Bootstrapping Extreme Value Statistics

Master Thesis Econometrics and Management Science
Quantitative Finance

Tim Groen (#402229)

Supervisor: Dr. Chen Zhou
Second reader: Dr. Alex J. Koning

December 16, 2015

Erasmus University Rotterdam
Erasmus School of Economics

Acknowledgements

I would like to thank dr. Chen Zhou and prof. dr. Laurens de Haan for providing the starting point of this thesis, introducing me to their research in extreme value theory and sharing their expertise in this field with me. Furthermore, I would like to thank dr. Alex J. Koning for his general advice and involvement in the project, always keeping an eye on the overall picture. To all three, I feel the need to thank you very much for your guidance, time and enjoyable meetings during this project.

ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Abstract

In this thesis we investigate the use of bootstrapping schemes to estimate the variance of estimators from extreme value theory. We consider estimators for the extreme value index, the central parameter in extreme value theory, and extreme quantiles for two fundamental approaches in extreme value theory. We analyse the Hill estimator for the extreme value index and the Weissman estimator for extreme quantiles from the peaks over threshold approach and the probability weighted moment estimators for the extreme value index and extreme quantiles from the block maxima approach. We find the limiting distributions of a bootstrapped sample and bootstrapped block maxima and subsequently determine the asymptotic behaviour of the bootstrapped Hill estimator and the bootstrapped probability weighted moment estimators. For the latter estimators, we provide an heuristic argument to show that one may use the sample variance of bootstrapped estimators to estimate the variance of the initial estimator.

Keywords: Extreme value index, bootstrapping, Hill estimator, probability weighted moment estimator, extreme quantile estimation.

Contents

1 Introduction

Extreme value theory deals with extreme and rare events. Therefore, it comes as no surprise that extreme value theory has received renewed interest in the past years due to the financial crisis. This branch of statistics models for instance unusually negative stock returns and tries to find the distribution of these very negative returns. However, extreme value theory is also extensively used in insurance, hydrology and earth sciences to determine the probability of default, extreme floods or earthquakes. In all of these cases, one is ultimately trying to determine (properties of) the distribution of the extreme events, i.e. the events in the tails of the original, underlying distribution. The challenge extreme value theory faces is the relative scarcity of extreme events. Therefore, extreme value theory tries to extract as much information as possible from the empirical distribution function.

The fundamental parameter in the distribution of the extreme events is the (first-order) extreme value index. This parameter is used to parametrize the extreme value distribution and to measure the heaviness of the tail of the original, underlying distribution. A positive extreme value index suggests a rather heavy tail and the converse holds for a negative extreme value index. In addition, the extreme value index has a substantial role in the estimation of extreme quantiles and probabilities. Hence, the estimation of the extreme value index and its asymptotic variance is a topic of great interest. A variety of estimators for the extreme value index is available. Most of them have explicit formulas for the asymptotic variance of the estimator. An alternative way to estimate the asymptotic variance of such an estimator is by means of bootstrapping. Particularly for estimators with an inexplicit variance the bootstrap procedure could provide a solution. In general, the idea of using a bootstrap procedure for estimators of the extreme value index and extreme quantiles is relatively unexplored and should provide new insights.

Clearly, a mathematical proof is needed to justify the bootstrapping procedure to estimate the variance of a particular estimator. In this thesis we aim to theoretically justify the use of bootstrapping methods to estimate the asymptotic variance of two extreme value index estimators and two associated extreme quantile estimators. We investigate the Hill estimator and the probability weighted moment estimator for the extreme value index. The analysis for these two estimators will be the core of this thesis and the study of the associated extreme quantile estimators follows as a natural application. The Hill estimator and the probability weighted moment estimator are both based on a number of extreme values from a random sample, however they differ fundamentally on the procedure of determining these extreme values. The Hill estimator is based on the largest values from a random sample corresponding to a peaks over threshold approach, in contrast to the probability weighted moment estimator which is based on block maxima of a random sample.

We mainly rely on theory presented in de Haan and Ferreira (2006) and this book acts as a central reference in this thesis. In addition we use results from recent work by Chen Zhou and Laurens de Haan. Often, a first step is to find a way to theoretically separate the randomness in the bootstrapping procedure into two parts; A first part due to the fact that the empirical distribution function is random and a second part due to the randomness caused by resampling. Resampling is defined as the act of drawing a bootstrap sample from a random sample. For a random sample consisting of i.i.d. random variables this separation is available and it will be the basis for the proofs.

Subsequently, we employ our theoretical results in a practical context by means of simulations and a data analysis to demonstrate their practical use. We consider the daily returns of several large stock market indices and estimate their extreme value indices and extreme quantiles.

We find that the bootstrap procedure asymptotically gives the appropriate variance for the Hill estimator and the probability weighted moment estimator as for their associated extreme quantile estimators. The often used Hill estimator has a variance which is easy to calculate and as a consequence the practical use of bootstrapping this estimator is limited. We do present the proof for the Hill estimator because it is illustrative and insightful for future endeavours. On the other hand, the variance of the probability weighted moment estimator is particularly involved and the bootstrap procedure gives a computationally easy way to calculate it. Furthermore, simulations indicate that the bootstrapping procedure gives a reliable variance estimate for the Hill estimator and probability weighted moment estimator even for a sample of 1000 observations.

2 Preliminaries

In this section we introduce concepts and theory needed for the proofs in Section 3 and Section 4. Preliminary definitions are given in Section 2.1. Subsequently, we introduce the convergence of random variables and corresponding notation in Section 2.2. We present basic extreme value theory in Section 2.3 and state successive theorems concerning intermediate order statistics in Section 2.4. Finally, we familiarize the reader with our bootstrapping schemes in Section 2.5.

2.1 Introductory definitions

We give the definitions of the standard Pareto distribution and standard Fréchet distribution, as they are used extensively in the coming sections.

Definition 2.1. Suppose X is a random variable with a standard Pareto distribution. Then its CDF $F_X(x)$ is given by

$$F_X(x) = 1 - \frac{1}{x} \quad \text{for } x \geq 1.$$

Definition 2.2. Suppose X is a random variable with a standard Fréchet distribution. Then its CDF $F_X(x)$ is given by

$$F_X(x) = \exp(-1/x) \quad \text{for } x > 0.$$

The standard Fréchet distribution is *max stable*, that is

$$\max_{j=1,\ldots,m} X_j \stackrel{d}{=} mX_0$$

with X_j, $j = 0, \ldots, m$ i.i.d. random variables from a standard Fréchet distribution. The symbol $\stackrel{d}{=}$ means equality in cumulative distribution.

An arbitrary distribution function F does not need to allow an inverse because it might not be injective on its domain. However we can always define its left-continuous inverse, which approximates the inverse function and coincides with the inverse function in case of injectivity.

Definition 2.3. Let $f : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function. Then its *left-continuous inverse* $f^{\leftarrow}$ is defined by

$$f^{\leftarrow}(x) = \inf\{y : f(y) \geq x\}.$$

This $f^{\leftarrow}$ is left continuous, non-decreasing and has the following property

$$u \leq f(x) \iff f^{\leftarrow}(u) \leq x$$

which is called the *switching formula*. Furthermore, because of its non-decreasing nature, the left-continuous inverse does preserve order. In the presence of a sample we may define the empirical distribution function.

Definition 2.4. Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution function F. Then the empirical cumulative distribution function F_n is defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \leq x}$$

where $\mathbf{1}$ is the indicator function.

2.2 Convergence of random variables

In a non-stochastic world the usual understanding, by means of a metric, of a limit suffices. For stochastic variables their distance cannot be measured by just the absolute value of their difference and the need arises to incorporate probability, which is exactly what the following notions of stochastic convergence do. The following definitions can all be found in Serfling (2008).

Definition 2.5. A sequence of random variables $X_1, X_2, \ldots$ is said to *converge in distribution* to a random variable X, if

$$\lim_{n\to\infty} P(X_n \le x) = P(X \le x) \tag{2.2.1}$$

for all x for which the function $x \mapsto P(X \le x)$ is continuous. Notation: $X_n \xrightarrow{d} X$ as $n \to \infty$.

Definition 2.6. Let $X_1, X_2, \ldots$ and X be random variables defined on the same probability space (i.e. functions $\Omega \mapsto \mathbb{R}$ for a certain sample space Ω). The sequence $\{X_n\}_{n=1}^{\infty}$ is said to *converge in probability* to X, if for any $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = 0. \tag{2.2.2}$$

Notation: $X_n \xrightarrow{P} X$ as $n \to \infty$. We note that Equation 2.2.2 is actually a shorter version of

$$\lim_{n\to\infty} P\left(\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\right) = 0 \quad \text{for any } \varepsilon > 0.$$

Definition 2.7. Let $X_1, X_2, \ldots$ and X be random variables defined on the same probability space (i.e. functions $\Omega \mapsto \mathbb{R}$ for a certain sample space Ω). The sequence $\{X_n\}_{n=1}^{\infty}$ is said to *converge almost surely* to a random variable X if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{2.2.3}$$

Notation: $X_n \xrightarrow{a.s.} X$ as $n \to \infty$. Again, Equation 2.2.3 denotes a shorter version of

$$P\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

Theorem 2.1. *Assume $X, X_1, X_2, \ldots$ are random variables, then*

(i) $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{d} X$;

(ii) $X_n \xrightarrow{a.s.} X$ implies $X_n \xrightarrow{P} X$.

A proof of this theorem can be found in any textbook on stochastic convergence. We refer to Serfling (2008, Theorem 1.3.1) and Serfling (2008, Theorem 1.3.3). In addition, we introduce stochastic small-o en big-O notation.

Definition 2.8. For a sequence of random variables $X_1, X_2, \ldots$ and a corresponding sequence of constants $a_1, a_2, \ldots$

(i) $X_n = o_P(a_n)$ is defined as $X_n / a_n \xrightarrow{P} 0$ as $n \to \infty$.

(ii) $X_n = O_P(a_n)$ means that for all $\varepsilon > 0$ there exists an $M > 0$ such that for all n we have

$$P(|X_n/a_n| > M) < \varepsilon.$$

We say that X_n/a_n is bounded in probability.

Obviously $X_n = o_P(a_n)$ implies $X_n = O_P(a_n)$. Often, it is more convenient to use small-o notation in comparison to a statement about convergence in probability as it is an easy way to give information about the speed of convergence. Going one step further, most of the random variables we encounter will be a function of a real number $s \in D \subset \mathbb{R}$ which gives rise to the notion of uniform convergence in probability.

Definition 2.9. Let $X_1(s), X_2(s), \dots$ and $X(s)$ be random variables defined for each $s \in D$ on the same probability space. The sequence $\{X_n(s)\}_{n=1}^{\infty}$ is said to *converge uniformly in probability for $s \in D \subset \mathbb{R}$* to $X(s)$, if

$$\sup_{s\in D} |X_n(s) - X(s)| = o_P(1). \tag{2.2.4}$$

Regularly, we encounter a slight variation in the form of

$$\sup_{s\in D} c(s)^{-1} |X_n(s) - X(s)| = o_P(1). \tag{2.2.5}$$

with $c(s)$ a non-stochastic function of s which is positive on D. In this case we also denote Equation 2.2.5 as

$$X_n(s) = X(s) + c(s)o_P(1) \tag{2.2.6}$$

uniformly for $s \in D$. In the next lemma we show that addition works as one would expect with respect to this notiation.

Lemma 2.1. *Suppose $X_1(s), X_2(s), \dots$ and $Y_1(s), Y_2(s), \dots$ are both sequences of random variables such that*

$$X_n(s) = c_x(s)o_P(1) \tag{2.2.7}$$

$$Y_n(s) = c_y(s)o_P(1) \tag{2.2.8}$$

both uniformly for $s \in D$ with c_x, c_y positive on D. Then

$$X_n(s) + Y_n(s) = \left(c_x(s) + c_y(s)\right) o_P(1) \tag{2.2.9}$$

uniformly for $s \in D$.

Proof.

$$0 \le \sup_{s\in D} \frac{|X_n(s) + Y_n(s)|}{c_x(s) + c_y(s)} \le \sup_{s\in D} \frac{|X_n(s)| + |Y_n(s)|}{c_x(s) + c_y(s)} \le \sup_{s\in D} \max\left(\frac{|X_n(s)|}{c_x(s)}, \frac{|Y_n(s)|}{c_y(s)}\right) \tag{2.2.10}$$

Here we use that for any $a, b \in \mathbb{R}_{\ge 0}$, $c, d \in \mathbb{R}_{>0}$ we have $\frac{a+b}{c+d} \le \max\left(\frac{a}{c}, \frac{b}{d}\right)$. Furthermore, we see that

$$\sup_{s\in D} \max(f(x), g(x)) = \max\left(\sup_{s\in D} f(x), \sup_{s\in D} g(x)\right) \tag{2.2.11}$$

for any real functions f, g with D a subset of their domain. Consequently

$$\sup_{s\in D} \max\left(\frac{|X_n(s)|}{c_x(s)}, \frac{|Y_n(s)|}{c_y(s)}\right) = \max\left(\sup_{s\in D}\frac{|X_n(s)|}{c_x(s)}, \sup_{s\in D}\frac{|Y_n(s)|}{c_y(s)}\right) = o_P(1). \tag{2.2.12}$$

□

In the next lemma we deal with multiplication with respect to small-o notation.

Lemma 2.2. *Suppose we have sequences of random variables $(W_n(s))_{n\geq 0}$, $(X_n(s))_{n\geq 0}$, $(Y_n(s))_{n\geq 0}$ and $(Z_n(s))_{n\geq 0}$ such that*

$$W_n(s) = X_n(s) + c_1(s)o_P(1)$$
$$Y_n(s) = Z_n(s) + c_2(s)o_P(1)$$

both uniformly for $s \in D$ with c_1, c_2 positive on D. Furthermore, suppose

$$\sup_{s\in D}\frac{|X_n(s)|}{c_1(s)} = O_P(1) \qquad \text{and} \qquad \sup_{s\in D}\frac{|Z_n(s)|}{c_2(s)} = O_P(1).$$

Then

$$W_n(s)Y_n(s) = X_n(s)Z_n(s) + c_1(s)c_2(s)o_P(1)$$

uniformly for $s \in D$.

Proof. For the sake of readability we suppress the (s) in the random variables and c_i. Notice that

$$W_nY_n - X_nZ_n = (W_n - X_n)(Y_n - Z_n) + W_nZ_n + X_nY_n - 2X_nZ_n$$

and therefore

$$\begin{aligned}\sup_{s\in D}\frac{|W_nY_n - X_nZ_n|}{c_1c_2} &\leq \sup_{s\in D}\frac{|(W_n - X_n)(Y_n - Z_n)|}{c_1c_2} + \sup_{s\in D}\frac{|W_nZ_n + X_nY_n - 2X_nZ_n|}{c_1c_2} \\ &:= \Delta_1 + \Delta_2.\end{aligned}$$

First we deal with Δ_1:

$$\Delta_1 = \sup_{s\in D}\frac{|W_n - X_n|}{c_1}\frac{|Y_n - Z_n|}{c_2} \leq \left(\sup_{s\in D}\frac{|W_n - X_n|}{c_1}\right)\left(\sup_{s\in D}\frac{|Y_n - Z_n|}{c_2}\right) = o_P(1).$$

And subsequently we consider Δ_2:

$$\begin{aligned}\Delta_2 &\leq \sup_{s\in D}\frac{|W_nZ_n - X_nZ_n|}{c_1c_2} + \sup_{s\in D}\frac{|X_nY_n - X_nZ_n|}{c_1c_2} \\ &= \sup_{s\in D}\frac{|Z_n|}{c_2}\frac{|W_n - X_n|}{c_1} + \sup_{s\in D}\frac{|X_n|}{c_1}\frac{|Y_n - Z_n|}{c_2} \\ &\leq \left(\sup_{s\in D}\frac{|Z_n|}{c_2}\right)\left(\sup_{s\in D}\frac{|W_n - X_n|}{c_1}\right) + \left(\sup_{s\in D}\frac{|X_n|}{c_1}\right)\left(\sup_{s\in D}\frac{|Y_n - Z_n|}{c_2}\right) \\ &= O_p(1)o_P(1) + O_p(1)o_P(1) \\ &= o_P(1).\end{aligned}$$

□

2.3 Extreme value distributions

Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed random variables with distribution function F. The first step in extreme value theory is to investigate the distribution of $\max(X_1, \ldots, X_n)$ as $n \to \infty$. Note that the latter quantity will converge to the right endpoint of the underlying distribution and therefore a normalization is required. Suppose there exist sequences of constants $(a_n)_{n \in \mathbb{Z}_{\geq 1}}$ and $(b_n)_{n \in \mathbb{Z}_{\geq 1}}$ with $a_n > 0$ such that

$$\frac{\max(X_1, \ldots, X_n) - b_n}{a_n}$$

has a non-degenerate limit distribution as $n \to \infty$, that is

$$\lim_{n \to \infty} F^n(a_n x + b_n) = G(x) \tag{2.3.1}$$

for every x where G is continuous, and G a non-degenerate distribution function. The Fisher-Tippett-Gnedenko theorem states that the only distributions G that occur (for suitably chosen a_n and b_n) as the previous limit are of the form

$$G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \quad 1 + \gamma x > 0, \tag{2.3.2}$$

with $\gamma \in \mathbb{R}$. For $\gamma = 0$ we have $G_\gamma(x) = \exp(-\exp(-x))$. The distributions in Equation 2.3.2 are called the extreme value distributions and the parameter γ is called the (first-order) extreme value index. Besides parametrizing the extreme value distributions, this extreme value index is of importance in the approximation of extreme probabilities and extreme quantiles. If F satisfies Equation 2.3.1 we say that F is in the domain of attraction of G_γ i.e. $F \in \mathcal{D}(G_\gamma)$. This is well-defined in the sense that $F \in \mathcal{D}(G_{\gamma_1})$ and $F \in \mathcal{D}(G_{\gamma_2})$ implies $\gamma_1 = \gamma_2$. Furthermore, *first-order conditions* for F to be in the domain of attraction of G_γ are available: For every distribution function F we may define the related function $U := \left(\frac{1}{1-F}\right)^{\leftarrow}$. Then $F \in \mathcal{D}_\gamma$ if and only if there is a positive function a such that for $x > 0$,

$$\lim_{t \to \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}. \tag{2.3.3}$$

In case $\gamma = 0$ the right-hand side is interpreted as $\lim_{\gamma \to 0} \frac{x^\gamma - 1}{\gamma} = \log x$. For $\gamma > 0$ the relation above is equivalent to

$$\lim_{t \to \infty} \frac{U(tx)}{U(t)} = x^\gamma \quad \text{for} \quad x > 0. \tag{2.3.4}$$

The discussion above gave us information about the maximum of a sample. Often, however, we are also interested in other statistics besides the maximum. Let us denote the n-th order statistics by $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. The asymptotic properties of the $X_{n-k,n}$ as $n \to \infty$ are of particular interest. Suppose we fix $k \in \mathbb{Z}_{>0}$ and let n tend to infinity, then one can show that the $X_{n-k,n}$ can be approximated by a Poisson point process. Another case, the so called central order statistics, considers $k(n)/n \to p \in (0, 1)$ as $n \to \infty$. In this thesis we study yet a third set of order statistics $X_{n-k,n}$ with $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$ called the *intermediate order statistics*.

In the context of intermediate order statistics we often need a second-order condition to ensure favourable behaviour of F. We state de Haan and Ferreira (2006, Definition 2.3.1) here.

The function U (or the probability distribution connected to it) is said to satisfy the *second-order condition* if for some positive function a and some positive or negative function A with $\lim_{t\to\infty} A(t) = 0$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma,\rho}(x), \quad x > 0, \tag{2.3.5}$$

where $H_{\gamma,\rho}$ is some function that is not a multiple of the function $\frac{x^\gamma-1}{\gamma}$. In particular $H_{\gamma,\rho}$ should not be identically zero. If U satisfies the second-order condition then a and A can be chosen such that $H_{\gamma,\rho}$ is of the following form

$$H_{\gamma,\rho} = \frac{1}{\rho}\left(\frac{x^{\gamma+\rho}-1}{\gamma+\rho} - \frac{x^\gamma-1}{\gamma}\right). \tag{2.3.6}$$

For $\gamma = 0$, $\rho = 0$ and/or $\gamma + \rho = 0$ Equation 2.3.6 is understood to be equal to the limit of $\gamma \to 0$, $\rho \to 0$ and/or $\gamma + \rho \to 0$. We also define the function $\Psi_{\gamma,\rho}$ which is closely related to the function $H_{\gamma,\rho}$. Actually, it is the result of the limit in Equation 2.3.5 for a different choice of a and A and has the following definition

$$\Psi_{\gamma,\rho}(x) = \begin{cases} \frac{x^{\gamma+\rho}-1}{\gamma+\rho}, & \gamma+\rho \neq 0,\ \rho < 0, \\ \log x, & \gamma+\rho = 0,\ \rho < 0, \\ \frac{1}{\gamma}x^\gamma \log x, & \rho = 0 \neq \gamma, \\ \frac{1}{2}(\log x)^2, & \rho = 0 = \gamma. \end{cases} \tag{2.3.7}$$

Additionally we need, for positive γ, a slightly stronger version of the second-order condition. For $\gamma > 0$ and $\rho \leq 0$ the following condition implies the second order condition;

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\gamma\rho} \quad \text{for } x > 0, \tag{2.3.8}$$

with A a possibly different function than in Equation 2.3.5.

If F satisfies the second-order condition for some γ and ρ then it is clear that F also satisfies the first-order condition (Equation 2.3.3) and therefore is in the domain of attraction of G_γ.

Finally we present a set of inequalities known as the *Potter inequalities*:

Theorem 2.2 (Potter, 1942). *Suppose U satisfies Equation 2.3.4. For $\delta_1, \delta_2 > 0$, there exists $t_0 = t_0(\delta_1, \delta_2)$ such that for $t \geq t_0$, $tx \geq t_0$,*

$$(1-\delta_1)x^\gamma \min\left(x^{\delta_2}, x^{-\delta_2}\right) < \frac{U(tx)}{U(t)} < (1+\delta_1)x^\gamma \max\left(x^{\delta_2}, x^{-\delta_2}\right). \tag{2.3.9}$$

2.4 Intermediate order statistics

The proofs in Section 3 and Section 4 build on theorems from extreme value theory and in particular on theorems involving approximations of the intermediate order statistics. We state the most important theorems in this subsection in an attempt to clarify the proofs in the coming sections. We start with two theorems employed in Section 3 and introduce the following notation; denote the smallest integer greater or equal to $x \in \mathbb{R}$ by $\lceil x \rceil$ and denote the largest integer less or equal to $x \in \mathbb{R}$ by $[x]$.

Theorem 2.3 (de Haan and Ferreira, 2006, Theorem 2.4.8)**.** *Assume $X_1, X_2, \dots$ are i.i.d. random variables with distribution function F and let $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ be its n-th order statistics. Suppose that F satisfies the second-order condition for some $\gamma > 0$ and $\rho \leq 0$ (Equation 2.3.8). Then there exists a sequence of Brownian motions $\{W_n(s)\}_{s\geq 0}$ and a function A_0 such that, for $\varepsilon > 0$ sufficiently small,*

$$\sqrt{k}\left(\frac{\log X_{n-[ks],n} - \log U(\frac{n}{k})}{\gamma} + \log s\right) = \frac{W_n(s)}{s} + \sqrt{k}A_0\left(\frac{n}{k}\right)\frac{1}{\gamma}\frac{s^{-\rho}-1}{\rho} + s^{-1/2-\varepsilon}o_p\left(1\right) \quad (2.4.1)$$

uniformly for $s \in (0,1]$ as $n \to \infty$, provided $k = k(n) \to \infty, k(n)/n \to 0$ and $\sqrt{k}A_0(n/k) = O(1)$.

Theorem 2.4 (de Haan and Ferreira, 2006, Lemma 2.4.10)**.** *Assume $Y_1, Y_2, \dots$ are i.i.d. random variables with a standard Pareto distribution and let $Y_{1,n} \leq Y_{2,n} \leq \dots Y_{n,n}$ be its n-th order statistics. For each $\gamma \in \mathbb{R}$ there exists a sequence of Brownian motions $\{W_n(s)\}_{s>0}$ such that for each $\varepsilon > 0$*

$$\sup_{k^{-1}\leq s\leq 1} s^{\gamma+1/2+\varepsilon}\left|\sqrt{k}\left(\frac{\left(\frac{k}{n}Y_{n-[ks],n}\right)^{\gamma}-1}{\gamma} - \frac{s^{-\gamma}-1}{\gamma}\right) - s^{-\gamma-1}W_n(s)\right| = o_P(1) \quad (2.4.2)$$

as $n \to \infty$, provided $k = k(n) \to \infty$ and $k/n \to 0$.

The proofs of Theorem 2.3 and Theorem 2.4 can be found in de Haan and Ferreira (2006). For our endeavours concerning the Hill estimator we need a corollary of Theorem 2.4.

Lemma 2.3. *Assume $Y_1, Y_2, \dots$ are i.i.d. random variables with a standard Pareto distribution and let $Y_{1,n} \leq Y_{2,n} \leq \dots Y_{n,n}$ be its n-th order statistics. Then for each $\xi \in \mathbb{R}$ and $\varepsilon > 0$*

$$\sup_{k^{-1}\leq s\leq 1} s^{\xi+1/2+\varepsilon}\left|\left(\frac{k}{n}Y_{n-[ks],n}\right)^{\xi} - s^{-\xi}\right| = o_P(1) \quad (2.4.3)$$

as $n \to \infty$, provided $k = k(n) \to \infty$ and $k/n \to 0$.

Proof.

$$
\begin{aligned}
&\sup_{k^{-1}\le s\le 1} s^{\xi+1/2+\varepsilon}\left|\left(\frac{k}{n}Y_{n-[ks],n}\right)^{\xi}-s^{-\xi}\right| \\
&=|\xi|\sup_{k^{-1}\le s\le 1} s^{\xi+1/2+\varepsilon}\left|\frac{\left(\frac{k}{n}Y_{n-[ks],n}\right)^{\xi}-1}{\xi}-\frac{s^{-\xi}-1}{\xi}\right| \\
&=|\xi|\sup_{k^{-1}\le s\le 1} s^{\xi+1/2+\varepsilon}\left|\frac{\left(\frac{k}{n}Y_{n-[ks],n}\right)^{\xi}-1}{\xi}-\frac{s^{-\xi}-1}{\xi}-\frac{s^{-\xi-1}}{k^{1/2}}W_n(s)+\frac{s^{-\xi-1}}{k^{1/2}}W_n(s)\right| \\
&\le|\xi|\sup_{k^{-1}\le s\le 1} s^{\xi+1/2+\varepsilon}\left|\frac{\left(\frac{k}{n}Y_{n-[ks],n}\right)^{\xi}-1}{\xi}-\frac{s^{-\xi}-1}{\xi}-k^{-1/2}s^{-\xi-1}W_n(s)\right| \\
&\quad+|\xi|\sup_{k^{-1}\le s\le 1} s^{\xi+1/2+\varepsilon}\left|k^{-1/2}s^{-\xi-1}W_n(s)\right| \\
&:=|\xi|\left(\Delta_1+\Delta_2\right)
\end{aligned}
$$

In the derivation above we take $W_n(s)$ as in Theorem 2.4. By the same theorem we see that $\Delta_1 = o_P\left(1/\sqrt{k}\right)$. For Δ_2 we observe the following

$$
\begin{aligned}
\Delta_2 =& \sup_{k^{-1}\le s\le 1} k^{-1/2}s^{-1/2+\varepsilon}\left|W_n(s)\right| \\
\le&\left(\sup_{k^{-1}\le s\le 1} k^{-1/2}s^{-1/2+\varepsilon}\right)\left(\sup_{k^{-1}\le s\le 1}\left|W_n(s)\right|\right) \\
\le&\left(k^{-1/2}(k^{-1})^{-1/2+\varepsilon}\right)O_P(1) \\
=&O_P(k^{-\varepsilon})
\end{aligned}
$$

We choose $\varepsilon \le 1/2$ such that the function $s^{-1/2+\varepsilon}$ is not increasing. In case $\varepsilon > 1/2$ the latter function is increasing, consequently finding $\Delta_2 = O_p(k^{-1/2})$. By assumption $k \to \infty$ and therefore $\Delta_2 = o_P(1)$ (in both cases) which implies $|\xi|(\Delta_1 + \Delta_2) = o_P(1)$. □

In case of the probability weighted moment estimator we need a slightly different theorem than Theorem 2.3 which is due to Drees (1998) and can also be found in de Haan and Ferreira (2006) as Theorem 2.4.2.

Theorem 2.5 (Drees, 1998, Theorem 2.1)**.** *Assume $X_1, X_2, \dots$ are i.i.d. random variables with distribution function F and let $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$ be its n-th order statistics. Suppose that F satisfies the second-order condition for some $\gamma \in \mathbb{R}$ and $\rho \le 0$ (Equation 2.3.5). Then there exists a sequence of Brownian motions $\{W_n(s)\}_{s>0}$ such that for suitably chosen functions a_0 and A_0 and each $\varepsilon > 0$*

$$
\sqrt{k}\left(\frac{X_{n-[ks],n}-U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)}-\frac{s^{-\gamma}-1}{\gamma}\right)=s^{-\gamma-1}W_n\left(s\right)+\sqrt{k}A_0\left(\frac{n}{k}\right)\Psi_{\gamma,\rho}(s^{-1})+s^{-\gamma-1/2-\varepsilon}o_P(1) \tag{2.4.4}
$$

uniformly for $s \in [k^{-1}, 1]$ as $n \to \infty$, provided $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A(n/k) = O(1)$.

We need an extended version of the latter theorem in the form of the following result:

Theorem 2.6 (de Haan and Ferreira, 2006, Extension of Corollary 2.4.5)**.** *Define*

$$B_0\left(\frac{n}{k}\right) := \begin{cases} U\left(\frac{n}{k}\right) & \text{if } \gamma \geq -\frac{1}{2} \\ X_{n,n} + \frac{a_0\left(\frac{n}{k}\right)}{\gamma} & \text{if } \gamma < -\frac{1}{2}. \end{cases} \tag{2.4.5}$$

Then, under the conditions of Theorem 2.5,

$$\sup_{0<s\leq\lambda(n)} \min\left(s^{\gamma+1/2+\varepsilon}, s^{\gamma+\rho-\varepsilon}\right) \left| \sqrt{k}\left(\frac{X_{n-[ks],n} - B_0\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} - \frac{s^{-\gamma}-1}{\gamma}\right) - s^{-\gamma-1}W_n(s) - \sqrt{k}A_0\left(\frac{n}{k}\right)\Psi_{\gamma,\rho}(s^{-1})\right| \xrightarrow{P} 0 \tag{2.4.6}$$

as $n \to \infty$, provided $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A_0(n/k) = O(1)$ and with $\lambda(n) = O\left((n/k)^{1-\tau}\right)$ for $\tau > 0$ such that $\tau > \frac{\varepsilon}{\varepsilon+1/2}$.

A proof of this theorem can be found in the appendix, where we extend the original proof of de Haan and Ferreira (2006, Corollary 2.4.5).

Theorem 2.7 (Ferreira and de Haan, 2015, Lemma 4.1.3)**.** *Assume $Z_1, Z_2, \dots, Z_k$ are i.i.d. random variables from a standard Fréchet distribution. In addition, assume that $\nu \in (0, 1/2)$ and $\xi \in \mathbb{R}$, then there exists an appropriate sequence of Brownian bridges $(E_k)_{k\geq1}$ such that*

$$\sqrt{k}s(-\log s)^{1+\xi}\left(\frac{Z^{\xi}_{\lceil ks\rceil,k}-1}{\xi} - \frac{(-\log s)^{-\xi}-1}{\xi}\right) - E_k(s) = (s(1-s))^{\nu}o_P(1) \tag{2.4.7}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$ as $k \to \infty$.

2.5 Bootstrapping schemes

Bootstrapping is a way of acquiring information about the sampling distribution of a given functional T, which is a function of the data. The bootstrap resamples from the given random sample and by resampling finds new values for T which make up an approximate sampling distribution for T. Bootstrapping was pioneered by Bradley Efron in Efron (1979) and Efron and Tibshirani (1993).

Assume $X_1, X_2, \dots, X_n$ are i.i.d. random variables from a distribution F. Then we are interested in the sampling distribution of a functional $T(X_1, \dots, X_n, F)$. However, the random sample gives just one value of T. By randomly picking an element from $(X_1, \dots, X_n)$ and repeating this n times we resample from the given random sample and generate a bootstrapped sample $X_1^*, \dots, X_n^*$. Note that this resampling is equivalent to drawing n times (independently) from the empirical distribution function F_n. With this new bootstrapped sample $X_1^*, \dots, X_n^*$ and F_n we may calculate $T(X_1^*, \dots, X_n^*, F_n)$. By repeating the resampling procedure we find m bootstrapped values for T, from which we can build an approximate sampling distribution for T. In this thesis we are only interested in an estimate for the variance of T for some specific T.

The bootstrap procedure is useful in case the theoretical sampling distribution is not known or hard to compute. In our case the theoretical variance of at least one estimator is cumbersome to calculate. As bootstrapping is a computationally light procedure it could provide an estimator for this variance which is easy to calculate.

We should formalise what it means for this bootstrapping procedure to work. Assume we have bootstrapped $T_i^* := T\left(X_1^{*,i}, \ldots, X_n^{*,i}, F_n\right)$ for $i = 1, \ldots, m$ with average $\overline{T^*}$ and denote the theoretical variance of $T(X_1, \ldots, X_n, F)$ by $\text{var}(T)$. Then, in case of the variance, we are looking for a statement of the sort

$$\frac{\frac{1}{m}\sum_{i=1}^{m}\left(T_i^* - \overline{T^*}\right)^2}{\text{Var}(T)} \xrightarrow{P} 1 \tag{2.5.1}$$

as $m \to \infty$ and $n \to \infty$. In this case the theoretical variance can be consistently estimated by the sample variance of the bootstrapped functionals. As will become clear in the next sections it is not trivial to rigorously prove this kind of statement in our extreme value theory context.

In applying the bootstrap procedure we allow for randomness in two ways. The first part of the randomness is due to the fact that the empirical distribution function is random and the second part is caused by the resampling. It will prove very convenient to split these two parts. We start by defining the empirical equivalent of U:

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

Suppose Y is standard Pareto distributed and independent from $X_1, X_2, \ldots, X_n$, then we consider the distribution of $U_n(Y)$:

$$\begin{aligned}
P(U_n(Y) \leq x) &= P\left(Y \leq \left(\frac{1}{1-F_n}\right)(x)\right) \\
&= P\left(Y \leq \left(\frac{1}{1-F_n}\right)(x) | F_n\right) \\
&= \Phi_{\text{Pareto}}\left(\frac{1}{1-F_n(x)}\right) \\
&= F_n(x).
\end{aligned}$$

The first step is an application of the switching formula. As Y and F_n are independent, $Y|F_n$ is just Pareto distributed and the second step follows. We conclude that $U_n(Y) \stackrel{d}{=} X^*$ with X^* drawn from the empirical distribution function and thus a bootstrap draw from the random sample.

Therefore, we may draw from F_n by drawing Y^* (independently) from a Pareto distribution and applying U_n to Y^*. Consequently, if $Y_1^*, \ldots, Y_n^*$ are i.i.d. random variables from the standard Pareto distribution which are also independent from $X_1, X_2, \ldots, X_n$, then $X_1^* := U_n(Y_1^*), \ldots, X_n^* := U_n(Y_n^*)$ are i.i.d. random variables from the empirical distribution function by the independence between Y^* and F_n. Additionally, $X^*_{n-[kx],n} = U_n\left(Y^*_{n-[kx],n}\right)$ for $x \in [0,1]$ since U_n preserves the ordering. U_n captures the randomness in the empirical distribution function and the Pareto draws capture the randomness due to the resampling.

In case the functional T depends directly on the data, i.e. no further operations on the data are necessary to calculate T, the previous approach suffices. However, in our block maxima approach we encounter functionals T which depend on the block maxima of a data set. Assume $\tilde{X}_1, \dots, \tilde{X}_n$ are i.i.d. random variables from distribution F. Define the block maxima X_i for $i = 1, \dots, k$ by

$$X_i = \max_{(i-1)m < j \leq im} \tilde{X}_j$$

with $\frac{n}{k} = m \in \mathbb{Z}_{>0}$ the block length. For the bootstrapping of the block maxima we do not resample from the block maxima, instead we resample from the entire random sample and construct bootstrapped block maxima from there. We define a function V and its empirical equivalent V_n as follows:

$$V(z) = F^{\leftarrow}\left(\Phi_{\mathrm{Fr}}(z)\right) \quad \text{and} \quad V_n(z) = F_n^{\leftarrow}\left(\Phi_{\mathrm{Fr}}(z)\right)$$

with $\Phi_{\mathrm{Fr}}(z) = \exp(-1/z)$ and Z follows a standard Fréchet distribution. Assume Z is independent from $\tilde{X}_1, \dots, \tilde{X}_n$ and consider $V_n(Z)$

$$\begin{aligned}
P\left(V_n(Z) \leq x\right) &= P\left(F_n^{\leftarrow}\left(\Phi_{\mathrm{Fr}}(Z)\right) \leq x\right) \\
&= P\left(\Phi_{\mathrm{Fr}}(Z) \leq F_n\left(x\right)\right) \\
&= P\left(Z \leq \Phi_{\mathrm{Fr}}^{-1}\left(F_n\left(x\right)\right) | F_n\right) \\
&= \Phi_{\mathrm{Fr}}\left(\Phi_{\mathrm{Fr}}^{-1}\left(F_n\left(x\right)\right)\right) \\
&= F_n(x).
\end{aligned}$$

The second step is again an application of the switching formula and the third step is by the independence of F_n and Z. It follows that $\tilde{X}^* \stackrel{d}{=} V_n(Z)$ with $\tilde{X}^*$ a draw from F_n. A bootstrapped block maximum is of the form $X^* = \max_{1 \leq j \leq m} \tilde{X}_j^*$. For such an X^* we find the following:

$$X^* = \max_{1 \leq j \leq m} \tilde{X}_j^* \stackrel{d}{=} \max_{1 \leq j \leq m} V_n(Z_j) = V_n\left(\max_{1 \leq j \leq m} Z_j\right) \stackrel{d}{=} V_n\left(mZ_0\right). \tag{2.5.2}$$

Here we use the fact that V_n is order preservering and the Fréchet distribution is max stable. We conclude that $X^* \stackrel{d}{=} V_n(mZ)$. Hence, we may draw bootstrapped block maxima by drawing Z^* from a Fréchet distribution, multiplying by m and applying V_n. Thus, if $Z_1^*, \dots, Z_k^*$ are i.i.d. random variables from the standard Fréchet distribution which are also independent from $\tilde{X}_1, \dots, \tilde{X}_n$, then $X_1^* := V_n(mZ_1^*), \dots, X_k^* := V_n(mZ_k^*)$ is a bootstrap sample of block maxima with $X^*_{\lceil ks \rceil, k} \stackrel{d}{=} V_n\left(mZ^*_{\lceil ks \rceil, k}\right)$ for $s \in (0, 1]$. In this case the randomness inherent to the empirical distribution function is captured by V_n and the randomness due to the resampling is caught by the Fréchet draws.

3 Bootstrapping the Hill estimator

This section is dedicated to bootstrapping the Hill estimator (Hill, 1975) for the extreme value index. The goal is to determine the asymptotic behaviour of a bootstrapped Hill estimator. From here, we show heuristically that the sample variance of bootstrapped Hill estimators is a consistent estimator for the theoretical variance of the initial Hill estimator. In this case the bootstrap procedure provides a reliable estimate for the Hill estimator's variance. We start with the definition of the Hill estimator in Section 3.1 and we prove the required result involving the sample variance of the bootstrapped estimators in Section 3.2. Subsequently, we apply the results from Section 3.2 in the context of the Weissman extreme quantile estimator in Section 3.3.

3.1 The Hill estimator

Suppose $\gamma > 0$ and $F \in \mathcal{D}(G_\gamma)$, then extreme value theory gives us that

$$\lim_{t\to\infty} \frac{\int_t^\infty (\log u - \log t) dF(u)}{1 - F(t)} = \gamma.$$

Now let $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ be the n-th order statistics from a random sample with distribution function F. We replace t and F by their empirical variant, that is the intermediate order statistic $X_{n-k,n}$ and the empirical cumulative distribution function F_n and consequently obtain the Hill estimator $\hat{\gamma}_H$,

$$\hat{\gamma}_H := \frac{\int_{X_{n-k,n}}^\infty (\log u - \log X_{n-k,n}) dF_n(u)}{1 - F_n(X_{n-k,n})} \tag{3.1.1}$$

or equivalently,

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}. \tag{3.1.2}$$

The following theorem gives the limiting distribution of the Hill estimator.

Theorem 3.1 (de Haan and Ferreira, 2006, Theorem 3.2.5)**.** *Suppose F satisfies the second-order condition for some $\gamma > 0$ and $\rho \leq 0$ and $\lim_{t\to\infty} A(t) = 0$ (Equation 2.3.8). Then*

$$\sqrt{k}\,(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right) \tag{3.1.3}$$

with N normal, provided $k(n) \to \infty, k(n)/n \to 0$ as $n \to \infty$ and

$$\lim_{n\to\infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda < \infty.$$

The latter theorem implies that $\hat{\gamma}_H$ is a consistent estimator for γ and $k \cdot \text{var}\,(\hat{\gamma}_H) \xrightarrow{d} \gamma^2$. As a result, it is trivial to estimate the variance of an estimator $\hat{\gamma}_H$ and the need for bootstrapping the Hill estimator is minimal. However, the associated proof is illustrative and provides insight and an appropriate start.

3.2 Asymptotic sample variance of the bootstrapped Hill estimator

In this subsection we present the main result of Section 3 in the form of Theorem 3.3. The crucial intermediate result concerning the limiting properties of bootstrapped intermediate order statistics deserves its own theorem and can be found beneath.

Theorem 3.2. *Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. random variables from distribution F. Assume that F satisfies the second-order condition for some $\gamma > 0$ and $\rho \leq 0$ (Equation 2.3.8). Suppose $X_1^*, X_2^*, \ldots, X_n^*$ is a bootstrapped sample, that is a random sample from distribution F_n, then we may define independent sequences of Brownian motions $\{W_n(s)\}_{s\geq 0}$ and $\{W_n^*(s)\}_{\geq 0}$ such that for a suitable function A_0,*

$$\frac{\log X^*_{n-[kx],n} - \log U(\frac{n}{k})}{\gamma} = -\log(x) + \frac{W_n^*(x) + W_n(x)}{\sqrt{kx}} + A_0\left(\frac{n}{k}\right)\frac{x^{-\rho}-1}{\gamma\rho} + c(x)o_p\left(\frac{1}{\sqrt{k}}\right) \tag{3.2.1}$$

uniformly for $x \in [k^{-1}, 1]$ as $n \to \infty$, provided $k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A_0\left(\frac{n}{k}\right) = O(1)$ with $c(x)$ a non-stochastic function integrable on $(0,1]$.

Proof. We start by recalling the definition of U and U_n

$$U = \left(\frac{1}{1-F}\right)^{\leftarrow} \quad \text{and} \quad U_n = \left(\frac{1}{1-F_n}\right)^{\leftarrow}.$$

Remark that $F_n^{\leftarrow}(x) = X_{\lceil nx \rceil, n}$ and it follows that

$$U_n\left(\frac{n}{kx}\right) = \left(\frac{1}{1-F_n}\right)^{\leftarrow}\left(\frac{n}{kx}\right) = F_n^{\leftarrow}\left(1 - \frac{kx}{n}\right) = X_{\lceil n-kx \rceil, n} = X_{n-[kx],n}. \tag{3.2.2}$$

We invoke Theorem 2.3 and Equation 3.2.2 to get that: There exists a sequence of Brownian motions $\{W_n(s)\}_{s\geq 0}$ and a function A_0 such that, for $\varepsilon > 0$ sufficiently small,

$$\frac{\log U_n(\frac{n}{ks}) - \log U(\frac{n}{k})}{\gamma} = -\log s + \frac{W_n(s)}{\sqrt{ks}} + A_0\left(\frac{n}{k}\right)\frac{1}{\gamma}\frac{s^{-\rho}-1}{\rho} + s^{-1/2-\varepsilon}o_p\left(\frac{1}{\sqrt{k}}\right) \tag{3.2.3}$$

uniformly for $s \in (0,1]$ as $n \to \infty$, provided $k(n) \to \infty$, $k(n)/n \to 0$ and $\sqrt{k}A_0(n/k) = O(1)$.

In Equation 3.2.3 we would like to substitute s with $s(x) = \frac{n}{kY^*_{n-[kx],n}}$. In doing so we would find, by the reasoning in Section 2.5,

$$U_n\left(\frac{n}{ks(x)}\right) = U_n\left(Y^*_{n-[kx],n}\right) \overset{d}{=} X^*_{n-[kx],n}.$$

We do have to justify this substitution as $s(x)$ is no longer non-stochastic. Let us rewrite Equation 3.2.3: for all $\varepsilon' > 0$

$$\lim_{n\to\infty} P\left(\left|\sup_{s\in(0,1]}\frac{|f(s)|}{c_0(s)}\right| > \varepsilon'\right) = 0$$

with $c_0(s) = s^{-1/2-\varepsilon}$ and

$$f(s) := \sqrt{k}\left(\frac{\log U_n(\frac{n}{ks}) - \log U(\frac{n}{k})}{\gamma} + \log s\right) - \frac{W_n(s)}{s} - \sqrt{k}A_0\left(\frac{n}{k}\right)\frac{1}{\gamma}\frac{s^{-\rho}-1}{\rho}.$$

Note that $f(s)$ actually depends on k and n. Now denote the event ($\forall x \in (0,1]$ holds $s(x) \in (0,1]$) by A. Hence follows by the law of total probability that

$$P\left(\left|\sup_{x\in(0,1]} \frac{|f(s(x))|}{c_0(s(x))}\right| > \varepsilon'\right) = P\left(\left|\sup_{x\in(0,1]} \frac{|f(s(x))|}{c_0(s(x))}\right| > \varepsilon' \bigcap A\right) + P\left(\left|\sup_{x\in(0,1]} \frac{|f(s(x))|}{c_0(s(x))}\right| > \varepsilon' \bigcap A^c\right)$$
$$= \underbrace{P\left(\left|\sup_{x\in(0,1]} \frac{|f(s(x))|}{c_0(s(x))}\right| > \varepsilon' | A\right)}_{\Pi} P(A) + P\left(\left|\sup_{x\in(0,1]} \frac{|f(s(x))|}{c_0(s(x))}\right| > \varepsilon' | A^c\right) P(A^c).$$

By Equation 3.2.3, $\Pi \to 0$ as $n \to \infty$ and what remains to be shown is that $P(A) \to 1$, or equivalently $P(A^c) \to 0$, as $n \to \infty$. If this is the case we get that

$$\lim_{n\to\infty} P\left(\left|\sup_{x\in(0,1]} \frac{|f(s(x))|}{c_0(s(x))}\right| > \varepsilon'\right) = 0$$

which implies that the substitution is justified. Suppose $P(A) \leq 1-\delta$ with $\delta > 0$ for a sequence $n_i \to \infty$ as $i \to \infty$. Then there exists an $x_0 \in (0,1]$ such that $s(x_0) > 1$ has a positive probability. In particular this implies that for sufficiently small ε_0

$$P\left(\left|s(x_0) - \frac{1}{x_0}\right| > \varepsilon_0\right) > 0$$

for all n_i. Note that this violates Lemma 2.3 at $\xi = -1$ and $\varepsilon = 1/2$. Therefore we conclude $P(A) \to 1$ as $n \to \infty$. Consequently, we apply the substitution for $x \in (0,1]$ and find

$$\frac{\log X^*_{n-[kx],n} - \log U(\frac{n}{k})}{\gamma} =$$
$$\underbrace{\log \frac{kY^*_{n-[kx],n}}{n}}_{\Lambda.1} + \underbrace{\frac{\sqrt{k}Y^*_{n-[kx],n} W_n\left(s(x)\right)}{n}}_{\Lambda.2} + \frac{1}{\gamma} \underbrace{A_0\left(\frac{n}{k}\right) \frac{s(x)^{-\rho}-1}{\rho}}_{\Lambda.3} + \underbrace{s(x)^{-1/2-\varepsilon}}_{\Lambda.4} o_p\left(\frac{1}{\sqrt{k}}\right) \tag{3.2.4}$$

uniformly for $x \in (0,1]$. We show that the following holds for the four parts:

$$\Lambda.1 = -\log(x) + \frac{W^*_n(x)}{\sqrt{kx}} + c_1(x) o_p\left(\frac{1}{\sqrt{k}}\right)$$
$$\Lambda.2 = \frac{W_n\left(x\right)}{\sqrt{kx}} + c_2(x) o_P\left(\frac{1}{\sqrt{k}}\right)$$
$$\Lambda.3 = A\left(\frac{n}{k}\right) \frac{x^{-\rho}-1}{\rho} + c_3(x) o_P\left(\frac{1}{\sqrt{k}}\right)$$
$$\Lambda.4 = x^{1/2+\varepsilon} + c_4(x) o_P(1)$$

uniformly for $x \in [k^{-1}, 1]$ as $n \to \infty$ with $c_1(x)$, $c_2(x)$, $c_3(x)$ and $c_4(x)$ integrable functions on $[0,1]$ and $\{W_n(s)\}_{s\geq 0}$, $\{W^*_n(s)\}_{\geq 0}$ independent sequences of Brownian motions.

We start with $\Lambda.1$. The function U related to the Pareto distribution is the identity function. From here it is easy to show that its first-order extreme value index γ is equal to 1, simply take $a(t) = t$ in Equation 2.3.3. Strictly speaking the Pareto distribution does not satisfy the second-order condition as the right-hand side of Equation 2.3.5 becomes zero, which is excluded from the definition. However, this indicates that the Pareto distribution behaves unusually good and can be interpreted as $\rho = -\infty$. By applying de Haan and Ferreira (2006, Theorem 2.4.8) to the Pareto distribution, there exists a sequence of Brownian motions $\{W_n^*(x)\}_{x\geq 0}$ such that, for $\varepsilon_1 > 0$ sufficiently small,

$$\log \frac{kY^*_{n-[kx],n}}{n} = -\log(x) + \frac{W_n^*(x)}{\sqrt{kx}} + x^{-1/2-\varepsilon_1}\, o_p\left(\frac{1}{\sqrt{k}}\right)$$

uniformly for $x \in (0,1]$ as $n \to \infty$, provided $k(n) \to \infty$ and $k(n)/n \to 0$. Here we use that $\lim_{\rho\to-\infty} \frac{s^{-\rho}-1}{\rho} = 0$ for $|s| < 1$ and as a result there is no bias for the Pareto distribution. Furthermore, it is obvious that W_n^* and W_n are independent and $c_1(x) = x^{-1/2-\varepsilon_1}$ is integrable.

Now we focus on $\Lambda.2$. To continue our endeavours we need to work with the Brownian motion in $\Lambda.2$ and need an extra tool here: the modulus of continuity of a Brownian motion. Suppose W is Brownian motion, then almost surely,

$$\lim_{h\downarrow 0} \sup_{0\leq t\leq 1-h} \frac{|W(t+h) - W(t)|}{\sqrt{2h\log(1/h)}} = 1.$$

This implies that the sample paths of the Brownian motion admit, with probability one, for δ small enough and $\varepsilon_w > 0$ a non-stochastic modulus of continuity $w(\delta)$ defined as

$$w(\delta) = \sqrt{2\delta\log(1/\delta)(1+\varepsilon_w)}.$$

We note that for δ small enough

$$w(\delta) < \delta^{1/2-\varepsilon_w}$$

By Lemma 2.3 at $\xi = -1$ and with $\tau > 0$ we know that

$$\frac{n}{kY^*_{n-[kx],n}} = x + x^{1/2-\tau} o_P(1).$$

By the modulus of continuity of the Brownian motion we find the following identity almost surely

$$\sup_{0\leq x<1} \left| W_n\left(\frac{n}{kY^*_{n-[kx],n}}\right) - W_n(x) \right| < \left(x^{1/2-\tau} o_P(1)\right)^{1/2-\varepsilon_w} = x^{(1/2-\tau)(1/2-\varepsilon_w)} o_P(1).$$

Hence we find, assuming $\tau \leq 1/2$, that uniformly for $x \in (0,1)$

$$W_n\left(\frac{n}{kY^*_{n-[kx],n}}\right) = W_n(x) + c_{21}(x) o_P(1). \tag{3.2.5}$$

By Lemma 2.3 we have that uniformly for $x \in [k^{-1}, 1]$ we have

$$\frac{k}{n} Y^*_{n-[kx],n} = x^{-1} + c_{22}(x)\, o_P(1) \tag{3.2.6}$$

with $c_{22}(x) = x^{-3/2-\tau}$ for $\tau > 0$. Now we remark that

$$\sup_{k^{-1}\leq x\leq 1} \frac{|W_n(x)|}{c_{21}} = O_P(1) \quad \text{and} \quad \sup_{k^{-1}\leq x\leq 1} \frac{|x^{-1}|}{c_{22}} = O_P(1)$$

as x^{-1} is not stochastic and the path of a Brownian motion is stochastically bounded. Hence, by Lemma 2.2 we see that

$$\frac{k}{n} Y^*_{n-[kx],n} W_n \left(\frac{n}{k Y^*_{n-[kx],n}} \right) = x^{-1} W_n(x) + c_2(x) o_P(1)$$

with $c_2(x) = c_{21}(x) c_{22}(x)$, which is integrable. Subsequently, we find for $\Lambda.2$

$$\frac{\sqrt{k}}{n} Y^*_{n-[kx],n} W_n \left(\frac{n}{k Y^*_{n-[kx],n}} \right) = \frac{W_n(x)}{\sqrt{k} x} + c_2(x) o_P\left(\frac{1}{\sqrt{k}} \right) \tag{3.2.7}$$

uniformly for $x \in [k^{-1}, 1)$, as $n \to \infty$ with $W_n^*(s)$ and $W_n(s)$ independent.

For $\Lambda.3$ we employ Lemma 2.3 again to see that the following holds uniformly for $x \in [k^{-1}, 1]$

$$\frac{s(x)^{-\rho} - 1}{\rho} = \frac{x^{-\rho} - 1}{\rho} + c_3(x) o_P(1). \tag{3.2.8}$$

as $n \to \infty, k(n) \to \infty$ and $k/n \to 0$, with $c_3(x) = x^{-\rho - 1/2 - \tau}$. Finally, we note that $\sqrt{k} A_0(n/k) = O(1)$ and thus

$$\sqrt{k} A_0\left(\frac{n}{k}\right) \left(\frac{s(x)^{-\rho} - 1}{\rho} - \frac{x^{-\rho} - 1}{\rho} \right) = O(1) \cdot c_3(x) o_P(1) = c_3(x) o_P(1)$$

uniformly for $x \in [k^{-1}, 1]$ as $n \to \infty$, which implies

$$A_0\left(\frac{n}{k}\right) \left(\frac{s(x)^{-\rho} - 1}{\rho} - \frac{x^{-\rho} - 1}{\rho} \right) = c_3(x) o_P\left(\frac{1}{\sqrt{k}} \right).$$

And hence we may state

$$A_0\left(\frac{n}{k}\right) \frac{s(x)^{-\rho} - 1}{\rho} = A\left(\frac{n}{k}\right) \frac{x^{-\rho} - 1}{\rho} + c_3(x) o_P\left(\frac{1}{\sqrt{k}} \right). \tag{3.2.9}$$

uniformly for $x \in [k^{-1}, 1]$ as $n \to \infty$. This result gives us the requested representation for $\Lambda.3$.

Finally considering $\Lambda.4$; in a very similar fashion as $\Lambda.3$ we find

$$s(x)^{-1/2-\varepsilon} = x^{1/2+\varepsilon} + c_4(x) o_P(1) \tag{3.2.10}$$

uniformly for $x \in [k^{-1}, 1]$ as $n \to \infty$, $k(n) \to \infty$ and $k/n \to 0$, where $c_4(x)$ is an integrable function.

We combine the expressions for $\Lambda.1$, $\Lambda.2$, $\Lambda.3$ and $\Lambda.4$, implicitly using Lemma 2.1, and we ultimately find

$$\frac{\log X^*_{n-[kx],n} - \log U(\frac{n}{k})}{\gamma} = -\log(x) + \frac{W_n^*(x) + W_n(x)}{\sqrt{k} x} + A_0\left(\frac{n}{k}\right) \frac{x^{-\rho} - 1}{\gamma \rho} + c(x) o_p\left(\frac{1}{\sqrt{k}} \right) \tag{3.2.11}$$

with $c(x) = c_1(x) + c_2(x) + c_3(x) + x^{1/2+\varepsilon} + c_4(x)$, which is integrable. □

The limiting distribution of the bootstrapped intermediate order statistics closely resembles that of the intermediate order statistics as presented in Theorem 2.3. The bootstrapping apparently introduces an extra random term in the form of $\frac{W_n^*(x)}{\sqrt{kx}}$. This also hints towards greater applicability of the latter theorem than just our purposes. As results based on Theorem 2.3 can often be adjusted without much effort to similar results in the bootstrap context by using Theorem 3.2 instead of Theorem 2.3.

The following theorem is the main result of Section 3. In bootstrapping the Hill estimator we bootstrap the entire sample and not just the highest order statistics the estimator is based on.

Theorem 3.3. *Suppose $X_1, X_2, \dots, X_n$ are i.i.d. random variables from distribution F. Assume that F satisfies the second-order condition for some $\gamma > 0$ and $\rho \leq 0$ (Equation 2.3.8). Let $\hat{\gamma}_H^*$ be a bootstrapped Hill estimator, then*

$$\frac{\hat{\gamma}_H^*}{\gamma} = 1 + \frac{1}{\sqrt{k}}(N^* + N) + A_0\left(\frac{n}{k}\right)\frac{1}{\gamma}\frac{1}{1-\rho} + o_p\left(\frac{1}{\sqrt{k}}\right) \tag{3.2.12}$$

as $n \to \infty$, provided $k(n) \to \infty$ and $k(n)/n \to 0$, with A_0 from Theorem 3.2, N^* and N independent normal random variables with mean 0 and variance 1.

Proof. The bootstrapped Hill estimator $\hat{\gamma}_H^*$ can be rewritten in the following way:

$$\begin{aligned}
\frac{\hat{\gamma}_H^*}{\gamma} &= \frac{1}{\gamma k}\sum_{i=0}^{k-1} \log X_{n-i,n}^* - \log X_{n-k,n}^* \\
&= \int_0^1 \frac{\log X_{n-[kx],n}^* - \log X_{n-k,n}^*}{\gamma}\,dx \\
&= \int_0^{k^{-1}} \frac{\log X_{n-[kx],n}^* - \log X_{n-k,n}^*}{\gamma}\,dx + \int_{k^{-1}}^1 \frac{\log X_{n-[kx],n}^* - \log X_{n-k,n}^*}{\gamma}\,dx \\
&:= \Delta_1 + \Delta_2.
\end{aligned}$$

Our aim is to show that $\Delta_1 = o_P\left(\frac{1}{\sqrt{k}}\right)$ and then use Theorem 3.2 to approximate Δ_2. The limiting distribution of $\hat{\gamma}_H^*$ will therefore be determined by Δ_2 as Δ_1 becomes small. First we remark that Δ_1 can be written in the following way:

$$\Delta_1 = \int_0^{k^{-1}} \frac{\log X_{n,n}^* - \log X_{n-k,n}^*}{\gamma}\,dx = \frac{1}{\gamma k}\left(\log X_{n,n}^* - \log X_{n-k,n}^*\right)$$

By Equation 3.2.1 at $x = 1$ we see that $|\log X_{n-k,n}^* - \log U\left(\frac{n}{k}\right)| \xrightarrow{P} 0$. Now we wish to show that

$$\frac{1}{\sqrt{k}}\left(\log X_{n,n}^* - \log U(n)\right) = o_P(1).$$

First of all, notice that $X_{n,n}^* \leq X_{n,n}$. We may show, following an approach similar to the one presented in Section 2.5, that $X_{n,n} \stackrel{d}{=} U\left(Y_{n,n}\right)$ with $Y_1, \dots, Y_n$ i.i.d. random variables from the standard Pareto distribution. This entails in particular

$$\frac{X_{n,n}^*}{U(n)} \leq \frac{X_{n,n}}{U(n)} \stackrel{d}{=} \frac{U\left(Y_{n,n}\right)}{U(n)}.$$

As U satisfies the first order condition with $\gamma > 0$ we may apply the Potter inequalities; for $\delta_1, \delta_2 > 0$ there exists $t_0 = t_0(\delta_1, \delta_2)$ such that for $n \geq t_0$ and $Y_{n,n} \geq t_0$,

$$\frac{U(Y_{n,n})}{U(n)} < (1+\delta_1) \cdot \left(\frac{Y_{n,n}}{n}\right)^{\gamma} \cdot \max\left(\left(\frac{Y_{n,n}}{n}\right)^{\delta_2}, \left(\frac{Y_{n,n}}{n}\right)^{-\delta_2}\right) := \Omega \tag{3.2.13}$$

and subsequently

$$\log U(Y_{n,n}) - \log U(n) < \log(\Omega). \tag{3.2.14}$$

Because $Y_{n,n} \geq t_0$ almost surely as $n \to \infty$ the latter inequalities are applicable in our situation.

The standard Pareto distribution satisfies Equation 2.3.1 with $a_n = b_n = n$ and $\gamma = 1$ which implies that $\frac{Y^*_{n,n}}{n} - 1 \xrightarrow{d} G_1$, with G_γ the extreme value distribution with parameter γ from Equation 2.3.2. As a result Ω follows a non-trivial distribution, as $n \to \infty$, and the same holds for $\log(\Omega)$. Combining the latter with Equation 3.2.14 we find $\log U(Y_{n,n}) - \log U(n) = O_P(1)$ and consequently

$$\frac{1}{\sqrt{k}}\left(\log X^*_{n,n} - \log U(n)\right) = o_p(1)$$

as $n, k \to \infty$. Using this information we may state

$$\begin{aligned} 0 \leq \Delta_1 =& \frac{1}{\gamma k}\left(\log X^*_{n,n} - \log U(n) - \log X^*_{n-k,n} + \log U(n/k)\right) + \frac{1}{\gamma k}\left(\log U(n) - \log U(n/k)\right) \\ =& o_P\left(\frac{1}{\sqrt{k}}\right) + o_P\left(\frac{1}{k}\right) + \frac{1}{\gamma k}\log\left(\frac{U(n)}{U(n/k)}\right) \\ =& \frac{1}{\gamma k}\log\left(\frac{U(n)}{U(n/k)}\right) + o_P\left(\frac{1}{\sqrt{k}}\right). \end{aligned}$$

Again, by the Potter inequalities, we have for any $\varepsilon, \delta > 0$

$$\frac{U(n)}{U\left(\frac{n}{k}\right)} < (1+\varepsilon) \cdot k^{\gamma+\delta}$$

as $n, k \to \infty$ in such a way that $n/k \to \infty$. Using the latter inequality we find for Δ_1:

$$\begin{aligned} \Delta_1 =& \frac{1}{\gamma k}\log\left(\frac{U(n)}{U(n/k)}\right) + o_P\left(\frac{1}{\sqrt{k}}\right) \\ <& \frac{1}{\gamma k}\log\left((1+\varepsilon) \cdot k^{\gamma+\delta}\right) + o_P\left(\frac{1}{\sqrt{k}}\right) \\ =& \frac{\gamma+\delta}{\gamma} \cdot \frac{\log k}{k} + \frac{\log(1+\varepsilon)}{\gamma k} + o_P\left(\frac{1}{\sqrt{k}}\right) \\ =& o_P\left(\frac{1}{\sqrt{k}}\right). \end{aligned}$$

Here, we use the fact that $\frac{1}{\sqrt{k}}\log k \to 0$ as $k \to \infty$. This entails in particular

$$\Delta_2 = \frac{\hat{\gamma}^*_H}{\gamma} + o_p\left(\frac{1}{\sqrt{k}}\right). \tag{3.2.15}$$

For Δ_2 we take Equation 3.2.1 and subtract the same equation at $x = 1$, which gives the following uniform convergence on $[k^{-1}, 1]$

$$\begin{aligned} \frac{\log X^*_{n-[kx],n} - \log X^*_{n-k,n}}{\gamma} =& -\log(x) + \frac{x^{-1}W^*_n(x) - W^*_n(1)}{\sqrt{k}} + \frac{x^{-1}W_n(x) - W_n(1)}{\sqrt{k}} \\ &+ A_0\left(\frac{n}{k}\right)\left(\frac{x^{-\rho}-1}{\gamma\rho}\right) + c(x)o_p\left(\frac{1}{\sqrt{k}}\right). \end{aligned} \tag{3.2.16}$$

We integrate both sides of this equation with respect to x from $t = k^{-1}$ to $t = 1$. The uniform convergence on this interval gives that the integral of the left-hand side and the right-hand side of the previous equation are equal. We find

$$\Delta_2 = \int_{k^{-1}}^{1} -\log(x) + \frac{x^{-1}W_n^*(x) - W_n^*(1)}{\sqrt{k}} + \frac{x^{-1}W_n(x) - W_n(1)}{\sqrt{k}} + A_0\left(\frac{n}{k}\right)\left(\frac{x^{-\rho}-1}{\gamma\rho}\right) + c(x)o_p\left(\frac{1}{\sqrt{k}}\right)\,dx$$

as $n \to \infty, k(n) \to \infty$ and $k(n)/n \to 0$. We focus on the integration of the latter integral. Remark that, by the fundamental theorem of calculus,

$$\lim_{k\to\infty}\int_{k^{-1}}^{1} -\log(x)\,dx = \int_0^1 -\log x\,dx = 1$$

and

$$\lim_{k\to\infty}\int_{k^{-1}}^{1} \frac{x^{-\rho}-1}{\gamma\rho}\,dx = \int_0^1 \frac{x^{-\rho}-1}{\gamma\rho}\,dx = \frac{1}{\gamma\rho}\cdot\left(\frac{x^{-\rho+1}}{-\rho+1} - x\right)\Bigg|_0^1 = \frac{1}{\gamma}\frac{1}{1-\rho},$$

thus it remains to determine the stochastic part of the integral. We consider

$$\lim_{k\to\infty}\int_{k^{-1}}^{1} x^{-1}W_n(x) - W_n(1)\,dx = \int_0^1 x^{-1}W_n(x) - W_n(1)\,dx$$

and show that this quantity follows a standard normal distribution. Viewed as a Riemann sum, this integral is the limit of normally distributed random variables and therefore itself normally distributed. As we may interchange the integral and the expectation operator the entire integral has expectation 0. Now for the variance:

$$\begin{aligned}
&\operatorname{var}\left(\int_0^1 x^{-1}W_n(x) - W_n(1)\,dx\right)\\
&= E\left[\left(\int_0^1 x^{-1}W_n(x)\,dx - W_n(1)\right)\left(\int_0^1 y^{-1}W_n(y)\,dy - W_n(1)\right)\right]\\
&= E\left[W_n^2(1)\right] + 2\int_0^1\int_x^1 E[W_n(x)W_n(y)]\frac{dy}{y}\frac{dx}{x} - 2\int_0^1 E\left[W_n(1)W_n(x)\right]\frac{dx}{x}\\
&= 1 + 2\int_0^1\int_x^1 x\frac{dy}{y}\frac{dx}{x} - 2\int_0^1 x\frac{dx}{x}\\
&= 1 + 2\int_0^1 -\log x\,dx - 2\\
&= 1
\end{aligned}$$

Here we use

$$\int_0^1\int_0^1 E\left[W_n(x)W_n(y)\right]\frac{dy}{y}\frac{dx}{x} = 2\int_0^1\int_x^1 E\left[W_n(x)W_n(y)\right]\frac{dy}{y}\frac{dx}{x}$$

and

$$E\left[W_n(1)W_n(x)\right] = E\left[W_n(x)(W_n(x) + (W_n(1) - W_n(x)))\right] = E[W_n(x)^2] = x.$$

We conclude that $\int_0^1 x^{-1}W_n(x) - W_n(1)\,dx$ and $\int_0^1 x^{-1}W_n^*(x) - W_n^*(1)\,dx$ are normally distributed random variables with mean 0 and variance 1. We remark that $\int_0^1 c(x)\,dx \cdot o_P(1/\sqrt{k}) =$

$o_P(1/\sqrt{k})$ as the integral is finite. Finally, we obtain by Equation 3.2.15

$$\frac{\hat{\gamma}_H^*}{\gamma} = 1 + \frac{1}{\sqrt{k}}(N^* + N) + A_0\left(\frac{n}{k}\right)\frac{1}{\gamma}\frac{1}{1-\rho} + o_p\left(\frac{1}{\sqrt{k}}\right) \tag{3.2.17}$$

as $k(n) \to \infty, k(n)/n \to 0, n \to \infty$, with N^* and N independent normal random variables with mean 0 and variance 1. □

As we showed in the previous theorem, a single bootstrapped estimator $\hat{\gamma}_{j,H}^*$ can be approximated using Equation 3.2.12. Assuming m bootstrapped estimators $\hat{\gamma}_{j,H}^*$, the following holds for the average $\overline{\gamma}_H^*$

$$\sqrt{k}\cdot\frac{\overline{\gamma}_H^*}{\gamma} \approx \sqrt{k} + \left(\overline{N}^* + N\right) + \sqrt{k}A_0\left(\frac{n}{k}\right)\frac{1}{\gamma}\frac{1}{1-\rho}$$

with $\overline{\gamma}_H^* = \frac{1}{m}\sum_{j=1}^m \hat{\gamma}_{j,H}^*$ and $\overline{N}^* = \frac{1}{m}\sum_{j=1}^m N_j^*$. It follows that

$$\sqrt{k}\left(\hat{\gamma}_{j,H}^* - \overline{\gamma}_H^*\right) \approx \gamma\left(N_j^* - \overline{N}^*\right)$$

as $n \to \infty, k(n) \to \infty$ and $k/n \to 0$. Finally, we are in the position to calculate the variance of the bootstrapped Hill estimators:

$$\frac{k}{m}\sum_{j=1}^m\left(\hat{\gamma}_{j,H}^* - \overline{\gamma}_H^*\right)^2 \approx \gamma^2\frac{1}{m}\sum_{j=1}^m\left(N_j^* - \overline{N}^*\right)^2$$

By the law of large numbers $\frac{1}{m}\sum_{j=1}^m\left(N_j^* - \overline{N}^*\right)^2 \xrightarrow{p} \text{var}(N_j^*) = 1$ as $m \to \infty$ and therefore

$$\frac{k}{m}\sum_{j=1}^m\left(\hat{\gamma}_{j,H}^* - \overline{\gamma}_H^*\right)^2 \approx \gamma^2 \tag{3.2.18}$$

As $m \to \infty, k \to \infty, n \to \infty$.

Deliberately, we employ an heuristic argument here and do not proceed in the rigorous way presented before. This is due to a caveat in the argument above. In order to substantiate statements about the average $\overline{\gamma}_H^*$ we actually need the result in Equation 3.2.12 to hold uniformly for $j = 1, \ldots, m$ bootstrapped Hill estimators. This is actually the case but lies outside the scope of this thesis.

3.3 The Weissman extreme quantile estimator

The result in the previous subsection has an elegant application in the context of extreme quantile estimation. As one can imagine, estimators of extreme quantiles usually depend on the extreme value index. The variance of such an extreme quantile estimator can often be linked to the variance of the extreme value index estimator. Suppose we use the Hill estimator to estimate the extreme value index, then we know we may employ bootstrapping methods to approximate the variance of the Hill estimator. And subsequently we can deduce the variance of the extreme quantile estimator from there.

We consider an extreme quantile estimator proposed in Weissman (1978) and since we wish to use the Hill estimator we restrict ourselves to positive γ. We want to estimate the $(1-p)$-quantile of F for p near 0, that is $x_p = F^{\leftarrow}(1-p) = U(1/p)$. This quantile should be an extreme quantile, in the sense that there should be very few observations to the right of this quantile. To deduce asymptotic statements about such an estimator we require that n, the number of observations, goes to infinity. If p would be fixed we would eventually find a numerous amount of observations to the right of x_p which is not desirable. Hence p should depend on n in such a way that (at least) $\lim_{n\to\infty} p_n \to 0$.

Theorem 3.4 (de Haan and Ferreira, 2006, Theorem 4.3.8)**.** *Suppose $X_1, X_2, \dots, X_n$ are i.i.d. random variables from distribution F. Assume that F satisfies the second-order condition for some $\gamma > 0$ and $\rho \leq 0$ (Equation 2.3.8) such that $A(t) \to 0$, as $n \to \infty$. Additionally assume:*

1. *$\hat{\gamma}$ is an estimator for the extreme value index and $\sqrt{k}\,(\hat{\gamma} - \gamma) \xrightarrow{d} \Gamma$ as $n \to \infty$ with Γ normally distributed with known mean depending on γ and ρ and variance only depending on γ,*
2. *$k \to \infty$, $k/n \to 0$, and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$, $n \to \infty$,*
3. *$np_n = o(k)$ and $\log np_n = o\left(\sqrt{k}\right)$, $n \to \infty$.*

Let $\hat{x}_{p_n}$ be the Weissman extreme quantile estimator defined as

$$\hat{x}_{p_n} = X_{n-k,n} \left(\frac{k}{np_n}\right)^{\hat{\gamma}} \quad \text{and} \quad x_{p_n} = U\left(\frac{1}{p_n}\right).$$

Then as $n \to \infty$,

$$\frac{\sqrt{k}}{\log d_n} \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1\right) \xrightarrow{d} \Gamma \tag{3.3.1}$$

with $d_n = k/(np_n)$.

The latter theorem particularly implies

$$\frac{k}{\left(x_{p_n} \log d_n\right)^2} \operatorname{var}\left(\hat{x}_{p_n}\right) \to \operatorname{var}(\Gamma) \tag{3.3.2}$$

which gives a method of determining the variance of $\hat{x}_{p_n}$ assuming we know (an estimate of) $\operatorname{var}(\Gamma)$. Suppose we use the Hill estimator as estimator for γ, which meets the requirements for the extreme value estimator as stated above. Then, by bootstrapping the Hill estimator we approximate $\operatorname{var}(\Gamma)$, from which we can deduce $\operatorname{var}\left(\hat{x}_{p_n}\right)$. Finally we notice that this procedure is not restricted to the Hill estimator and has much wider applicability.

4 Bootstrapping the probability weighted moment estimator

In Section 3 we considered the Hill estimator for the extreme value index based on the peaks over threshold method. We will now investigate the properties of the bootstrapped probability weighted moment (PWM) estimator for the extreme value index using a block maxima approach. When we refer to the PWM estimator we mean the PWM estimator for the extreme value index. Again, our aim is to determine the asymptotic behaviour of a bootstrapped PWM estimator. Subsequently, we set up an heuristic argument to show that the sample variance of bootstrapped PWM estimators is a consistent estimator for the theoretical variance of the PWM estimator. The theoretical variance of the PWM estimator is not easy to calculate as it involves multiple stochastic integrals. Therefore, it would be of practical use to show that the sample variance of the bootstrapped PWM estimators consistently estimates the theoretical variance as this would give a straightforward way to approximate the variance of a PWM estimator. An introduction to the PWM estimator follows in Section 4.1, followed by the proof in Section 4.2. As in the previous section we conclude with a subsection on extreme quantile estimation in Section 4.3.

4.1 The probability weighted moment estimator for block maxima

Assume $\tilde{X}_1, \ldots, \tilde{X}_n$ are i.i.d. random variables from distribution F. Define the block maxima X_i for $i = 1, \ldots, k$ by

$$X_i = \max_{(i-1)m < j \leq im} \tilde{X}_j$$

with $\frac{n}{k} = m \in \mathbb{Z}_{>0}$ the block length. Let $X_{1,k}, \ldots, X_{k,k}$ be its order statistics and define $\beta_0 = \frac{1}{k} \sum_{i=1}^{k} X_{i,k}$ and

$$\beta_r = \frac{1}{k} \sum_{i=1}^{k} \frac{(i-1)\ldots(i-r)}{(k-1)\ldots(k-r)} X_{i,k}$$

for $r \in \mathbb{Z}_{>0}$, $r < k$. The PWM estimator (Hosking, Wallis, and Wood, 1985) for γ, denoted $\hat{\gamma}_{k,m}$, is defined as the solution of

$$\frac{3^{\hat{\gamma}_{k,m}} - 1}{2^{\hat{\gamma}_{k,m}} - 1} = \frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0}.$$

In addition we define estimators for the sequences a_n and b_n from the first-order condition:

$$\hat{a}_{k,m} = \frac{\hat{\gamma}_{k,m}}{2^{\hat{\gamma}_{k,m}} - 1} \frac{2\beta_1 - \beta_0}{\Gamma(1 - \hat{\gamma}_{k,m})}$$

$$\hat{b}_{k,m} = \beta_0 + \hat{a}_{k,m} \frac{1 - \Gamma(1 - \hat{\gamma}_{k,m})}{\hat{\gamma}_{k,m}},$$

with $\Gamma(x)$ the usual gamma function. The following theorem describes the asymptotic behaviour of the three estimators defined above.

Theorem 4.1 (Ferreira and de Haan, 2015, Theorem 2.3)**.** *Assume F satisfies the second-order condition for $\gamma < \frac{1}{2}$ and $\rho \in \mathbb{R}$. Assume $k(n) \to \infty$ and $m(n) \to \infty$ as $n \to \infty$ such that $\sqrt{k}A(m) \to \lambda \in \mathbb{R}$*

(with $A(m)$ as in Equation 2.3.5) then

$$\sqrt{k}\left(\hat{\gamma}_{k,m}-\gamma\right) \xrightarrow{d} \frac{1}{\Gamma(1-\gamma)}\left(\frac{\log 3}{1-3^{-\gamma}}-\frac{\log 2}{1-2^{-\gamma}}\right)^{-1}\left(\frac{\gamma}{3^{\gamma}-1}(Q_2-Q_0)-\frac{\gamma}{2^{\gamma}-1}(Q_1-Q_0)\right) := \Delta \tag{4.1.1}$$

$$\sqrt{k}\left(\frac{\hat{a}_{k,m}}{a_m}-1\right) \xrightarrow{d} \frac{\gamma}{(2^{\gamma}-1)\Gamma(1-\gamma)}(Q_1-Q_0)+\Delta\left(\frac{\log 2}{\gamma}\left(\frac{-\gamma}{1-2^{-\gamma}}+\frac{1}{\log 2}\right)+\frac{\Gamma'(1-\gamma)}{\Gamma(1-\gamma)}\right) =: \Lambda \tag{4.1.2}$$

$$\sqrt{k}\left(\frac{\hat{b}_{k,m}-b_m}{a_m}\right) \xrightarrow{d} Q_0+\frac{\gamma\Gamma'(1-\gamma)-1+\Gamma(1-\gamma)}{\gamma^2}\Delta+\frac{1-\Gamma(1-\gamma)}{\gamma}\Lambda =: \Xi \tag{4.1.3}$$

as $n \to \infty$. With

$$Q_r = (r+1)\int_0^1 s^{r-1}(-\log s)^{-1-\gamma}E(s)ds+\lambda I_r(\gamma,\rho),$$

$E(s)$ a Brownian bridge and

$$I_r(\gamma,\rho) = (r+1)\int_0^1 H_{\gamma,\rho}\left(\frac{1}{-\log s}\right)s^r\,ds.$$

From this theorem we can deduce that the estimator $\hat{\gamma}_{k,m}$ is biased due to the (non-stochastic) factor $\lambda I_r(\gamma,\rho)$. Note that Q_i is stochastic and quite involved and it is its presence that makes the theoretical variance hard to calculate.

4.2 Asymptotic sample variance of the bootstrapped PWM estimator

The following theorem is the crucial stochastic approximation which enables us to prove the main theorem of Section 4.2 in Theorem 4.3. The structure of this proof displays similarities with the proof presented in the previous section and we attempt to let the structures align as much as possible. In addition, we recall the definition of a bootstrapped sample of block maxima. We construct a bootstrapped sample of block maxima by resampling from the entire random sample and then calculating the associated block maxima.

In the following proof there is a subtlety related to whether $\gamma \geq -\frac{1}{2}$. In case $\gamma \geq -\frac{1}{2}$ we are in the relatively favorable situation. By introducing the B_0 term from Theorem 2.6 we are also able to handle $\gamma < -\frac{1}{2}$.

Theorem 4.2. *Let $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ be i.i.d. random variables from distribution F with block maxima $X_1, \ldots, X_k$ and block length $m := \frac{n}{k}$. Assume that F satisfies the second-order condition for some $\gamma \in \mathbb{R}$ and $\rho \leq 0$. Suppose $X_1^*, X_2^*, \ldots, X_k^*$ is a bootstrapped sample of block maxima, then we may define a sequence of Brownian motions $\{W_n(s)\}_{s\geq 0}$ and a sequence of Brownian bridges $\{E_k^*(s)\}_{s\geq 0}$, independent from the Brownian motions, such that for suitable functions a_0 and A_0,*

$$\sqrt{k}\left(\frac{X^*_{\lceil ks\rceil,k}-B_0(m)}{a_0(m)}-\frac{(-\log s)^{-\gamma}-1}{\gamma}\right) = \frac{E_k^*(s)}{s(-\log s)^{1+\gamma}}+\frac{W_n(-\log s)}{(-\log s)^{1+\gamma}}+\sqrt{k}A_0(m)\Psi_{\gamma,\rho}\left(\frac{1}{-\log s}\right)+c(s)o_P(1) \tag{4.2.1}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$ *as* $n \to \infty$*, provided* $k(n) \to \infty$*,* $n \geq k(n)^{\frac{2-\tau}{1-\tau}}$ *for each* n *with* $\tau \in \left(\frac{1}{2}, 1\right)$*,* $\sqrt{k} A_0\left(\frac{n}{k}\right) = O(1)$ *with* $c(s)$ *non-stochastic and integrable on* $[0,1]$ *and* B_0 *as defined in Theorem 2.6.*

Proof. We start by invoking Theorem 2.6; We may define a sequence of Brownian motions $\{W_n\}_{s\geq 0}$ such that for suitably chosen functions a_0 and A_0 and each $\varepsilon \in (0, 1/2)$,

$$\sqrt{k}\left(\frac{\tilde{X}_{n-\left[\frac{n}{m}\cdot x\right],n} - B_0(m)}{a_0(m)} - \frac{x^{-\gamma}-1}{\gamma}\right) = x^{-\gamma-1} W_n(x) + \sqrt{k} A_0(m) \Psi_{\gamma,\rho}(x^{-1}) + h(x)^{-1} o_P(1) \tag{4.2.2}$$

uniformly for $x \in (0, \lambda(n)]$ as $n \to \infty$, provided $k = k(n) \to \infty$, $k/n \to 0$, $\sqrt{k} A(m) = O(1)$ and with $\lambda(n) = O\left((n/k)^{1-\tau}\right)$ for $\tau > 0$ such that $\tau > \frac{\varepsilon}{\varepsilon+1/2}$, with $h(x)$ defined as follows:

$$h(x) := \min\left(x^{\gamma+1/2+\varepsilon}, x^{\gamma+\rho-\varepsilon}\right).$$

Also note that for positive x the multiplicative inverse of $h(x)$ is given by

$$h(x)^{-1} = \max\left(x^{-\gamma-1/2-\varepsilon}, x^{-\gamma-\rho+\varepsilon}\right).$$

As we wish to gain information about the limiting distribution of bootstrapped block maxima, the need arises to substitute x in Equation 4.2.2 with a suitably chosen function. First we motivate our substitution function after which we justify the substitution.

By definition of the function V_n (see Section 2.5) we obtain $V_n(z) = \tilde{X}_{\lceil n\Phi(z)\rceil,n}$ and consequently

$$V_n(mz) = \tilde{X}_{\lceil n\Phi(mz)\rceil,n} = \tilde{X}_{n-\left[\frac{n}{m}m(1-\Phi(mz))\right],n}. \tag{4.2.3}$$

Subsequently, assume Z_i^*, for $i = 1, \ldots, k$ to be a random sample from the standard Fréchet distribution, then we know by the reasoning in Section 2.5 that

$$V_n\left(mZ^*_{\lceil ks\rceil,k}\right) \stackrel{d}{=} X^*_{\lceil ks\rceil,k}, \quad \text{for} \quad s \in (0,1]. \tag{4.2.4}$$

Hence we aim to substitute x with $x(z(s)) = m\left(1 - \Phi(mz(s))\right)$ with $z(s) = Z^*_{\lceil ks\rceil,k}$ for $s \in (0,1]$, resulting in

$$\tilde{X}_{n-\left[\frac{n}{m}\cdot x(z(s))\right],n} = \tilde{X}_{n-\left[\frac{n}{m}m(1-\Phi(mz(s)))\right],n} = V_n\left(mZ^*_{\lceil ks\rceil,k}\right) \stackrel{d}{=} X^*_{\lceil ks\rceil,k} \quad \text{for} \quad s \in (0,1] \tag{4.2.5}$$

with $X^*_{\lceil ks\rceil,k}$ the order statistics of bootstrapped block maxima.

For the justification we expand $\Phi(mz)$ around $\frac{1}{mz} = 0$, yielding

$$\Phi(mz) = \exp\left(-\frac{1}{mx}\right) = 1 - \frac{1}{mz} + \frac{1}{2m^2z^2} + \ldots,$$

and therefore

$$m\left(1 - \Phi(mz)\right) = \frac{1}{z} - \frac{1}{2mz^2} + \ldots.$$

Assuming $\frac{1}{z} \leq O\left((n/k)^{1-\tau}\right)$ with $\tau \in (1/2, 1)$ we see that $x(z) = m\left(1 - \Phi(mz)\right) \to \frac{1}{z}$ since $\frac{1}{mz^2} \to 0$. The assumption that $\frac{1}{z} \leq O\left((n/k)^{1-\tau}\right)$ is also required in the next step of the proof. Hence the only new requirement is that τ must be in $(1/2, 1)$.

By the continuity of the functions x^γ, $W_n(x)$ and $\Psi_{\gamma,\rho}(x)$ we henceforth find

$$\sqrt{k}\left(\frac{V_n(mz)-B_0\left(m\right)}{a_0\left(m\right)}-\frac{z^\gamma-1}{\gamma}\right)=z^{\gamma+1}W_n\left(\frac{1}{z}\right)+\sqrt{k}A_0\left(m\right)\Psi_{\gamma,\rho}(z)+h(z^{-1})^{-1}o_P(1) \quad (4.2.6)$$

uniformly for $1/z \in (0,\lambda(n)]$ as $n \to \infty$ and with $\lambda(n) = O\left((n/k)^{1-\tau}\right)$. In particular, notice that

$$0<\frac{1}{z}\leq O\left((n/k)^{1-\tau}\right) \iff O\left((k/n)^{1-\tau}\right)\leq z<\infty. \quad (4.2.7)$$

Now we wish to substitute z with $z(s) = Z^*_{\lceil ks\rceil,k}$ for $s \in (0,1]$ with Z^* standard Fréchet distributed. In doing so we need to satisfy Equation 4.2.7, which restricts the speed of which $Z_{1,k}$ may go to zero and therefore the growth of k. We require $P\left(Z^*_{1,k} > M\left(k/n\right)^{1-\tau}\right) = 1$ for fixed $M > 0$ as $k, n \to \infty$. Consider the following:

$$\begin{aligned}
P\left(Z^*_{1,k}\leq\frac{M}{k}\right)&=P\left(-\max\left(-Z^*_1,\dots,-Z^*_k\right)\leq\frac{M}{k}\right)\\
&=1-P\left(\max\left(-Z^*_1,\dots,-Z^*_k\right)\leq-\frac{M}{k}\right)\\
&=1-P\left(-Z^*\leq-\frac{M}{k}\right)^k\\
&=1-\left(1-\exp\left(-\frac{k}{M}\right)\right)^k.
\end{aligned}$$

For $M > 0$ we see that

$$\lim_{k\to\infty}\left(1-\exp\left(-\frac{k}{M}\right)\right)^k=1. \quad (4.2.8)$$

Hence follows that $P\left(Z^*_{1,k}\leq\frac{M}{k}\right)\to 0$ as $k\to\infty$ and consequently $P\left(Z^*_{1,k}>\frac{M}{k}\right)\to 1$. Thus it suffices to have $\frac{1}{k}\geq(k/n)^{1-\tau}$, which may be achieved by assuming $n \geq k(n)^{\frac{2-\tau}{1-\tau}}$. We apply the substitution in order to find

$$\sqrt{k}\left(\frac{X^*_{\lceil ks\rceil,k}-B_0(m)}{a_0(m)}\right)=$$

$$\underbrace{\sqrt{k}\frac{\left(Z^*_{\lceil ks\rceil,k}\right)^\gamma-1}{\gamma}}_{\Lambda.1}+\underbrace{\left(Z^*_{\lceil ks\rceil,k}\right)^{\gamma+1}W_n\left(\frac{1}{Z^*_{\lceil ks\rceil,k}}\right)}_{\Lambda.2}+\underbrace{\sqrt{k}A_0(m)\Psi_{\gamma,\rho}\left(Z^*_{\lceil ks\rceil,k}\right)}_{\Lambda.3}+\underbrace{h\left(z(s)^{-1}\right)^{-1}}_{\Lambda.4}o_P\left(1\right). \quad (4.2.9)$$

uniformly for $s \in (0,1]$ as $n \to \infty$. We show that the following holds

$$\begin{aligned}
\Lambda.1&=\sqrt{k}\frac{(-\log s)^{-\gamma}-1}{\gamma}+\frac{E^*_k(s)}{s(-\log s)^{1+\gamma}}+c_1(s)o_P\left(1\right)\\
\Lambda.2&=\frac{W_n\left(-\log s\right)}{(-\log s)^{\gamma+1}}+c_2(s)o_P(1)\\
\Lambda.3&=\sqrt{k}A_0(m)\Psi_{\gamma,\rho}\left(\frac{1}{-\log s}\right)+c_3(s)o_P\left(1\right)\\
\Lambda.4&=h\left(-\log s\right)^{-1}+c_4(s)o_P\left(1\right)
\end{aligned}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$ as $n \to \infty$, with the functions $c_i(s)$ positive and integrable on $[0,1]$ and $\{E_k^*(s)\}_{s\geq 0}$ a sequence of Brownian bridges, independent from the sequence $\{W_n(s)\}_{s\geq 0}$.

We start with Λ.1.; By Theorem 2.7 there exists, for $\nu \in (0, 1/2)$, an appropriate sequence of Brownian bridges $(E_k^*)_{k\geq 1}$ such that

$$\sqrt{k}\frac{\left(Z^*_{\lceil ks\rceil,k}\right)^{\gamma} - 1}{\gamma} = \sqrt{k}\frac{(-\log s)^{-\gamma} - 1}{\gamma} + \frac{E_k^*(s)}{s(-\log s)^{1+\gamma}} + c_1(s) o_P(1) \tag{4.2.10}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$ as $k \to \infty$ with $c_1(s) = \frac{(s(1-s))^{\nu}}{s(-\log s)^{1+\gamma}}$.

For the remaining approximations we need an intermediate result. Again, we start with the statement from Theorem 2.7; for $\xi \in \mathbb{R}$ we have

$$\sup_{1/(k+1)\leq s\leq k/(k+1)} (s(1-s))^{-\nu} \left| s(-\log s)^{1+\xi}\left(\frac{Z^{\xi}_{\lceil ks\rceil,k} - 1}{\xi} - \frac{(-\log s)^{-\xi} - 1}{\xi}\right) - \frac{E_k(s)}{\sqrt{k}} \right| = o_P\left(\frac{1}{\sqrt{k}}\right). \tag{4.2.11}$$

We see that

$$\begin{aligned}\sup_{1/(k+1)\leq s\leq k/(k+1)} (s(1-s))^{-\nu} \left|\frac{E_k(s)}{\sqrt{k}}\right| &\leq \sup_s \left(\frac{1}{k+1}\cdot\frac{k}{k+1}\right)^{-\nu} \left|\frac{E_k(s)}{\sqrt{k}}\right| \\ &= \sup_s (k+1)^{\nu}\cdot k^{-1/2}\cdot\left(\frac{k}{k+1}\right)^{-\nu} |E_k(s)| \\ &= o_P(1).\end{aligned}$$

In the last line we use that $\lim_{k\to\infty}(k+1)^{\nu}\cdot k^{-1/2} = 0$ for $\nu \in (0,1/2)$. Therefore, Equation 4.2.11 implies

$$\sup_{1/(k+1)\leq s\leq k/(k+1)} (s(1-s))^{-\nu} \left| s(-\log s)^{1+\xi}\left(\frac{Z^{\xi}_{\lceil ks\rceil,k} - 1}{\xi} - \frac{(-\log s)^{-\xi} - 1}{\xi}\right)\right| = o_P(1).$$

Hence we find for $\xi \in \mathbb{R}$ and $\nu \in (0, 1/2)$

$$Z^{\xi}_{\lceil ks\rceil,k} = (-\log s)^{-\xi} + c_{z,\xi}\cdot o_P(1) \tag{4.2.12}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$ as $k \to \infty$ with $c_{z,\xi} = \frac{(s(1-s))^{\nu}}{s(-\log s)^{1+\xi}}$.

Now we use Equation 4.2.12 at $\xi = -1$ and, as in the proof of Theorem 3.2, we employ the modulus of continuity of the Brownian motion to see that

$$W_n\left(\frac{1}{Z^*_{\lceil ks\rceil,k}}\right) = W_n(-\log s) + c_{22}(s) o_P(1). \tag{4.2.13}$$

We combine Equation 4.2.12 at $\xi = \gamma + 1$ and Equation 4.2.13 via Lemma 2.2 to find for Δ.2

$$\left(Z^*_{\lceil ks\rceil,k}\right)^{\gamma+1} W_n\left(\frac{1}{Z^*_{\lceil ks\rceil,k}}\right) = \left(\frac{1}{-\log s}\right)^{\gamma+1} W_n(-\log s) + c_2(s) o_P(1) \tag{4.2.14}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$ with $c_2(s) = c_{22}(s)c_{z,\gamma+1}(s)$, which is positive and integrable since both $c_{22}(s)$ and $c_{z,\gamma+1}(s)$ are.

For $\Lambda.3$ we recall the definition of $\Psi_{\gamma,\rho}(x)$:

$$\Psi_{\gamma,\rho}(x) = \begin{cases} \frac{x^{\gamma+\rho}-1}{\gamma+\rho}, & \gamma+\rho \neq 0,\ \rho < 0, \\ \log x, & \gamma+\rho = 0,\ \rho < 0, \\ \frac{1}{\gamma}x^{\gamma}\log x, & \rho = 0 \neq \gamma, \\ \frac{1}{2}(\log x)^2, & \rho = 0 = \gamma. \end{cases} \tag{4.2.15}$$

Our aim is to show that

$$\Psi_{\gamma,\rho}\left(Z^*_{\lceil ks\rceil,k}\right) = \Psi_{\gamma,\rho}\left(\frac{1}{-\log s}\right) + c_3(s)o_P(1) \tag{4.2.16}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$. The first case of Equation 4.2.15 is already done as it is an implication of Equation 4.2.12. For the second case we employ the Taylor expansion of $\log x$ at $x = 1$, which is given by

$$\log x = (x-1) - \frac{1}{2}(x-1)^2 + O(x^3).$$

Subsequently we define

$$\Upsilon := \sup_{1/(k+1)\leq s\leq k/(k+1)} \frac{c_{z,1}(s)^{-1}}{(-\log s)} \left|\log\left(Z^*_{\lceil ks\rceil,k}\right) - \log\left(\frac{1}{-\log(s)}\right)\right|.$$

Then we see the following for Υ by using the Taylor expansion of $\log x$

$$\begin{aligned} \Upsilon &= \sup_{1/(k+1)\leq s\leq k/(k+1)} \frac{c_z(s)^{-1}}{(-\log s)} \left|\log\left((-\log s)Z^*_{\lceil ks\rceil,k}\right)\right| \\ &= \sup_{1/(k+1)\leq s\leq k/(k+1)} \frac{c_z(s)^{-1}}{(-\log s)} \left|\left((-\log s)Z^*_{\lceil ks\rceil,k} - 1\right) + \dots\right| \\ &= o_P(1) \end{aligned}$$

which means that

$$\log\left(Z^*_{\lceil ks\rceil,k}\right) = \log\left(\frac{1}{-\log(s)}\right) + (-\log s)c_{z,1}(s)o_P(1) \tag{4.2.17}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$. The other two cases of Equation 4.2.15 can be deduced from the first two cases by means of Lemma 2.2. This shows Equation 4.2.16 with $c_3(s)$ positive, integrable and depending on ρ and γ. Since $\sqrt{k}A_0(n/k) = O(1)$ we find

$$\sqrt{k}A_0(n/k)\Psi_{\gamma,\rho}\left(Z^*_{\lceil ks\rceil,k}\right) = \sqrt{k}A_0(n/k)\Psi_{\gamma,\rho}\left(\frac{1}{-\log s}\right) + c_3(s)o_P(1) \tag{4.2.18}$$

uniformly for $s \in [1/(k+1), k/(k+1)]$ as $n \to \infty$, $k \to \infty$ and $k(n)/n \to 0$.

Finally, for $\Lambda.4$ we notice that

$$h\left(z(s)^{-1}\right)^{-1} = \max\left(z(s)^{\gamma+1/2+\varepsilon}, z(s)^{\gamma+\rho-\varepsilon}\right) \tag{4.2.19}$$

and hence Equation 4.2.12 implies the requested identity. By the expressions for $\Lambda.1$, $\Lambda.2$, $\Lambda.3$ and $\Lambda.4$ we find, implicitly using Lemma 2.1,

$$\sqrt{k}\frac{X^*_{\lceil ks\rceil,k} - B_0(m)}{a_0(m)} =$$
$$\sqrt{k}\frac{(-\log s)^{-\gamma}-1}{\gamma} + \frac{E^*_k(s)}{s(-\log s)^{1+\gamma}} + \frac{W_n(-\log s)}{(-\log s)^{\gamma+1}} + \sqrt{k}A_0(m)\Psi_{\gamma,\rho}\left(\frac{1}{-\log s}\right) + c(s)o_P(1) \tag{4.2.20}$$

uniformly for $1/k \le s \le k/(k+1)$ as $n \to \infty$, $k \to \infty$ and $k(n)/n \to 0$ with

$$c(s) = c_1(s) + c_2(s) + c_3(s) + h\left(-\log s\right)^{-1} + c_4(s).$$

□

Before continuing we clarify the meaning of a bootstrapped PWM estimator, denoted by $\hat{\gamma}^*_{k,m}$. Suppose $X^*_1, X^*_2, \ldots, X^*_k$ is a bootstrapped sample of block maxima then we may define β^*_r for $r \in \mathbb{Z}_{\ge 0}$ by $\beta^*_0 = \frac{1}{k}\sum_{i=1}^k X^*_{i,k}$ and

$$\beta^*_r = \frac{1}{k}\sum_{i=1}^k \frac{(i-1)\ldots(i-r)}{(k-1)\ldots(k-r)}X^*_{i,k}$$

for $r \in \mathbb{Z}_{>0}$, $r < k$. Then $\hat{\gamma}^*_{k,m}$ is defined as the solution of

$$\frac{3^{\hat{\gamma}^*_{k,m}}-1}{2^{\hat{\gamma}^*_{k,m}}-1} = \frac{3\beta^*_2 - \beta^*_0}{2\beta^*_1 - \beta^*_0}.$$

Theorem 4.3. *Suppose $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ are i.i.d. random variables from distribution F with block maxima $X_1, \ldots, X_k$ and block length m. Assume that F satisfies the second-order condition for some $\gamma < \frac{1}{2}$ and $\rho \le 0$ and $\sqrt{k}A(m) \to \lambda \in \mathbb{R}$ (with $A(m)$ as in Equation 2.3.5). Let $\hat{\gamma}^*_{k,m}$ be a bootstrapped PWM estimator, then*

$$\sqrt{k}\left(\hat{\gamma}^*_{k,m} - \gamma\right) = c_2(\gamma)(Q^*_2 - Q^*_0) - c_1(\gamma)(Q^*_1 - Q^*_0) + o_P(1) \tag{4.2.21}$$

as $n \to \infty$, provided $k(n) \to \infty$, $n \ge k(n)^{\frac{2-\tau}{1-\tau}}$ for each n with $\tau \in (1/2, 1)$ and with

$$c_i(\gamma) = \frac{1}{\Gamma(1-\gamma)}\left(\frac{\log 3}{1-3^{-\gamma}} - \frac{\log 2}{1-2^{-\gamma}}\right)^{-1}\left(\frac{\gamma}{(i+1)^\gamma - 1}\right) \quad i = 1, 2,$$

$$Q^*_r = (r+1)\int_0^1 s^r\left(\frac{E^*_k(s)}{s(-\log s)^{1+\gamma}} + \frac{W_n(-\log s)}{(-\log s)^{\gamma+1}}\right)ds + \lambda I^*_r(\gamma,\rho)$$

and

$$I^*_r(\gamma,\rho) = (r+1)\int_0^1 \Psi_{\gamma,\rho}\left(\frac{1}{-\log s}\right)s^r\,ds.$$

Proof. We are now in a position to invoke Theorem 4.2 and note the resemblance between this theorem and Ferreira and de Haan (2015, Theorem 2.1). The idea is to use the arguments made in Ferreira and de Haan (2015) to formulate analogous corollaries, with respect to Ferreira and de Haan (2015, Theorem 2.1), of Theorem 4.2. Although the similarities between the theorems are great, we focus on the differences and their consequences for the analogous corollaries.

First of all, Ferreira and de Haan (2015, Theorem 2.1) assumes that $\sqrt{k}A(m) \to \lambda \in \mathbb{R}$. For us it is also necessary to make this assumption as it limits growth of the number of blocks $k(n)$. Next we find a $\Psi_{\gamma,\rho}$ in our theorem where the original theorem shows a $H_{\gamma,\rho}$. This gives a slightly different definition of $I_r(\gamma,\rho)$ compared to the original proof. Subsequently, the use of $B_0(m)$ in our theorem in comparison to the use of b_m in the original theorem does give an extra term $\Pi(m)$ in the approximation of β_r^* due to the difference $B_0(m) - b_m$.

Finally, the most striking difference is the additional term $\frac{W_n(-\log s)}{(-\log s)^{\gamma+1}}$ due to the randomness introduced by the bootstrapping procedure. Considering all this, the analogous version of Ferreira and de Haan (2015, Theorem 2.2) is

$$\sqrt{k}\left(\frac{(r+1)\beta_r^* - b_m}{a_m} - D_r(\gamma)\right) = Q_r^* + \sqrt{k}\Pi(m) + o_P(1)$$

as $n \to \infty$, jointly for $r = 0, 1, 2, \ldots$ with $\Pi(m) = \frac{B_0(m)-b_m}{a_m}$, B_0 as defined in Theorem 2.6 and a_m, b_m as defined in Ferreira and de Haan (2015, Theorem 2.1). The β_r^* denotes the bootstrapped version of β_r with

$$D_r(\xi) = \frac{(r+1)^{\xi}\Gamma(1-\xi) - 1}{\xi}, \quad \xi < 1$$

and $\Gamma(x)$ the usual gamma function. For the interested reader we added the proof of the latter statement in the appendix, Theorem A.5. This proof is a copy of the proof presented in Ferreira and de Haan (2015) adjusted to fit our situation and therefore we do not present it to be our proof. The extra factor $\sqrt{k}\Pi(m)$ drops out at the next step of the proof as we are interested in the following quantities:

$$\sqrt{k}\left(\frac{2\beta_1^* - \beta_0^*}{a_m} - \frac{2^{\gamma}-1}{\gamma}\Gamma(1-\gamma)\right) = Q_1^* - Q_0^* + o_P(1)$$

and

$$\sqrt{k}\left(\frac{3\beta_2^* - \beta_0^*}{a_m} - \frac{3^{\gamma}-1}{\gamma}\Gamma(1-\gamma)\right) = Q_2^* - Q_0^* + o_P(1).$$

Following the exact lines of the proof for Ferreira and de Haan (2015, Theorem 2.3) we find for the bootstrapped PWM estimator $\hat{\gamma}_{k,m}^*$ that

$$\sqrt{k}\left(\hat{\gamma}_{k,m}^* - \gamma\right) = c_2(\gamma)(Q_2^* - Q_0^*) - c_1(\gamma)(Q_1^* - Q_0^*) + o_P(1) \tag{4.2.22}$$

as $n \to \infty$, provided $k(n) \to \infty$, $n \geq k(n)^{\frac{2-\tau}{1-\tau}}$ for each n with $\tau \in (1/2, 1)$. □

As for the Hill estimator, we explain heuristically why the bootstrap enables us to estimate the variance of the PWM estimator. Let $\hat{\gamma}_{k,m}^{*,l}$ be bootstrapped PWM estimators for $l = 1, \ldots, t$, and define $\overline{\gamma}_{k,m}^* = \frac{1}{t}\sum_{l=1}^{t} \hat{\gamma}_{k,m}^{*,l}$. For the average $\overline{\gamma}_{k,m}^*$ we find

$$\sqrt{k}\left(\overline{\gamma}_{k,m}^* - \gamma\right) \approx a_0(\gamma)\overline{Q}_0^* + a_1(\gamma)\overline{Q}_1^* + a_2(\gamma)\overline{Q}_2^*$$

with $a_0(\gamma) = c_1(\gamma) - c_2(\gamma)$, $a_1(\gamma) = c_1(\gamma)$, $a_2(\gamma) = c_2(\gamma)$ and

$$\overline{Q}_r^* = (r+1)\int_0^1 s^r\left(\frac{\overline{E}_k^*(s)}{s(-\log s)^{1+\gamma}} + \frac{W_n(-\log s)}{(-\log s)^{\gamma+1}}\right) ds + \lambda I_r^*(\gamma,\rho).$$

We may state the latter result because $W_n(-\log(s))$ and $I_r^*(\gamma,\rho)$ do not depend on a particular bootstrap draw and thus we do not need to take the average over this part. As a result

$$\sqrt{k}(\hat{\gamma}^{*,l}_{k,m} - \overline{\gamma^*}_{k,m}) \approx a_0(\gamma)Q_0^{\dagger,*} + a_1(\gamma)Q_1^{\dagger,*} + a_2(\gamma)Q_2^{\dagger,*}$$

with

$$Q_r^{\dagger,*} := Q_r^* - \overline{Q}_r^* = (r+1)\int_0^1 s^{r-1}\left(\frac{E_k^*(s) - \overline{E}_k^*(s)}{(-\log s)^{1+\gamma}}\right) ds.$$

A Brownian bridge has mean zero and therefore $\overline{E}_k^*(s) \xrightarrow{p} 0$ as $t \to \infty$. As a consequence

$$Q_r^{\dagger,*} = Q_r^* - \overline{Q}_r^* \xrightarrow{P} (r+1)\int_0^1 s^{r-1}\frac{E_k(s)}{(-\log s)^{1+\gamma}}\, ds \tag{4.2.23}$$

$$= Q_r - E[Q_r] \tag{4.2.24}$$

for $r = 0,1,2$ as $t \to \infty$ with

$$E[Q_r] = E\left[(r+1)\int_0^1 s^{r-1}\frac{E_k(s)}{(-\log s)^{1+\gamma}}ds + \lambda I_r(\gamma,\rho)\right] = \lambda I_r(\gamma,\rho).$$

Define $Q_i^\dagger := Q_i - E[Q_i]$, then we may conclude

$$\begin{aligned} k\cdot\frac{1}{t}\sum_{l=1}^{t}\left(\hat{\gamma}^{*,l}_{k,m} - \overline{\gamma^*}_{k,m}\right)^2 &\approx E\left[\left(a_0(\gamma)Q_0^{\dagger,*} + a_1(\gamma)Q_1^{\dagger,*} + a_2(\gamma)Q_2^{\dagger,*}\right)^2\right] \\ &\xrightarrow{P} E\left[\left(a_0(\gamma)Q_0^{\dagger} + a_1(\gamma)Q_1^{\dagger} + a_2(\gamma)Q_2^{\dagger}\right)^2\right] \\ &\xrightarrow{d} \operatorname{var}\left(\sqrt{k}(\hat{\gamma}_{k,m} - \gamma)\right) \\ &= k\cdot \operatorname{var}(\hat{\gamma}_{k,m}). \end{aligned}$$

It follows that

$$\frac{1}{t}\sum_{l=1}^{t}\left(\hat{\gamma}^{*,l}_{k,m} - \overline{\gamma^*}_{k,m}\right)^2 \approx \operatorname{var}(\hat{\gamma}_{k,m}) \tag{4.2.25}$$

as $t \to \infty, n \to \infty, k(n) \to \infty, n \geq k(n)^{\frac{2-\tau}{1-\tau}}$.

4.3 The probability weighted moment extreme quantile estimator for block maxima

As for the Hill estimator, the blockmaxima and PWM context provides an estimator for extreme quantiles. For a quantile $x_{p_n} = F^{\leftarrow}(1-p_n)$ with p_n small, the PWM extreme quantile estimator is given by

$$\hat{x}_{p_n,k,m} = \hat{b}_{k,m} + \hat{a}_{k,m}\frac{(mp_n)^{-\hat{\gamma}_{k,m}} - 1}{\hat{\gamma}_{k,m}}$$

with $\hat{\gamma}_{k,m}$, $\hat{b}_{k,m}$ and $\hat{a}_{k,m}$ as defined in Section 4.1. The following theorem gives the limiting distribution of $\hat{x}_{p_n,k,m}$.

Theorem 4.4 (Ferreira and de Haan, 2015, Theorem 2.4). *Assume the conditions of Theorem 4.1 with $\rho < 0$ or $\rho = 0$ and $\gamma < 0$. In addition, suppose that the probabilities p_n satisfy*

$$\lim_{n\to\infty} mp_n = 0 \quad \textit{and} \quad \lim_{n\to\infty} \frac{\log(mp_n)}{\sqrt{k}} = 0.$$

Then

$$\sqrt{k}\frac{\left(\hat{x}_{p_n,k,m} - x_{p_n}\right)}{a_m q_\gamma(1/(mp_n))} \xrightarrow{d} \Delta + (\gamma_-)^2\,\Xi - \gamma_-\Lambda - \lambda\frac{\gamma_-}{\gamma_- + \rho} \tag{4.3.1}$$

as $n \to \infty$ with $\gamma_- = \min(0,\gamma)$ and $q_\gamma(t) = \int_1^t s^{-\gamma-1}\log s\, ds$ and Δ, Ξ as in Theorem 4.1.

As in the proof of Theorem 4.3 in the previous subsection we may formulate bootstrap variants of Equation 4.1.2 and Equation 4.1.3, replacing $\hat{a}_{k,m}$, $\hat{b}_{k,m}$ and the Q_i by their bootstrap equivalents $\hat{a}^*_{k,m}$, $\hat{b}^*_{k,m}$ and the Q^*_i. Additionally, the need arise to introduce a stochastic function $g_q(k,m)$. Consequently we find for a bootstrapped PWM extreme quantile estimator $\hat{x}^*_{p_n,k,m}$ that

$$\sqrt{k}\frac{\left(\hat{x}^*_{p_n,k,m} - x_{p_n}\right)}{a_m q_\gamma(1/(mp_n))} = \Delta^* + (\gamma_-)^2\,\Xi^* - \gamma_-\Lambda^* - \lambda\frac{\gamma_-}{\gamma_- + \rho} + g_q(k,m) + o_P(1) \tag{4.3.2}$$

as $n \to \infty$ with $g_q(k,m)$ stochastic but not depending on a bootstrap draw. The important observation here is that the right-hand side of the latter equation is linear in the Q^*_i. Hence by the argument for the PWM estimator on the previous page we find that bootstrapping also works for this estimator; Let $\hat{x}^{*,l}_{p_n,k,m}$ be bootstrapped PWM extreme quantile estimators for $l = 1,\ldots,t$, then

$$\frac{1}{t}\sum_{l=1}^{t}\left(\hat{x}^{*,l}_{p_n,k,m} - \overline{x^*}_{p_n,k,m}\right)^2 \approx \operatorname{var}\left(\hat{x}_{p_n,k,m}\right) \tag{4.3.3}$$

as $t \to \infty, n \to \infty, k(n) \to \infty, n \geq k(n)^{\frac{2-\tau}{1-\tau}}$ with $\overline{x^*}_{p_n,k,m} = \frac{1}{t}\sum_{l=1}^{t}\hat{x}^{*,l}_{p_n,k,m}$.

5 Simulation Results

In this section we demonstrate the presented theorems in a simulation exercise. We show by means of simulation that our bootstrapping schemes actually provide reliable estimates and compare the peaks over threshold method with the block maxima approach. Section 5.1 discusses the choice for the underlying distribution and Section 5.2 provides the results and interpretation.

5.1 Burr distribution

To ensure comparability between the peaks over threshold method and the block maxima approach we need to start with the same sample for both methods. This places restrictions on the underlying distribution. To use our bootstrapping theorems concerning the Hill estimator and successive Weissman estimator we need $\gamma > 0$ and $\rho \leq 0$ to hold for the underlying distribution. On the other hand we need $\gamma < \frac{1}{2}$ and $\rho < 0$ or $\gamma < 0$ and $\rho = 0$ to satisfy the conditions of the block maxima bootstrapping theorems. Hence the underlying distribution needs to satisfy $\gamma \in (0, 1/2)$ and $\rho < 0$. To meet these specific restrictions we choose to sample from the Burr distribution. The CDF of the Burr distribution has parameters $\beta > 0$, $\lambda > 0$, $\tau > 0$ and has the following definition

$$F(x) = 1 - \left(1 + \frac{x^{\tau}}{\beta}\right)^{-\lambda}, \quad \text{for } x > 0. \tag{5.1.1}$$

The related function $U = \left(\frac{1}{1-F}\right)^{\leftarrow}$ has the following form

$$U(x) = \beta^{\frac{1}{\tau}} \left(x^{\frac{1}{\lambda}} - 1\right)^{\frac{1}{\tau}}, \quad \text{for } x > 0, \tag{5.1.2}$$

and subsequently we find for the Burr(β, λ, τ) distribution

$$\gamma = \frac{1}{\lambda\tau} \quad \text{and} \quad \rho = -\frac{1}{\lambda}. \tag{5.1.3}$$

Hence we take $(\beta, \lambda, \tau) = (1, 1, 4)$ for the Burr distribution in order to have $\gamma = \frac{1}{4}$ and $\rho = -1$ and meet our conditions.

5.2 Results

Initially we produce a sample of n independent draws from the Burr$(1, 1, 4)$ distribution and fix k. This allows us to calculate the Hill estimator and the PWM estimator for the extreme value index. For the block maxima approach we use block length $m = n/k$ to get a sample of k block maxima. In this way, both the Hill estimator and the PWM estimator are calculated using k extreme observations. Consecutively we fix $p = 1/n$ and calculate the Weissman estimator (based on the Hill estimator) and the PWM extreme quantile estimator. In order for the quantile to be "extreme" we need to choose p this small, which guarantees that there will be only a few observations near the quantile x_p and we are (almost) outside the range of the data. For both the extreme value index γ and the extreme quantile x_p we have the theoretical value en we report

Table 1: Ratio of the estimator to its true value for different values of n and k, with n the total sample size and k the amount of extreme data points used to calculate the extreme value index estimators. For the Hill estimator and the PWM estimator the ratio $\hat{\gamma}/\gamma$ is presented and for the Weissman estimator and the PWM extreme quantile estimator the table depicts the value $\hat{x}_p/x_p$, with $p = 1/n$.

	Estimator ratios			
(n,k)	*Hill est.*	*Weissman est.*	*PWM est.*	*PWM quant. est*
(1000,50)	1.12	1.06	0.82	0.94
(1000,100)	1.08	1.02	1.17	0.99
(1000,150)	0.98	0.93	1.20	1.00
(5000,100)	0.92	0.87	0.74	0.78
(5000,200)	0.97	0.91	0.71	0.78
(5000,500)	0.99	0.93	0.76	0.80
(10000,200)	0.98	0.94	1.14	0.97
(10000,500)	0.98	0.94	0.94	0.90

the ratios of the estimator to the true value: $\hat{\gamma}/\gamma$ and $\hat{x}_p/x_p$. The results can be found in Table 1. We see that both the Hill estimator and the PWM estimator provide accurate estimates for the extreme value index, even for a sample of size $n = 1000$. However, the PWM estimator appears more sensitive to the choice of k than the Hill estimator. Additionally, both extreme quantile estimators seem to perform well.

Subsequently we apply t bootstrap procedures. For each bootstrap procedure we calculate the Hill estimator and both the PWM estimators. The empirical variance of the bootstrapped Hill estimators gives an estimate for the variance of the original Hill estimator. From this estimate for the variance of the Hill estimator we may deduce an estimate for the variance of the Weissman estimator $\hat{x}_p$ via:

$$\operatorname{var}\left(\hat{x}_p\right) \approx \operatorname{var}\left(\hat{\gamma}_H\right) \frac{\left(x_p \log d_n\right)^2}{k}$$

with $d_n = k/(np)$. We use the empirical variance of the bootstrapped Hill estimators (following Equation 3.2.18) as proxy for $\operatorname{var}\left(\hat{\gamma}_H\right)$ and use the Weissman estimator as proxy for x_p to produce an estimate for $\operatorname{var}\left(\hat{x}_p\right)$. For the PWM estimators we take the empirical variance of the bootstrapped estimators as an estimate for $\operatorname{var}\left(\hat{\gamma}_{k,m}\right)$ and $\operatorname{var}\left(\hat{x}_{p,k,m}\right)$. In all of these cases we denote the estimate of the variance (of a particular estimator) by $\widehat{\operatorname{var}}\left(\cdot\right)$ to distinguish from the theoretical variance, denoted by $\operatorname{var}\left(\cdot\right)$.

The next step is compare the estimate for a particular variance to the theoretical value. For the Hill estimator and the Weissman estimator (asymptotic) theoretical values are available as presented in Theorem 3.1 and Theorem 3.4. For the PWM estimators these are not available and therefore we employ a pre-simulation. We produce 10 000 samples, each sample consisting of n independent draws from the Burr$(1, 1, 4)$ distribution, and for each of these samples we calculate both PWM estimators with the corresponding parameters k and p. Subsequently, for

Table 2: Ratio of the variance estimated by bootstrapping to the theoretical variance, that is $\widehat{\text{var}}\left(\hat{\Theta}\right)/\text{var}\left(\hat{\Theta}\right)$, for the four estimators $\hat{\Theta}$ for different triples (n,k,t). With n the total sample size, k the amount of extreme data points used to calculate the extreme value index estimators and t the amount of bootstrapping procedures. For the extreme quantile estimators we use $p = 1/n$.

	Variance ratios			
(n,k,t)	*Hill est.*	*Weissman est.*	*PWM est.*	*PWM quant. est*
(1000,50,1000)	1.13	1.27	0.54	0.23
(1000,50,5000)	1.08	1.21	0.53	0.22
(1000,100,1000)	1.23	1.28	0.34	0.31
(1000,150,1000)	1.18	1.03	0.34	0.39
(5000,100,1000)	0.68	0.51	1.24	0.39
(5000,100,5000)	0.70	0.52	1.23	0.39
(5000,100,10000)	0.71	0.53	1.19	0.39
(5000,200,1000)	0.82	0.68	0.90	0.36
(5000,500,1000)	0.92	0.80	0.69	0.40
(10000,200,1000)	0.88	0.77	0.79	0.58
(10000,200,5000)	0.93	0.81	0.82	0.61
(10000,200,10000)	0.92	0.81	0.82	0.61
(10000,500,1000)	1.00	0.88	0.83	0.63

both PWM estimators, we determine the empirical variance of these 10 000 estimators and treat this empirical variance as the true variance. Hence we have produced for all four estimators a theoretical value of the variance. In Table 2 we depict the ratio of the estimated variance to the theoretical variance for the four estimators for different triples (n,k,t). The first observation we make is that the estimates for the variance hardly depend on the amount of bootstrap procedures. Even for a larger sample of 10000 observations the improvement in the estimate caused by increasing the amount of bootstrap procedures from 1000 to 10000 is minimal. This indicates that even for larger samples a relatively modest amount of bootstrapping procedures may be used. We do notice large differences in performance between different estimators. In general, we see two things; The estimated variances of the extreme value index estimators seem to be more accurate then the estimated variances of the extreme quantile estimators and the variances of the peaks over threshold estimators appear to be estimated more accurately than for the block maxima estimators under consideration. The accuracy of the variance estimates for the PWM estimators appears to improve relatively much by increasing n. This might indicate that for the PWM estimators the sample sizes presented here are actually too small to let the bootstrap perform optimally. Across the board, the bootstrapping procedure seems to provide estimates of the variance that are rather good (provided n is large enough), especially for the peaks over threshold estimators under consideration.

6 Empirical application

In this section we employ our bootstrapping methods in the context of financial data and in particular estimate the value at risk for a small number of stock exchange indices and estimate the corresponding variance.

The data consists of daily returns quoted as a percentage for several large stock exchange indices. This dataset is extracted from Bloomberg and the specific ticker names can be found in Table 3 along with their start date and end date.

Table 3: Daily return series quoted as a percentage for which we bootstrap estimators for the extreme value index and extreme quantiles.

Index (Bloomberg ticker)	*start date*	*end date*	*number of obs.*
AEX Index (AEX)	4-1-1983	31-8-2015	8304
DAX Index (DAX)	2-10-1959	31-8-2015	14058
Nikkei 225 (NKY)	6-1-1970	31-8-2015	11260
Shanghai Stock Exch. Composite Index (SHCOMP)	20-12-1990	31-8-2015	6041
S&P 500 Index (SPX)	3-1-1950	31-8-2015	16523
FTSE 100 Index (UKX)	4-1-1984	28-8-2015	8015

For each stock exchange index i we implicitly assume their daily returns are i.i.d. draws from a distribution F_i satisfying the second order extreme value condition with $\gamma \in (0, 1/2)$ and $\rho < 0$. We fix $k = 200$ and calculate the Hill estimator and PWM estimator, for each index i, to measure the tail heaviness of the underlying distribution F_i. The simulations indicate that the estimates are not sensitive to the choice of k and that $k = 200$ seems an appropriate choice for a sample of 5000 observations as well as for a sample of 10 000 observations. Therefore we use $k = 200$ for all indices. Furthermore, we produce for each stock exchange index i estimates for an extreme quantile. We choose $p = 0.0001 \approx 1/n$ and estimate $x_{p,i} = F_i^{\leftarrow}(1 - p)$ by means of the Weissman estimator and the PWM extreme quantile estimator. In other words, we estimate the daily return that is exceeded only by a chance of p. As for the simulations, this choice of p guarantees that the quantile is (almost) outside the scope of the data. Subsequently we employ 5000 bootstrap procedures and with each bootstrap procedure we calculate the Hill estimator and PWM estimators. We finally find an estimate for the variances of all four estimators as we did for the simulations.

We report the values of the Hill estimator and the PWM estimator along with the square root of the estimated variance in Table 4. We notice large differences between the indices with respect to their extreme value index. The Shanghai Stock Exchange composite index (SHCOMP) has the heaviest tail as opposed to the Nikkei 225 index. We also note substantial differences between the Hill estimator and PWM estimator, where the PWM estimator usually gives a lower estimate. On the other hand, the variance of the PWM estimator is always higher than the variance of the Hill estimator. The PWM estimate for the SHCOMP is near 0.5, which is close to our theoretical

Table 4: Values of the Hill estimator and PWM estimator for the stock exchange indices with corresponding square root of the estimated variance denoted by *Boot. Std.* For every estimator, 5000 bootstrap sequences are used.

	Hill estimator		**PWM estimator**	
Index	*estimate*	*Boot. Std.*	*estimate*	*Boot. Std.*
AEX	0.356	0.024	0.316	0.041
DAX	0.323	0.022	0.279	0.050
NKY	0.282	0.020	0.166	0.066
SHCOMP	0.483	0.034	0.497	0.106
SPX	0.323	0.020	0.216	0.065
UKX	0.318	0.023	0.263	0.057

Table 5: Quantile estimates of the daily returns distribution at $p = 0.0001$ for the stock indices based on the Weissman estimator and the PWM extreme quantile estimator with corresponding square root of the estimated variance denoted by *Boot. Std.* For every estimator, 5000 bootstrap sequences are used.

	Weissman estimator		**PWM quantile estimator**	
Index	*estimate (%)*	*Boot. Std. (%)*	*estimate (%)*	*Boot. Std. (%)*
AEX	18.38	2.54	14.15	1.54
DAX	14.41	1.63	11.96	1.23
NKY	12.65	1.31	11.36	1.67
SHCOMP	61.14	12.11	56.92	22.19
SPX	11.36	1.25	8.05	1.07
UKX	12.31	1.53	9.76	1.43

boundary. Hence the associated variance might not be accurate.

We report the values of the Weissman estimator and the PWM extreme quantile estimator along with the square root of the estimated variance in Table 5. We note that the Weissman estimates are higher than the PWM extreme quantile estimates, perhaps due to the higher Hill estimates relative to the PWM estimates. The estimated variances of the Weissman estimates are also higher, except for the SHCOMP and Nikkei 225 index. In alignment with our findings on the extreme value index we see that the extreme quantile for the SHCOMP is by far the largest.

So far we have investigated the properties of the right tail of F_i, the tail associated with high returns. However, economic theory suggests that we fear downside risk more than we appreciate upside potential. Hence the left tail and the Value at Risk (VaR) are of particular financial interest. To determine a VaR we wish to estimate $F^{\leftarrow}(p)$ for p very small. This presents us with a challenge as both of the extreme quantile estimators can only deal with quantiles of the form $U(1/p) = F^{\leftarrow}(1-p)$. To solve this we propose to flip the sign of the entire dataset and

Table 6: 1 day 0.01% Value at Risk estimates for the indices based on the Weissman estimator and the PWM extreme quantile estimator with corresponding square root of the estimated variance denoted by *Boot. Std.* For every estimator, 5000 bootstrap sequences are used.

	Weissman estimator		**PWM quantile estimator**	
Index	*estimate (%)*	*Boot. Std. (%)*	*estimate (%)*	*Boot. Std. (%)*
AEX	-20.92	3.01	-11.10	1.18
DAX	-14.88	1.70	-10.73	0.93
NKY	-14.52	1.62	-12.13	1.65
SHCOMP	-35.55	5.57	-19.93	1.67
SPX	-11.11	1.19	-10.07	1.80
UKX	-13.36	1.69	-9.78	1.75

then construct for this flipped dataset the estimates for the quantile $F^{\leftarrow}(1-p)$ which coincides (up to a sign change) with the requested $F^{\leftarrow}(p)$ quantile of the untouched dataset. In doing so we also have to re-estimate the extreme value index. For $p = 0.0001$, the calculated quantity can be interpreted as a 1 day 0.01% Value At Risk for the particular index. Table 6 contains the results of this procedure. We note some remarkable differences, primarily for the SHCOMP. The absolute values of the quantiles for the SHCOMP are much lower than the values presented in Table 5, indicating substantial differences between the extreme negative and extreme positive returns on this index. In particular there should be more extreme positive returns than negative ones. This is substantiated by lower estimates for the extreme value index for the flipped dataset. For the other indices the VaR values are in accordance with the results presented in Table 5, signalling similarities between the (extreme) left tail and (extreme) right tail of their underlying distribution.

7 Concluding remarks

This study investigates the use of bootstrapping schemes to estimate the variance of estimators from extreme value theory. The core of this thesis is devoted to the analysis of the bootstapped Hill estimator and the bootstrapped probability weighted moment estimator for the extreme value index. This analysis finds a natural application in extreme quantile estimation, and more specifically in the Weissman estimator and the probability weighted moment estimator for extreme quantiles.

The first step is to split the randomness in the bootstrapping procedure into two parts; one part due to the randomness of the empirical distribution function and one part due to resampling. Consecutively, we investigate the limiting distribution of the intermediate order statistics of a bootstrapped sample and bootstrapped block maxima. By employing this information about a bootstrapped sample, we show heuristically that the sample variance of bootstrapped Hill estimators is a consistent estimator for the theoretical variance of the original Hill estimator. It is worth noting that this statement about the limiting distribution of a bootstrapped sample has greater applicability because it indicates that the differences in the limiting distribution between a random sample and a corresponding bootstrapped sample are small. Corollaries based on the limiting distribution of a random sample can therefore often be adjusted to fit the bootstrap framework.

We follow exactly this reasoning for the PWM estimator. The theorem concerning the limiting distribution of the PWM estimator is based on a statement about the intermediate order statistics of block maxima. We prove a very similar statement about the intermediate order statistics of bootstrapped block maxima and then follow the lines of the original proof to construct the limiting distribution of a bootstrapped PWM estimator. From here we show heuristically that the sample variance of bootstrapped PWM estimators provides a consistent estimator for the theoretical variance of the original PWM estimator for the extreme value index.

Additionally, we expand our scope towards extreme quantile estimation. We may base the Weissman estimator on the Hill estimator. In this context is the variance of the Weissman estimator a function of the variance of the Hill estimator. Hence, by bootstrapping the Hill estimator we may estimate the variance of the Weissman estimator. The theory behind the limiting distribution of the PWM estimator for extreme quantiles is very similar to that of the PWM estimator for the extreme value index. Hence, we employ a very similar argument as for the PWM estimator for the extreme value index to find that bootstrapping the PWM extreme quantile estimator also provides the correct variance.

Finally, we perform a simulation exercise and a data analysis to review the practical use of our theory. The simulation exercise indicates that our bootstrapping schemes allow for accurate estimation of the variance of these estimators even for a relatively small sample of 1000 observations. Signalling that our bootstrapping schemes are applicable for real datasets.

Naturally, in both the peaks over threshold and block maxima context, there exist more estimators for the extreme value index and extreme quantiles than the ones considered in this thesis. Armed with our theorems concerning the limiting distribution of the intermediate order statistics of a bootstrapped sample and bootstrapped block maxima it would be reasonable to investigate whether bootstrapping also works to estimate their variance.

Furthermore, we showed heuristically that our bootstrapping schemes worked for our four estimators. In order to make such an argument completely rigorous we need the limiting distribution of a bootstrapped estimator uniform for any number of these bootstrapped estimators. To extend the presented theorems in such a way would present a substantial challenge and would pose an apparent subject for future research.

References

Drees, Holger (1998). On smooth statistical tail functionals. *Scandinavian Journal of Statistics 25*(1), 187–210.

Efron, B. (1979, 01). Bootstrap methods: Another look at the jackknife. *Ann. Statist. 7*(1), 1–26.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Ferreira, Ana and Laurens de Haan (2015, 02). On the block maxima method in extreme value theory: Pwm estimators. *Ann. Statist. 43*(1), 276–298.

de Haan, Laurens and Ana Ferreira (2006). *Extreme Value Theory: An Introduction*. New York: Springer-Verlag.

Hill, Bruce M. (1975, 09). A simple general approach to inference about the tail of a distribution. *Ann. Statist. 3*(5), 1163–1174.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics 27*(3), pp. 251–261.

Potter, H. S. A. (1942). The mean values of certain dirichlet series, ii. *Proceedings of the London Mathematical Society s2-47*(1), 1–19.

Serfling, Robert J. (2008). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc.

Weissman, Ishay (1978). Estimation of parameters and larger quantiles based on the k largest observations. *Journal of the American Statistical Association 73*(364), pp. 812–815.

Appendices

A Additional proofs

In the next theorems we extend results presented in de Haan and Ferreira (2006). The first four theorems all work towards an extension of de Haan and Ferreira (2006, Corollary 2.4.5) which is essential in Section 4. The proofs and their structure presented here rely heavily on the work of Laurens de Haan.

Theorem A.1 (de Haan and Ferreira, 2006, Extension of Lemma 2.4.10)**.** *Assume $Y_1, Y_2, \dots$ are i.i.d. random variables with a standard Pareto distribution and let $Y_{1,n} \leq Y_{2,n} \leq \dots Y_{n,n}$ be its n-th order statistics. For each $\gamma \in \mathbb{R}$ we may define a sequence of Brownian motions $\{W_n(s)\}_{s>0}$ such that for each $\varepsilon \in (0, 1/2)$,*

$$\sup_{k^{-1} \leq s \leq \lambda(n)} s^{\gamma+1/2+\varepsilon} \left| \sqrt{k} \left(\frac{\left(\frac{k}{n} Y_{n-[ks],n}\right)^{\gamma} - 1}{\gamma} - \frac{s^{-\gamma} - 1}{\gamma} \right) - s^{-\gamma-1} W_n(s) \right| = o_P(1) \tag{A.0.1}$$

as $n \to \infty$, $k(n) \to \infty$ and $k/n \to 0$ and with $\lambda(n) = O\left((n/k)^{1-\tau}\right)$ for $\tau > 0$ such that $\tau > \frac{\varepsilon}{\varepsilon + 1/2}$.

Proof. The sequence $\frac{Y_1^{\gamma}-1}{\gamma}, \frac{Y_2^{\gamma}-1}{\gamma}, \dots$ has distribution function $F(x) = 1 - (1+\gamma x)^{-\frac{1}{\gamma}}$ for which the conditions of de Haan and Ferreira (2006, Proposition 2.4.9) hold.

In general, let $f(x) = F'(x)$ and $Q(t) = F^{\leftarrow}(t)$. Suppose the conditions of de Haan and Ferreira (2006, Proposition 2.4.9) hold, then for $0 \leq \varepsilon < \frac{1}{2}$ we may define a sequence of Brownian bridges $\{B_n(t)\}$ such that

$$\sup_{1/(n+1) \leq t \leq n/(n+1)} n^{\varepsilon} t^{\varepsilon-1/2} (1-t)^{\varepsilon-1/2} \left| \sqrt{n} f(Q(t)) \left(Q(t) - X_{\lceil nt \rceil, n} \right) - B_n(t) \right| = O_P(1) \tag{A.0.2}$$

as $n \to \infty$ and $X_1, X_2, \dots$ i.i.d. random variables from distribution F.

In our case

$$\begin{aligned} F(x) &= 1 - (1+\gamma x)^{-\frac{1}{\gamma}} \\ f(x) &= (1+\gamma x)^{-\frac{\gamma+1}{\gamma}} \\ Q(t) &= \frac{(1-t)^{-\gamma} - 1}{\gamma}. \end{aligned}$$

By Equation A.0.2

$$\sup_{1/(n+1) \leq t \leq n/(n+1)} n^{\varepsilon} t^{\varepsilon-1/2} (1-t)^{\varepsilon-1/2} \left| \sqrt{n} (1-t)^{\gamma+1} \left(\frac{Y_{\lceil nt \rceil, n}^{\gamma} - 1}{\gamma} - \frac{(1-t)^{\gamma} - 1}{\gamma} \right) - B_n(t) \right| = O_P(1). \tag{A.0.3}$$

We interchanged $Q(t)$ and $X_{\lceil nt \rceil, n}$. This is justified as $B_n(t)$ and $-B_n(t)$ have the same probability. Now we wish to substitute t with $t(s) = 1 - \frac{ks}{n}$. The boundaries of $t(s)$ are given by

$1/(n+1) \leq t(s) \leq n/(n+1)$. This entails for s:

$$1/(n+1) \leq 1 - \frac{ks}{n} \leq n/(n+1) \tag{A.0.4}$$

$$1 - 1/(n+1) \geq \frac{ks}{n} \geq 1 - (n/n+1) \tag{A.0.5}$$

$$\frac{n}{k}\frac{n}{n+1} \geq s \geq \frac{n}{k}\frac{1}{n+1} \tag{A.0.6}$$

The following restrictions imply the latter restrictions and are therefore sufficient:

$$\frac{1}{k} \leq s \leq \frac{n}{k}\frac{n}{n+1}. \tag{A.0.7}$$

Additionally, we wish to have the following for $\varepsilon \in (0, 1/2)$

$$\sup_{1/k \leq s \leq \lambda(n)} \left(1 - \frac{ks}{n}\right)^{\varepsilon - 1/2} - 1 \to 0 \tag{A.0.8}$$

as $n \to \infty$. For $\frac{ks}{n} < 1$, which should be the case by the previous restrictions, the supremum is achieved at $s = \lambda(n)$. Hence it is necessary and sufficient to have $\lambda(n) \cdot \frac{k}{n} \to 0$ as $n \to \infty$. Which is the case for $\lambda(n) = O\left((n/k)^{1-\tau}\right)$.

After some rearrangements we get

$$\sup_{k^{-1} \leq s \leq \lambda(n)} k^{\varepsilon} \Bigg| \left(\frac{k}{n}\right)^{\gamma} s^{\gamma+1/2+\varepsilon} \sqrt{k} \left(\frac{Y_{n-[ks],n}^{\gamma} - 1}{\gamma} - \frac{\left(\frac{ks}{n}\right)^{-\gamma} - 1}{\gamma} \right) - \left(\frac{k}{n}\right)^{-1/2} s^{\varepsilon - 1/2} B_n \left(1 - \frac{ks}{n}\right) \Bigg| = O_p(1) \tag{A.0.9}$$

Now we follow the proof as given in de Haan and Ferreira (2006). The only additional result we need is

$$\sup_{k^{-1} \leq s \leq \lambda(n)} \left| \left(\frac{ks}{n}\right)^{1/2} s^{\varepsilon}\, W_n(1) \right| = o_p(1). \tag{A.0.10}$$

For this to hold, it suffices to have

$$\left(\frac{k}{n}\right)^{1/2} \lambda(n)^{\varepsilon + 1/2} \to 0 \quad \text{as} \quad n \to \infty. \tag{A.0.11}$$

For $\lambda(n) = O\left((n/k)^{1-\tau}\right)$ there exists $M > 0$ and $m_0 \in \mathbb{Z}_{>0}$ such that

$$\lambda(n) \leq M \cdot (n/k)^{1-\tau} \quad \text{for} \quad n > m_0. \tag{A.0.12}$$

Hence follows

$$\left(\frac{k}{n}\right)^{1/2} \lambda(n)^{\varepsilon+1/2} \leq \left(\frac{k}{n}\right)^{1/2} M^{\varepsilon+1/2} \cdot \left((n/k)^{1-\tau}\right)^{\varepsilon+1/2} \tag{A.0.13}$$

$$= \left(\frac{k}{n}\right)^{1/2} M^{\varepsilon+1/2} \cdot \left(\frac{k}{n}\right)^{-1/2-\varepsilon+\tau(\varepsilon+1/2)} \tag{A.0.14}$$

$$= \left(\frac{k}{n}\right)^{-\varepsilon+\tau(\varepsilon+1/2)} M^{\varepsilon+1/2}. \tag{A.0.15}$$

Hence it suffices to have $-\varepsilon + \tau(\varepsilon + 1/2) > 0$ or equivalently $\tau > \frac{\varepsilon}{\varepsilon+1/2}$. □

Theorem A.2 (de Haan and Ferreira, 2006, Extension of Lemma 2.4.11). *Let $X_1, X_2, \dots$ be i.i.d. random variables with distribution function F. Suppose $U = (1/(1-F))^{\leftarrow}$ satisfies von Mises' second order condition of de Haan and Ferreira, 2006, Theorem 2.3.12 for some $\gamma \in \mathbb{R}$ and $\rho \le 0$. Then we may define a sequence of Brownian motions $\{W_n(s)\}_{s>0}$ such that for suitably chosen functions a_0 and A_0 and each $\varepsilon \in (0, 1/2)$,*

$$\sup_{k^{-1} \le s \le \lambda(n)} \min\left(s^{\gamma+1/2+\varepsilon}, s^{\gamma+\rho-\varepsilon}\right) \left| \sqrt{k}\left(\frac{X_{n-[ks],n} - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} - \frac{s^{-\gamma}-1}{\gamma}\right) - s^{-\gamma-1}W_n(s) - \sqrt{k}A_0\left(\frac{n}{k}\right)\Psi_{\gamma,\rho}(s^{-1}) \right| \xrightarrow{P} 0 \quad \text{(A.0.16)}$$

as $n \to \infty$, provided $k = k(n) \to \infty$, $k/n \to 0$, $\sqrt{k}A_0(n/k) = O(1)$ and with $\lambda(n) = O\left((n/k)^{1-\tau}\right)$ for $\tau > 0$ such that $\tau > \frac{\varepsilon}{\varepsilon+1/2}$.

Proof. The function F satisfies the conditions of de Haan and Ferreira (2006, Proposition 2.4.9) for the right tail. We employ this proposition and use the same $\lambda(n)$, that is $\lambda(n) = O\left((n/k)^{1-\tau}\right)$, as in the previous proof to find

$$\sup_{k^{-1} \le s \le \lambda(n)} s^{1/2+\varepsilon} \left| \sqrt{k}\frac{X_{n-[ks],n} - U\left(\frac{n}{ks}\right)}{a\left(\frac{n}{ks}\right)} - \frac{W_n(s)}{s} \right| = o_p(1) \quad \text{(A.0.17)}$$

as $n \to \infty$ and with $a(t) = tU'(t)$. We take $a_0 \sim a$ with a_0 from de Haan and Ferreira (2006, Proposition 2.3.6).

For $s \in [k^{-1}, 1]$ the result is already proved in de Haan and Ferreira (2006), hence we proceed for $s \in (1, \lambda(n)]$. Note that for $s > 1$

$$\min\left(s^{\gamma+1/2+\varepsilon}, s^{\gamma+\rho-\varepsilon}\right) = s^{\gamma+\rho-\varepsilon} \quad \text{(A.0.18)}$$

because $\gamma + \rho - \varepsilon < \gamma + \frac{1}{2} + \varepsilon$. The expression from the theorem becomes

$$s^{\gamma+\rho-\varepsilon}\left\{\sqrt{k}\left(\frac{X_{n-[ks],n} - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} - \frac{s^{-\gamma}-1}{\gamma}\right) - s^{-\gamma-1}\sqrt{\frac{n}{k}}W_n\left(\frac{ks}{n}\right) - \sqrt{k}A_0\left(\frac{n}{k}\right)\Psi_{\gamma,\rho}\left(s^{-1}\right)\right\} \quad \text{(A.0.19)}$$

which can be rewritten towards

$$\stackrel{d}{=} s^{\gamma+\rho-\varepsilon}\left\{\sqrt{k}\left(\frac{U\left(\frac{n}{ks}\right) - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} - \frac{s^{-\gamma}-1}{\gamma}\right) - \sqrt{k}A_0\left(\frac{n}{k}\right)\Psi_{\gamma,\rho}\left(s^{-1}\right)\right\} \quad \text{(A.0.20)}$$

$$+ s^{\gamma+\rho-1/2-2\varepsilon}\frac{a_0\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{k}\right)}\left\{s^{1/2+\varepsilon}\left(\sqrt{k}\frac{X_{n-[ks],n} - U\left(\frac{n}{ks}\right)}{a\left(\frac{n}{ks}\right)} - \frac{W_n(s)}{s}\right)\right\} \quad \text{(A.0.21)}$$

$$+ s^{\gamma+\rho-\varepsilon}\left(\frac{a_0\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{k}\right)} - s^{-\gamma}\right)\frac{W_n(s)}{s} \quad \text{(A.0.22)}$$

where we use the identity

$$\frac{X_{n-[ks],n} - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} = \frac{U\left(\frac{n}{ks}\right) - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} + \frac{a_0\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{k}\right)}\left(\frac{X_{n-[ks],n} - U\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{ks}\right)}\right). \quad \text{(A.0.23)}$$

By de Haan and Ferreira (2006, Proposition 2.3.6) we see that for every $\varepsilon', \delta > 0$ the absolute value of Equation A.0.20 is bounded by:

$$s^{\gamma+\rho-\varepsilon}\sqrt{k} \cdot A_0\left(\frac{n}{k}\right) \varepsilon' s^{-\gamma-\rho} \max\left(s^{\delta}, s^{-\delta}\right). \tag{A.0.24}$$

Here, we implicitly use that $s \leq \lambda(n) = O\left((n/k)^{1-\tau}\right)$. Without loss of generality we drop $\sqrt{k}A_0\left(\frac{n}{k}\right)$ from the latter equation, as this term is bounded and such a bound can easily be incorporated by suitable choice of ε'. Now notice that

$$\max\left(s^{\delta}, s^{-\delta}\right) = \begin{cases} s^{\delta} & \text{if} \quad s > 1, \\ s^{-\delta} & \text{if} \quad s \leq 1. \end{cases} \tag{A.0.25}$$

Hence for $s > 1$ the upper bound is given by

$$\varepsilon' \cdot s^{-\varepsilon+\delta}, \tag{A.0.26}$$

subsequently choosing $\delta = \varepsilon$ gives that Equation A.0.20 converges uniformly to zero for $s \in (1, \lambda(n)]$. Now we consider Equation A.0.21 and immediately use Potter's inequality. For $\delta_1, \delta_2 > 0$

$$s^{\gamma+\rho-1/2-2\varepsilon} \cdot \frac{a_0\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{k}\right)} < s^{\gamma+\rho-1/2-2\varepsilon} \cdot (1+\delta_1)s^{-\alpha} \max\left(s^{\delta_2}, s^{-\delta_2}\right) \tag{A.0.27}$$

with $a_0 \in \mathrm{RV}_\alpha$. Here, we implicitly use that $s \leq \lambda(n) = O\left((n/k)^{1-\tau}\right)$. By theory on regular variation we know that $\alpha = \gamma$. Hence we find

$$s^{\gamma+\rho-1/2-2\varepsilon}\frac{a_0\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{k}\right)} < c_0 \cdot s^{\rho-1/2-2\varepsilon+\delta_2} \tag{A.0.28}$$

with c_0 a constant. Therefore, the latter quantity is bounded. By using Theorem A.1 it follows that Equation A.0.21 converges uniformly to zero for $s \in (1, \lambda(n)]$. For Equation A.0.22 we employ de Haan and Ferreira (2006, Proposition 2.3.6) to find

$$\left(s^{\gamma}\frac{a_0\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{k}\right)} - 1\right) < A_0\left(\frac{n}{k}\right)\left(\frac{s^{-\rho}-1}{\rho} + \varepsilon' s^{-\rho} \max\left(s^{\delta}, s^{-\delta}\right)\right). \tag{A.0.29}$$

Hence we find for $s > 1$

$$s^{\rho-\varepsilon} \cdot \left(s^{\gamma}\frac{a_0\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{k}\right)} - 1\right) < s^{\rho-\varepsilon} \cdot A_0\left(\frac{n}{k}\right)\left(\frac{s^{-\rho}-1}{\rho} + \varepsilon' s^{-\rho+\delta}\right) \tag{A.0.30}$$

$$= A_0\left(\frac{n}{k}\right)\left(\frac{s^{-\varepsilon}-s^{\rho-\varepsilon}}{\rho} + \varepsilon' s^{\delta-\varepsilon}\right). \tag{A.0.31}$$

As $A_0\left(\frac{n}{k}\right) = o(1)$ and $\left(\frac{s^{-\varepsilon}-s^{\rho-\varepsilon}}{\rho} + \varepsilon' s^{\delta-\varepsilon}\right)$ is bounded (by suitable choice of δ) for $s > 1$ we see that

$$s^{\rho-\varepsilon} \cdot \left(s^{\gamma}\frac{a_0\left(\frac{n}{ks}\right)}{a_0\left(\frac{n}{k}\right)} - 1\right) = o(1). \tag{A.0.32}$$

Finally, the paths of a Brownian motion are stochastically bounded and in particular

$$\sup_{1<s\leq\lambda(n)} \frac{W_n(s)}{s} < \infty \quad \text{almost surely.} \tag{A.0.33}$$

Therefore Equation A.0.22 also converges to zero for $s \in (1, \lambda(n)]$. □

Theorem A.3 (Drees, 1998, Extension of Theorem 2.1)**.** *Assume $X_1, X_2, \dots$ are i.i.d. random variables with distribution function F and let $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ be its n-th order statistics. Suppose that F satisfies the second-order condition for some $\gamma \in \mathbb{R}$ and $\rho \leq 0$ (de Haan and Ferreira, 2006, (2.3.21)). Then we may define a sequence of Brownian motions $\{W_n(s)\}_{s>0}$ such that for suitably chosen functions a_0 and A_0 and each $\varepsilon \in (0, 1/2)$,*

$$\sup_{k^{-1} \leq s \leq \lambda(n)} \min\left(s^{\gamma+1/2+\varepsilon}, s^{\gamma+\rho-\varepsilon}\right) \left| \sqrt{k} \left(\frac{X_{n-[ks],n} - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} - \frac{s^{-\gamma}-1}{\gamma} \right) \right.$$
$$\left. - s^{-\gamma-1} W_n(s) - \sqrt{k} A_0\left(\frac{n}{k}\right) \Psi_{\gamma,\rho}(s^{-1}) \right| \xrightarrow{P} 0 \quad \text{(A.0.34)}$$

as $n \to \infty$, provided $k = k(n) \to \infty$, $k/n \to 0$, $\sqrt{k}A_0(n/k) = O(1)$ and with $\lambda(n) = O\left((n/k)^{1-\tau}\right)$ for $\tau > 0$ such that $\tau > \frac{\varepsilon}{\varepsilon+1/2}$.

Proof. We may follow the exact lines of the proof presented in de Haan and Ferreira (2006). Instead of using de Haan and Ferreira (2006, Lemma 2.4.10) and de Haan and Ferreira (2006, Lemma 2.4.11) we need to employ Theorem A.1 and Theorem A.2. □

Theorem A.4 (de Haan and Ferreira, 2006, Extension of Corollary 2.4.5)**.** *Define*

$$B_0\left(\frac{n}{k}\right) := \begin{cases} U\left(\frac{n}{k}\right) & \text{if } \gamma \geq -\frac{1}{2} \\ X_{n,n} + \frac{a_0\left(\frac{n}{k}\right)}{\gamma} & \text{if } \gamma < -\frac{1}{2}. \end{cases} \quad \text{(A.0.35)}$$

Then, under the conditions of Theorem A.3,

$$\sup_{0 < s \leq \lambda(n)} \min\left(s^{\gamma+1/2+\varepsilon}, s^{\gamma+\rho-\varepsilon}\right) \left| \sqrt{k} \left(\frac{X_{n-[ks],n} - B_0\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} - \frac{s^{-\gamma}-1}{\gamma} \right) \right.$$
$$\left. - s^{-\gamma-1} W_n(s) - \sqrt{k} A_0\left(\frac{n}{k}\right) \Psi_{\gamma,\rho}(s^{-1}) \right| \xrightarrow{P} 0 \quad \text{(A.0.36)}$$

as $n \to \infty$ provided $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A_0(n/k) = O(1)$ and with $\lambda(n) = O\left((n/k)^{1-\tau}\right)$ for $\tau > 0$ such that $\tau > \frac{\varepsilon}{\varepsilon+1/2}$.

Proof. For $\gamma \geq -\frac{1}{2}$ we combine the original statement of de Haan and Ferreira (2006, Corollary 2.4.5) with Theorem A.3 to find the requested result. For $\gamma < -\frac{1}{2}$ the original theorem deals with $s \in (0, 1]$ and we only need to prove the statement for $s \in (1, \lambda(n)]$. Hence, by Theorem A.3 it suffices to show that

$$\sup_{1 < s \leq \lambda(n)} s^{\gamma+\rho-\varepsilon} \sqrt{k} \left| \frac{X_{n,n} - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} + \frac{1}{\gamma} \right| \xrightarrow{P} 0. \quad \text{(A.0.37)}$$

As $\gamma + \rho - \varepsilon < 0$ it follows that

$$\sup_{1 < s \leq \lambda(n)} s^{\gamma+\rho-\varepsilon} \sqrt{k} \left| \frac{X_{n,n} - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} + \frac{1}{\gamma} \right| \leq 1 \cdot \sqrt{k} \left| \frac{X_{n,n} - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} + \frac{1}{\gamma} \right| \quad \text{(A.0.38)}$$
$$\leq k^{-(\gamma+1/2+\varepsilon)} \cdot \sqrt{k} \left| \frac{X_{n,n} - U\left(\frac{n}{k}\right)}{a_0\left(\frac{n}{k}\right)} + \frac{1}{\gamma} \right| := \Delta. \quad \text{(A.0.39)}$$

The original proof of de Haan and Ferreira (2006, Corollary 2.4.5) shows that $\Delta \xrightarrow{P} 0$ and hence we are done. □

The following theorem and proof are entirely due to Ana Ferreira and Laurens de Haan. We merely adjusted it to fit our bootstrap approach.

Theorem A.5 (Bootstrap version of Ferreira and de Haan, 2015, Theorem 2.2.)**.** *Suppose $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ are i.i.d. random variables from distribution F with block maxima $X_1, \ldots, X_k$ and block length m. Assume that F satisfies the second-order condition for some $\gamma < \frac{1}{2}$ and $\rho \leq 0$ and $\sqrt{k}A(m) \to \lambda \in \mathbb{R}$ (with $A(m)$ as in Equation 2.3.5). For $r = 0, 1, 2, \ldots$ let β_r^* be a bootstrapped version of β_r, that is $\beta_0^* = \sum_{i=1}^k X_{i,k}^*$ and for $r \in \mathbb{Z}_{>0}$, $r < k$*

$$\beta_r^* = \frac{1}{k} \sum_{i=1}^{k} \frac{(i-1)\ldots(i-r)}{(k-1)\ldots(k-r)} X_{i,k}^*$$

with $X_{1,k}^*, \ldots, X_{k,k}^*$ bootstrapped block maxima. Then

$$\sqrt{k}\left(\frac{(r+1)\beta_r^* - b_m}{a_m} - D_r(\gamma)\right) \xrightarrow{d} Q_r^* + \sqrt{k}\Pi(m)$$

as $n \to \infty, k(n) \to \infty, n \geq k(n)^3$, jointly for $r = 0, 1, 2, \ldots$ with Q_r^* as in Theorem 4.3, $\Pi(m) = \frac{B_0(m) - b_m}{a_m}$ and

$$D_r(\xi) = \frac{(r+1)^{\xi}\Gamma(1-\xi) - 1}{\xi}, \quad \xi < 1.$$

Proof. Let, for $r = 0, 1, 2, 3, \ldots$,

$$J_k^{(r)}(s) = \frac{(\lceil ks \rceil - 1)\ldots(\lceil ks \rceil - r)}{(k-1)\ldots(k-r)}, \quad s \in [0,1].$$

Note that $J_k^{(r)}(s) \to s^r$, as $k \to \infty$, uniformly for in $s \in [0,1]$, and,

$$\frac{1}{k}\sum_{i=1}^{k} \frac{(i-1)\ldots(i-r)}{(k-1)\ldots(k-r)} = \int_0^1 J_k^{(r)}(s)\, ds = \frac{1}{r+1} = \int_0^1 s^r\, ds.$$

Moreover, note that

$$(r+1)\int_0^1 s^r \frac{(-\log s)^{-\xi} - 1}{\xi}\, ds = \frac{(r+1)^{\xi}\Gamma(1-\xi) - 1}{\xi}, \quad \xi < 1.$$

Then,

$$
\begin{aligned}
&\sqrt{k}\left(\frac{(r+1)\beta_r^* - b_m}{a_m} - \frac{(r+1)^\gamma\Gamma(1-\gamma)-1}{\gamma}\right) \\
&= \sqrt{k}\left(\frac{(r+1)\int_0^1 X^*_{\lceil ks\rceil,k} J_k^{(r)}(s)\, ds - b_m}{a_m} - (r+1)\int_0^1 \frac{(-\log s)^{-\gamma}-1}{\gamma} s^r\, ds\right) \\
&= \sqrt{k}(r+1)\int_0^1 \left(\frac{X^*_{\lceil ks\rceil,k} - b_m}{a_m} - \frac{(-\log s)^{-\gamma}-1}{\gamma}\right) J_k^{(r)}(s) ds \\
&\quad - \sqrt{k}(r+1)\int_0^1 \frac{(-\log s)^{-\gamma}-1}{\gamma}\left(s^r - J_k^{(r)}(s)\right) ds \\
&= \sqrt{k}(r+1)\int_0^{1/(k+1)} \left(\frac{X^*_{\lceil ks\rceil,k} - b_m}{a_m} - \frac{(-\log s)^{-\gamma}-1}{\gamma}\right) J_k^{(r)}(s) ds \\
&\quad \sqrt{k}(r+1)\int_{1/(k+1)}^{k/(k+1)} \left(\frac{X^*_{\lceil ks\rceil,k} - b_m}{a_m} - \frac{(-\log s)^{-\gamma}-1}{\gamma}\right) J_k^{(r)}(s) ds \\
&\quad \sqrt{k}(r+1)\int_{k/(k+1)}^{1} \left(\frac{X^*_{\lceil ks\rceil,k} - b_m}{a_m} - \frac{(-\log s)^{-\gamma}-1}{\gamma}\right) J_k^{(r)}(s) ds \\
&\quad - \sqrt{k}(r+1)\int_0^1 \frac{(-\log s)^{-\gamma}-1}{\gamma}\left(s^r - J_k^{(r)}(s)\right) ds \\
&:= I_1 + I_2 + I_3 + I_4
\end{aligned}
$$

For I_4: Since $\left(s^r - J_k^{(r)}(s)\right) = O_p\left(1/k\right)$ uniformly in s, $I_4 = O_p\left(1/\sqrt{k}\right)$.

For I_1, we first define $I_{1,1}$ and $I_{1,2}$:

$$
\begin{aligned}
\sqrt{k}\int_0^{1/(k+1)} \frac{X^*_{\lceil ks\rceil,k} - b_m}{a_m} s^r ds &= \sqrt{k}\int_0^{1/(k+1)} \frac{X_{\lceil ks\rceil,k} - b_m}{a_m} s^r ds + \sqrt{k}\int_0^{1/(k+1)} \frac{X^*_{\lceil ks\rceil,k} - X_{\lceil ks\rceil,k}}{a_m} s^r ds \\
&:= I_{1,1} + I_{1,2}.
\end{aligned}
$$

The original proof gives that $I_{1,1} = o_P(1)$. For $I_{1,2}$ we notice that

$$
I_{1,2} = \frac{1}{(r+1)} \frac{\sqrt{k}}{(k+1)^{r+1}} \frac{X^*_{1,k} - X_{1,k}}{a_m}.
$$

Now we take Equation 4.2.1 at $s_k = 1/(k+1)$ and subtract the equation in Ferreira and de Haan (2015, Theorem 2.1) at $s_k = 1/(k+1)$ to find

$$
\begin{aligned}
I_{1,2} = \frac{1}{(r+1)} \frac{1}{(k+1)^{r+1}} \Bigg(&\sqrt{k}\frac{b_m - B_0(m)}{a_m} + \frac{E_k^*(s_k)}{s_k(-\log s_k)^{1+\gamma}} - \frac{E_k\left(s_k\right)}{s_k(-\log s_k)^{1+\gamma}} + \frac{W_n\left(-\log s_k\right)}{(-\log s_k)^{1+\gamma}} \\
&- \sqrt{k}A_0(m)H_{\gamma,\rho}\left(\frac{1}{-\log s_k}\right) + \sqrt{k}A_0(m)\Psi_{\gamma,\rho}\left(\frac{1}{-\log s_k}\right)\Bigg) + o_P(1).
\end{aligned}
$$

Since $r \geq 0$, all the components on the right-hand side will vanish when multiplied by the term $\frac{1}{(k+1)^{r+1}}$ and therefore $I_{1,2} = o_P(1)$. We conclude $I_{1,1} + I_{1,2} = o_P(1)$ and thus $I_1 = o_P(1)$.

Again we define

$$
\begin{aligned}
\int_{k/(k+1)}^1 \frac{X^*_{\lceil ks\rceil,k} - b_m}{a_m} J_k^{(r)}(s) ds &= \int_{k/(k+1)}^1 \frac{X_{\lceil ks\rceil,k} - b_m}{a_m} J_k^{(r)}(s) ds + \int_{k/(k+1)}^1 \frac{X^*_{\lceil ks\rceil,k} - X_{\lceil ks\rceil,k}}{a_m} J_k^{(r)}(s) ds \\
&:= I_{3,1} + I_{3,2}
\end{aligned}
$$

The original proof shows that $I_{3,1} = o_P(1)$. In this part of the proof the fact that $\gamma < 1/2$ is needed.

For $I_{3,2}$ we see

$$I_{3,2} = \frac{X^*_{\lceil ks \rceil,k} - X_{\lceil ks \rceil,k}}{a_m}.$$

Now we use similar reasoning as for $I_{1,2}$ to show that $I_{3,2} = o_P(1)$ by taking $s_k = k/(k+1)$. As a result $I_3 = o_P(1)$.

Finally, I_2 has the same asymptotic behaviour as

$$(r+1)\int_{1/k}^{k/(k+1)} \sqrt{k}\left(\frac{X^*_{\lceil ks \rceil,k} - b_m}{a_m} - \frac{(-\log s)^{-\gamma} - 1}{\gamma}\right) s^r ds$$

which by Theorem 4.2 tends to

$$(r+1)\int_0^1 s^r \left(\frac{E^*_k(s)}{s(-\log s)^{1+\gamma}} + \frac{W_n(-\log s)}{(-\log x)^{1+\gamma}}\right) ds \\ + (r+1)\lambda \int_0^1 \Psi_{\gamma,\rho}\left(\frac{1}{-\log s}\right) s^r \, ds + (r+1)\sqrt{k}\int_0^1 s^r \frac{B_0(m) - b_m}{a_m}\, ds.$$

Hence we ultimately find

$$\sqrt{k}\left(\frac{(r+1)\beta^*_r - b_m}{a_m} - \frac{(r+1)^\gamma \Gamma(1-\gamma) - 1}{\gamma}\right) \xrightarrow{d} Q^*_r + \sqrt{k}\frac{B_0(m) - b_m}{a_m} \tag{A.0.40}$$

□