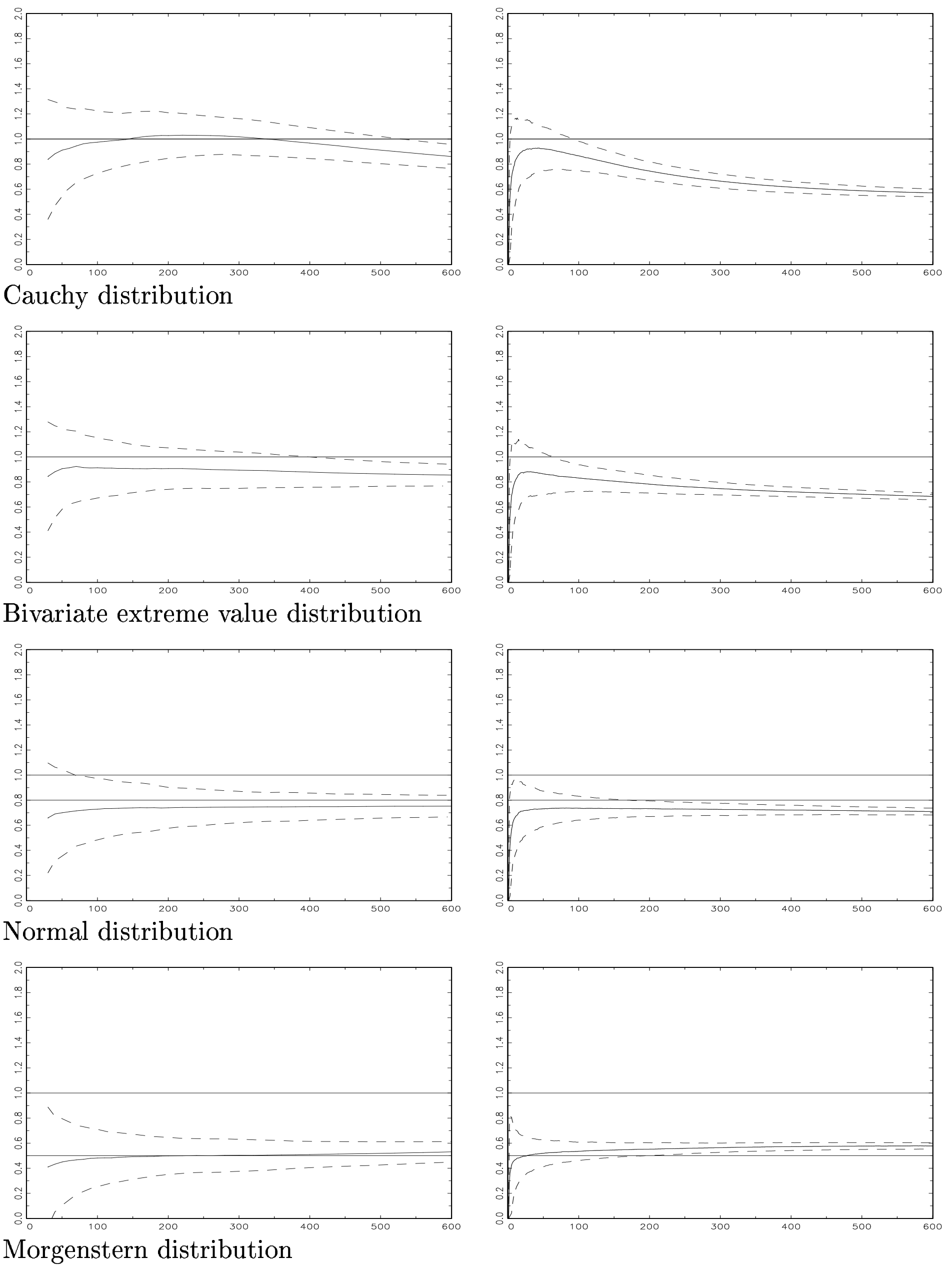


Figure 2: The ML-estimator, $\hat{\eta}_1$, on the left and the Hill estimator, $\hat{\eta}_2$, on the right as a function of m_n (sample size $n = 1000$). The graphs show the average over 250 samples (solid line). Observed standard errors are indicated by the dashed lines ($\pm$ 1.64 st.deviations). The horizontal lines indicate $\eta = 1$ and the true η for each distribution.

Table 2: Pengs estimator, $\hat{\eta}_3$, and our estimator, $\hat{\eta}_4$ (sample size $n = 1000$). Tabulated are mean and observed standard deviation of the estimator, and mean of estimates $\hat{\sigma}_{(i)}$ and $\hat{\sigma}_{(d)}$; the proportion of samples in which asymptotic dependence hypothesis is accepted in size 5% tests, based on $\hat{\sigma}_{(i)}$ resp. $\hat{\sigma}_{(d)}$; the last column gives the number of samples (out of 250) in which either $\hat{\eta}$ or $\hat{\sigma}^2$ could not be calculated.

		$\hat{\eta}$	Standard deviation			$\eta = 1$ accepted		
	k	avg.	observed	$\hat{\sigma}_{(i)}$	$\hat{\sigma}_{(d)}$	Indep.	Dep.	Missing
Peng, $\hat{\eta}_3$								
Cauchy	40	1.05	0.361	0.231	0.249	0.92	1.00	6
	80	0.97	0.178	0.158	0.176	0.88	1.00	1
	120	0.88	0.120	0.114	0.145	0.67	0.97	1
BEV	40	0.96	0.228	0.196	0.232	0.90	1.00	5
	80	0.85	0.124	0.123	0.170	0.60	0.97	2
	120	0.80	0.086	0.093	0.137	0.28	0.67	0
Normal	40	0.78	0.194	0.181	0.304	0.60	1.00	2
	80	0.75	0.093	0.116	0.192	0.27	0.94	0
	120	0.74	0.072	0.086	0.144	0.05	0.27	0
Morgenstern	40	0.55	0.221	0.239	0.741	0.32	1.00	10
	80	0.54	0.108	0.122	0.372	0.03	1.00	0
	120	0.55	0.070	0.088	0.253	0.00	0.10	0
This paper, $\hat{\eta}_4$								
Cauchy	80	1.03	0.230	0.202	0.191	0.891	0.94	1
	160	0.97	0.144	0.133	0.136	0.832	0.92	0
	240	0.89	0.098	0.095	0.109	0.612	0.72	0
BEV	80	0.96	0.174	0.171	0.176	0.83	0.911	2
	160	0.86	0.099	0.098	0.115	0.55	0.680	0
	240	0.82	0.067	0.072	0.090	0.23	0.304	0
Normal	80	0.79	0.146	0.144	0.187	0.500	0.700	0
	160	0.76	0.080	0.083	0.113	0.124	0.260	0
	240	0.75	0.058	0.061	0.084	0.040	0.056	0
Morgenstern	80	0.55	0.187	0.181	0.343	0.228	0.66	0
	160	0.54	0.085	0.087	0.172	0.020	0.04	0
	240	0.55	0.055	0.060	0.117	0.000	0.00	0

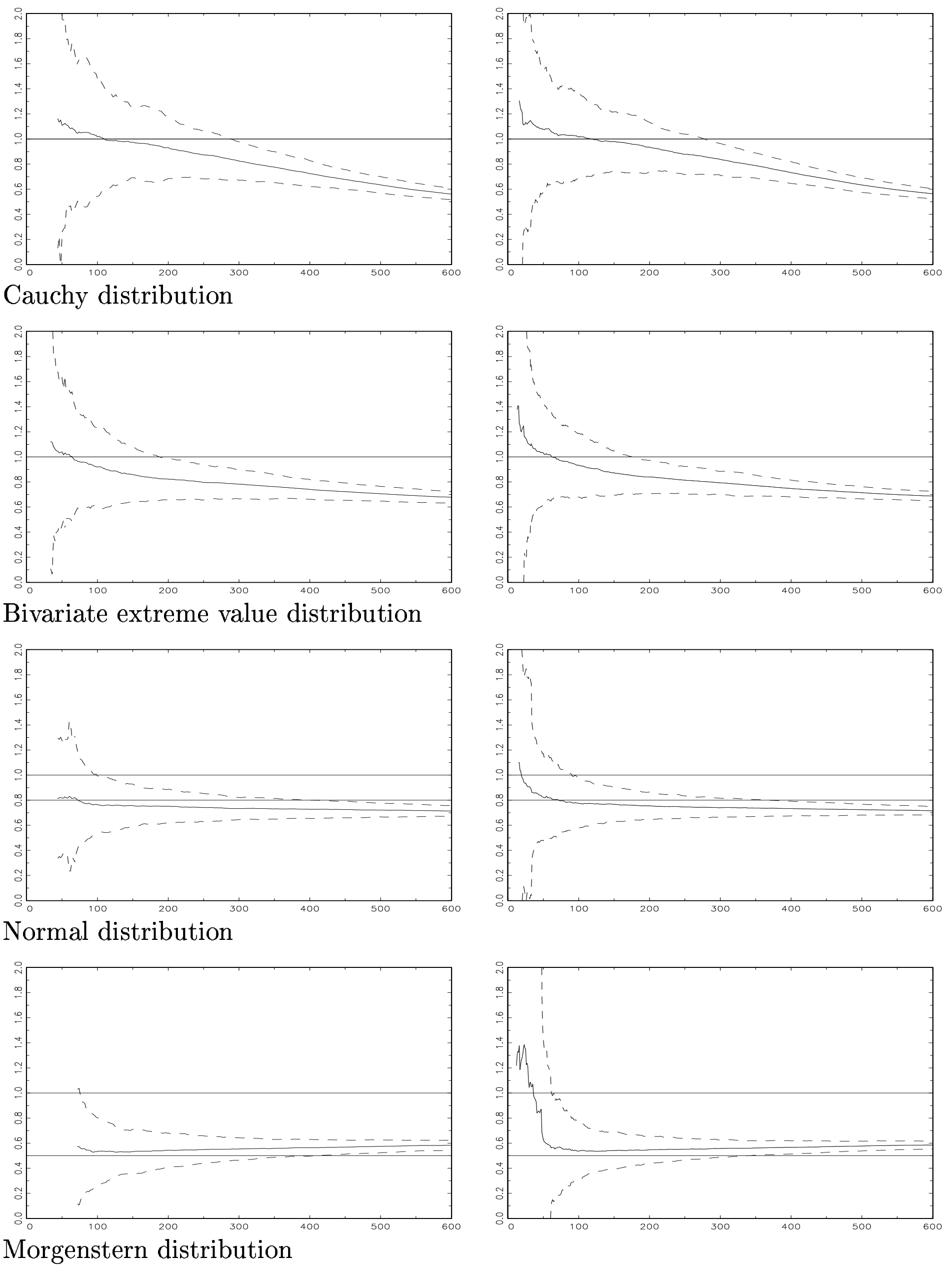
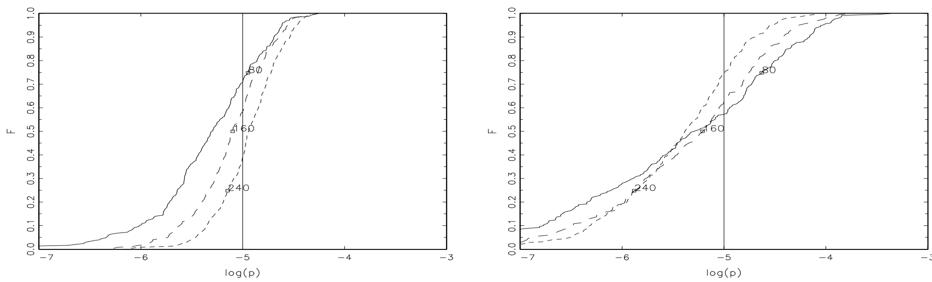


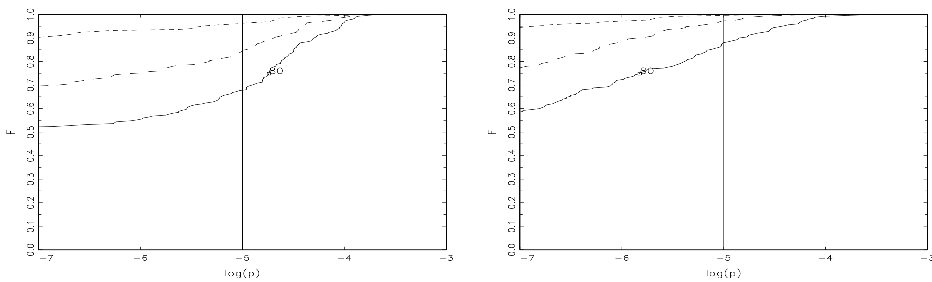
Figure 3: Peng's estimator, $\hat{\eta}_3$, as a function of $2k$ on the left, and our estimator, $\hat{\eta}_4$, as a function of k on the right (sample size $n = 1000$). The graphs show the average over 250 samples (solid line). Observed standard errors are indicated by the dashed lines ($\pm$ 1.64 st.deviations). The horizontal lines indicate $\eta = 1$ and the true η for each distribution.

Table 3: Estimating failure probabilities. The table lists the median values of the estimates. The probability estimates are the estimate for the general case ($\hat{p}_{\hat{\eta}}$), for the asymptotic dependent case ($\hat{p}_1$), and $\hat{p} = \hat{p}_{\hat{\eta}}$ or $= \hat{p}_1$, depending on whether asymptotic dependence is rejected resp. accepted.

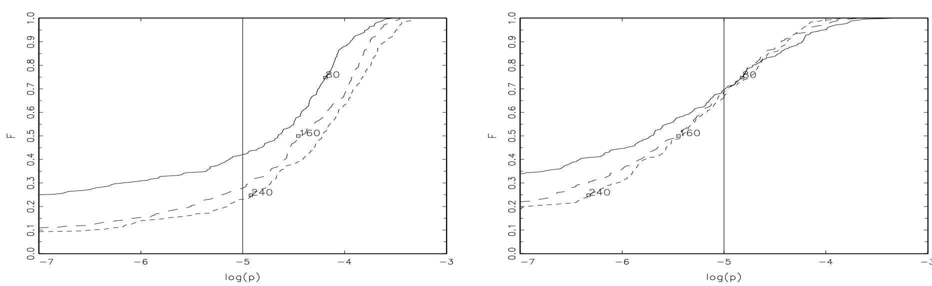
k	γ_1	γ_2	η	$\hat{p}_{\hat{\eta}}$	$\hat{p}_1$	$\hat{p}$
Cauchy	1	1	1		$\times 10^{-5}$	
80	0.95	0.98	1.01	0.52	0.51	0.47
160	1.02	0.96	0.94	0.63	0.80	0.75
240	1.00	1.04	0.90	0.42	1.14	0.99
Normal	0	0	0.8		$\times 10^{-5}$	
80	-0.15	-0.17	0.77	0.00003	0.00240	0.00005
160	-0.20	-0.13	0.75	0.00000	0.00000	0.00000
240	-0.17	-0.20	0.74	0.00000	0.00000	0.00000
Exponential/Normal	0	0	0.8		$\times 10^{-5}$	
80	0.016	0.040	0.77	0.20	2.2	0.61
160	0.062	0.036	0.75	0.36	3.7	0.44
240	0.045	0.058	0.74	0.39	6.1	0.43
Morgenstern	1	1	0.5		$\times 10^{-5}$	
80	0.99	1.01	0.54	1.1	27	19.7
160	1.03	0.97	0.53	1.3	58	1.3
240	1.00	1.02	0.55	2.0	84	2.0



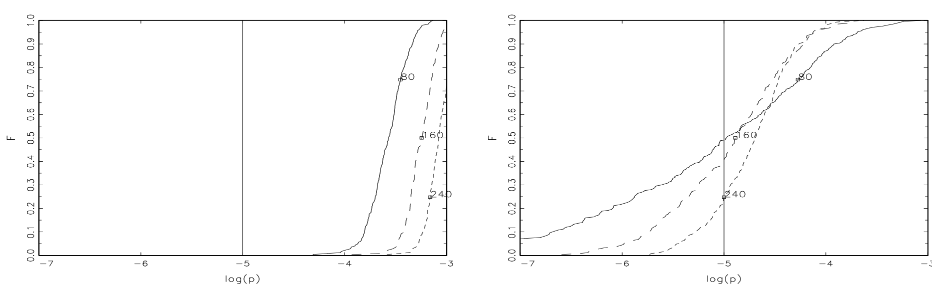
Cauchy distribution



Normal distribution



Normal distribution; exponential margins



Morgenstern distribution

Figure 4: Failure probabilities. The graphs show the empirical distributions of the estimates for different k. The graphs on the left refer to $\hat{p}_1$ (the estimate assuming dependence) and the graphs on the right to $\hat{p}_{\hat{\eta}}$ (the general estimate). The true value $p = 10^{-5}$ is indicated by the vertical line ($n = 1000$).

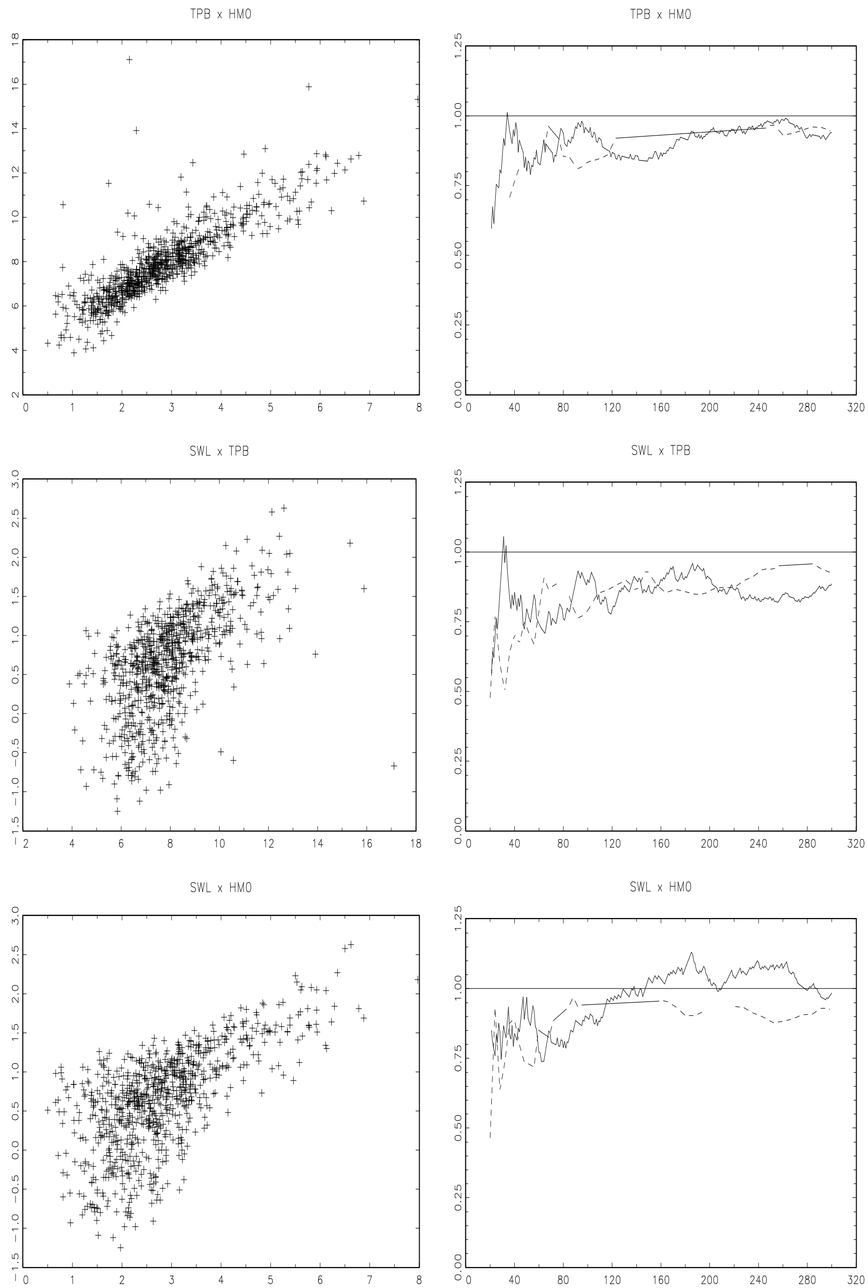


Figure 5: Estimating dependence between waveheight Hm0, waveperiod Tpb and still water level SWL. On the left a scatterplot and on the right $\hat{\eta}_4$ as a function of k. The solid lines display the estimates for various k. The horizontal lines indicate the $\eta = 1$ level. The area below the dotted line is the critical area of a one sided, size 5% test for asymptotic dependence.

Bibliography

[1] AARSSEN, K. AND DE HAAN, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* **4**, 259-281.

[2] ALEXANDER, K.S. (1987) Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probab. Theory Relat. Fields* **75**, 379-423.

[3] ARNOLD, B.C., BALAKRISHNAN, N. AND NAGARAJA, H. N. (1992). *A first course in order statistics.* John Wiley & Sons Inc.

[4] BALKEMA, A. A. AND DE HAAN, L. (1975). Residual life time at great age. *Ann. Probab.* **2**, 792-804.

[5] BILLINGSLEY, P. (1971). *Weak convergence of measures: Applications in Probability.* Regional Conference Series on Applied Mathematics, vol. 5, SIAM, Philadelphia, Pennsylvania.

[6] BINGHAM, N.H., GOLDIE C.M. AND TEUGELS, J.L. (1987). *Regular Variation.* Cambridge University Press.

[7] BOOS, D.D. (1984). Using extreme value theory to estimate large percentiles. *Technometrics* **26**, 33-39.

[8] CAERS, J., BEIRLANT, J. AND VYNCKIER, P. (1998). Bootstrap confidence intervals for tail indices. *Comput. Stat. and Data Analysis* **26**, 259-277.

[9] CHENG, S. AND DE HAAN, L. (2001). Penultimate approximation for Hill's estimator. *Scand. J. Statist.* **28**, 1-7.

[10] CHENG, S. AND PAN, J. (1998). Asymptotic expansions of estimators for the tail index with applications. *Scand. J. Statist.* **25**, 717-728.

[11] CHENG, S. AND PENG, L. (2001). Confidence intervals for the tail index. *Bernoulli* **7**, 751-760.

[12] COLES, S., HEFFERNAN, J. AND TAWN, J. (1999). Dependence measures for extreme value analysis. *Extremes* **2**, 339-366.

[13] Cuntz, A. and Haeusler, E. (2001). On Edgeworth expansions for the distribution function of the Hill estimator. Report.

[14] Danielsson, J., de Haan, L., Peng, L. and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Analysis* **76**, 226-248.

[15] Danielsson, J. and de Vries, C.G. (1997). Tail index and quantile estimation with very high frequency data. *J. of Empirical Finance* **4**, 241-257.

[16] Davis, R. and Resnick, S. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12**, 1467-1487.

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

[18] Dekkers, A.L.M. and de Haan, L. (1989b). On the estimation of the extreme value index and large quantile estimation. *Ann. Statist.* **17**, 1795-1832.

[19] Dekkers, A.L.M. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Analysis* **47**, 173-195.

[20] Dijk, V. and de Haan, L. (1992). On the estimation of the exceedance probability of a high level. In:*Order Statist. and Nonparametrics: Theory and Applications* (P.K. Sen and I.A. Salama, eds.), 79-92. North-Holland, Amsterdam.

[21] Dillingh, D., de Haan, L., Helmers, R., Können, G.P. and van Malde, I.J. (1993). De basispeilen langs de Nederlandse kust. Technical Report DGW-93.023, Rijks Instituut voor Kust en Zee, Rijkswaterstaat, 's-Gravenhage.

[22] Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2001). Tail dependence in independence. *Submitted.* EURANDOM Technical Report 2001-014, Eindhoven.

[23] Draisma, G., de Haan, L., Peng, L. and Pereira, T.T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes* **2**, 367-404.

[24] Drees, H. (1998a). On smooth statistical tail functionals. *Scand. J. Statist.* **25**, 187-210.

[25] Drees, H. (1998b). A general class of estimators of the extreme value index. *J. Statist. Plann. Inference* **66**, 95-112.

[26] Drees, H. de Haan, L. and Resnick, S. (2000). How to make a Hill plot. *Ann. Statist.* **28**, 254-274.

[27] DREES, H. AND KAUFMANN, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes Appl.* **75**, 149-172.

[28] EINMAHL, J.H.J. (1990). The empirical distribution function as a tail estimator. *Statistica Neerlandica* **44**, 79-82.

[29] EINMAHL, J.H.J. (1997). Poisson and Gaussian approximation of weighed local empirical processes. *Stoch. Proc. Appl.* **70**, 31-58.

[30] FERREIRA, A. (2002). Optimal asymptotic estimation of small exceedance probabilities. *J. Statist. Plann. Inference* **104**, 83-102.

[31] FERREIRA, A. DE HAAN, L. AND PENG, L. (1999). On optimizing the estimation of high quantiles of a probability distribution. To appear in Statistics. EURANDOM Technical Report 1999-042, Eindhoven.

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

[33] FRAGA ALVES, M.I., GOMES, M.I. AND DE HAAN, L. (2001). A new class of semi-parametric estimators of the second order parameter. *Submitted.*

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

[35] DE HAAN, L. (1990). Fighting the arch-enemy with mathematics. *Statistica Neerlandica* **44**, 45-68.

[36] DE HAAN, L. (1994). Extreme value statistics. In:*Extreme Value Theory and Applications* (J. Galambos, J. Lechner and E. Simiu, eds.), 93-122. Kluwer Academic, Dordrecht.

[37] DE HAAN, L. AND PEREIRA, T. T. (1999). Estimating the index of a stable distribution. *Statistics and Probability Letters* **41**, 39-55.

[38] DE HAAN, L. AND RESNICK, S.I. (1993). Estimating the limit distribution of multivariate extremes. *Commun. Statist. - Stochastic Models* **9**, 275-309.

[39] DE HAAN, L. AND DE RONDE, J. (1998). Sea and wind: multivariate extremes at work. *Extremes* **1**, 7-45.

[40] DE HAAN, L. AND ROOTZÉN, H. (1993). On the estimation of high quantiles. *J. Statist. Plann. Inference* **35**, 1-13.

[41] DE HAAN, L. AND SINHA, A.K. (1999). Estimating the probability of a rare event. *Ann. Statist.* **27**, 732-759.

[42] DE HAAN, L. AND STADTMÜLLER, U. (1996). Generalized regular variation of second order. *J. Australian Math. Soc. Ser. A* **61**, 381-395.

[43] HALL, P. (1982). On some simple estimators of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44**, 37-42.

[44] HALL, P. AND WEISSMAN, I. (1997). On the estimation of extreme tail probabilities. *Ann. Statist.* **25**, 1311-1326.

[45] HALL, P. AND WELSH, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.

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

[47] HUANG, X. (1992). *Statistics of Bivariate Extreme Values.* PhD thesis, Erasmus University Rotterdam. Tinbergen Institute, TIR series 22. The Netherlands.

[48] JANSEN, D.W. AND DE VRIES, C.G. (1991). On the frequency of the large stock returns: putting booms and busts into perpective. *Review of Economics and Statistics* **73**, 18-24.

[49] JOE, H. (1987). Estimation of quantiles of the maximum of N observations. *Biometrika* **74**, 347-354.

[50] LEDFORD, A. AND TAWN, J.A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169-187.

[51] LEDFORD, A. AND TAWN, J.A. (1997). Modelling dependence within joint tail regions. *J. Royal Statist. Soc. Ser. B* **59**, 475-499.

[52] LEDFORD, A. AND TAWN, J.A. (1998). Concomitant tail behaviour for extremes. *Adv. Appl. Prob.* **30**, 197-215.

[53] MARTINS, M.J. (2000). *Heavy tails estimation - variants to the Hill estimator.* PhD thesis (in Portuguese), University of Lisbon. Portugal

[54] PENG, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics and Probability Letters* **43**, 399-409.

[55] PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131.

[56] POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

[57] RESNICK, S.I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag, New York.

[58] ROSEN, O. AND WEISSMAN, I. (1996). Comparison of estimation methods in extreme value theory. *Commun. Statist.-Theory Meth.* **25**, 759-773.

[59] RUBEN, H. (1964) An asymptotic expansion for the multivariate normal distribution and Mill's ratio. *Journal of Research of the National Bureau of Standards* **68**, 3-11.

[60] SCHOENBERG, R. (1996). *Maximum Likelihood Estimation.* Aptech Systems Inc. Gauss maxlik application.

[61] SHORACK, G. AND WELLNER, J. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[62] SMITH, R.L. (1984). Threshold methods for sample extremes. In:*Statistical Extremes and Applications* (J. Tiago de Oliveira, ed.), 621-638. Reidel, Dordrecht.

[63] SMITH, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174-1207.

[64] WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73**, 812-815.

[65] VERVAAT, W. (1972) Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. und verw. Gebiete* **23**, 249-253.

Summary

Three chapters of the thesis are concerned with optimality problems in univariate extreme value statistics. The last two chapters are concerned with consistency and asymptotic normality of estimators, the first chapter in the univariate setting and the second one in bivariate extreme value theory.

Estimators in extreme value theory (for the extreme value index, for a high quantile or for the probability of an extreme set) are based on a number, say k, of upper order statistics from a sample of n observations. We consider the problem: which sequences $k = k(n) \to \infty$, $k(n)/n \to 0$ are optimal in the sense of balancing the variance and bias components of the estimators. This problem has been solved for high quantile estimators, for estimators of the endpoint of the distribution and for estimators of the probability of an extreme set (outside or at the boundary of the range of the available observations). This is the content of Chapters 2 and 3.

The problem of optimisation has been solved already in the case of the estimation of the extreme value index. In that case one further step has been taken in this thesis namely to construct a confidence interval for the tail index in the case of a non-zero bias, which is the optimal situation. Moreover this has also been extended to high quantiles. These are the contents of Chapter 4.

The remaining two chapters are concerned with the more basic problem of consistency and asymptotic normality of estimators, albeit in more involved situations. Chapter 5 offers a careful treatment of the asymptotic normality of maximum likelihood estimators for the extreme value index as well as the scale. Chapter 6 deals with an estimator of the probability of an extreme set, that comes up when a two dimensional probability distribution function has approximate independent components in the far tail.

Nederlandse samenvatting

De eerste drie hoofdstukken van dit proefschrift gaan over optimaliteitsproblemen in ééndimensionale extreme-waarden theorie. De laatste twee hoofdstukken gaan over consistentie en asymptotische normaliteit van schatters, het eerste van de twee in de ééndimensionale opzet en het tweede in tweedimensionale extreme-waarden theorie.

Schatters in extreme-waarden theorie (n.l. voor de extreme waarde index, voor een hoog kwantiel of voor de kans op een extreme verzameling) zijn gebaseerd op een aantal, zeg k, van bovenste order statistics van een steekproef van n waarnemingen. Wij beschouwen het probleem: welke rijen $\{k\}$ met $k = k(n) \to \infty$, $k(n)/n \to 0$, zijn optimaal in de zin van het afwegen van de variantie- en bias-componenten van de schatter? Dit probleem wordt opgelost voor schatters van een hoog kwantiel, van het eindpunt van de verdeling en voor de kans op een extreme verzameling (buiten of op de grens van het gebied waar waarnemingen beschikbaar zijn). Dit is de inhoud van de hoofdstukken 2 en 3.

In het geval van het schatten van de extreme-waarden-index was dit probleem van optimalisering al opgelost. In dit proefschrift wordt in dat geval een stap verder gezet: een betrouwbaarheidsinterval wordt geconstrueerd voor de staart-index in het geval wanneer bias optreedt. Dat is namelijk de optimale situatie. Die bias is meegenomen in het betrouwbaarheidsinterval. Bovendien wordt dit uitgebreid naar het geval van hoge kwantielen. Dit is de inhoud van hoofdstuk 4.

In de overige twee hoofdstukken wordt teruggegaan naar het basisprobleem van asymptotische normaliteit van schatters, maar dan in meer complexe gevallen. Hoofdstuk 5 geeft een zorgvuldige behandeling van de asymptotische normaliteit van grootste aannemelijkheidsschatters voor zowel de extreme-waarden-index als de schaalfactor. Hoofdstuk 6 behandelt een schatter voor de kans op een extreme verzameling die aan de orde is als een twee-dimensionale kansverdeling ongeveer onafhankelijke componenten heeft in de verre staart.