ERASMUS UNIVERSITY ROTTERDAM

Master Thesis Quantitative Finance

Return level estimation: a theoretical comparison between the peak-over-threshold approach and the block maxima method

August 16, 2019

Author:
Vos Vorstermans (457753)

Supervisor:
Chen Zhou

Abstract Two fundamental approaches in extreme value theory are the peak-over-threshold (POT) approach and the block maxima (BM) method. We provide a theoretical comparison between these methods for estimating the return level. The two methods are studied both for independent and identically distributed observations and for serial dependent observations. For both dependence structures and both methods, we propose an estimator for the return level and prove its asymptotic normality. Explicit calculations for specific time series models show that the BM method outperforms the POT approach in terms of having a lower asymptotic variance.

Keywords extreme value theory, stationary time series, Hill estimator, heavy tails, maximum likelihood estimation.

Contents

1 Introduction

Natural disasters such as earthquakes and floods often have a catastrophic societal impact. Modeling and analyzing such events is thus essential, particularly for risk management. For example, risk measures are used to determine the height of the dikes in the Netherlands of which approximately 30% is below sea level. In environmental risk management, a common used risk measure is the so-called return level defined as follows. Let $\{X_1, X_2, ...\}$ be a sequence of observations. Denote $M_m = \max(X_1, ..., X_m)$ as the maximum over m observations. Let $F_m(x) = \mathbb{P}(M_m \leq x)$. For $y \geq 1$, the return level $R_{m,y}$ is the quantile satisfying

$$\mathbb{P}\left(M_m > R_{m,y}\right) = 1/y. \tag{1}$$

In other words, the y-return level is defined as the $1/y$-th quantile of F_m,

$$R_{m,y} = F_m^{\leftarrow}\left(1 - 1/y\right) = \inf\{x \in \mathbb{R} : F_m(x) \geq 1 - 1/y\}.$$

It will take on average y blocks of size m until encountering the first such block whose maximum exceeds $R_{m,y}$. For example, the Dutch government sets a return level for the dikes such that the sea level may only exceed this level once in ten thousand years, on average. In this case yearly maxima are considered, i.e. $m = 365$. Here, "once in ten thousand years" corresponds to $y = 10000$. This thesis studies the estimation of the return level.

Estimating the return level relies on modeling the tail region of the distribution function of the observations. Making statistical inference for the tail is challenging due to the scarcity of data. Extreme value theory (EVT) offers a solid theoretical basis and framework for such a purpose. Different from the central limit theorem which studies the limit behavior of partial sums, extreme value theory is concerned with the limit behavior of sample extremes. Probabilistic EVT was developed by Fréchet (1927), Fisher and Tippett (1928) and von Mises (1936). The limit theory was completed by Gnedenko (1943) and the statistical theory was initiated by Pickands (1975).

By examining data in an intermediate region close to the tail, extreme value statistics exploit models to extrapolate intermediate properties to the tail region. Two fundamental approaches in classical extreme value statistics are the peak-over-threshold (POT) approach and the block maxima (BM) method. The POT approach consists of selecting the observations in a sample that exceed a certain high threshold. Such excesses are approximately generalized Pareto distributed (Pickands, 1975). The BM method consists of dividing the observed sample into non-overlapping blocks of equal size and restricting attention to the largest observation in each block (Gumbel, 1958). The block maxima follow approximately an extreme value distribution. Both methods often assume independent and identically distributed (i.i.d.) observations. However, this assumption is usually violated in practice since many financial and environmental data typically exhibit serial dependence.

The aim of this thesis is to determine which of the two methods is theoretically better for estimating the return level both for i.i.d. observations and for observations which exhibit serial dependence. For both dependence structures and both methods, we take the same steps. First, we motivate the estimator for the return level. Then, we present the (regularity) conditions and auxiliary lemmas necessary for the asymptotic theory and prove the asymptotic normality for the return level estimator. Finally, we provide a theoretical comparative study of these methods for specific time series models by explicitly calculating the asymptotic variances for the return level estimators based on the two methods.

Under independence, it is a general consensus among researchers in extreme value statistics that the POT approach makes use of extreme observations more efficiently than the BM method. The major heuristic reason is that the POT approach uses all large observations, while the BM method may miss some large observations falling into the same block (Bücher & Zhou, 2018). On the other hand, the available data often consists of block maxima alone, e.g. yearly maxima of a sea level. Then a researcher might only rely on the BM method. Under serial dependence, both methods are still valid; only the asymptotic variance of the estimators may differ from that in the i.i.d. case. Estimators based on the POT approach usually have a higher asymptotic variance (Drees, 2003), whereas the asymptotic variance of estimators based on the BM method stays the same, because the block maxima are usually distant from each other with very weak dependence and thus can be regarded as i.i.d (Bücher & Segers, 2018b).

Our theoretical comparison shows that both under independence and serial dependence, the BM method outperforms the POT approach in terms of having a lower asymptotic variance. This paper contributes to the existing literature of extreme value statistics by making at least three theoretical improvements. Firstly, for the POT approach, we reveal that the asymptotic distribution of the return level estimator is not dominated by the asymptotic distribution derived from estimating the extreme value index. This is often the case for estimating other tail related characteristics such as high quantiles; see, for example, Theorem 4.2 in De Haan, Mercadier, and Zhou (2016) and Theorem 2.2 in Drees (2003). We derive the joint asymptotic distribution of the respective estimators for different components of the return level estimator as shown in the proof of Theorem 3.4 and Theorem 3.9. Secondly, for the BM method under serial dependence, a difficult condition in Bücher and Segers (2018b) needs to be verified to prove the asymptotic normality for the return level estimator. In Section 4.4.4, we provide a different, slightly stronger, condition which is easier verifiable. Thirdly, to the best of our knowledge, there is no theoretical comparison between the BM method and the POT approach for return level estimation yet.

The paper is organized as follows. Section 2 reviews the fundamentals of EVT. Section 3 and 4 study the POT approach and the BM method respectively. Section 5 provides a theoretical comparison between the POT approach and the BM method for specific time series models. These sections all make the distinction between a simplified model with i.i.d. observations and a general model with serial dependent observations. Section 6 concludes the paper.

2 Extreme value theory

We introduce EVT under the assumption of i.i.d. observations (Section 2.1). Next, we present EVT for time series (Section 2.2).

2.1 I.i.d. case

Let $\{X_1, X_2, ...\}$ be an i.i.d. sequence of random variables with a common distribution function F. EVT studies the limit behavior of sample extremes $M_m = \max(X_1, ..., X_m)$ and relies on the following fundamental domain of attraction condition: there exist a constant $\gamma \in \mathbb{R}$ and sequences $a_m > 0$ and b_m, $m \in \mathbb{N}$, such that for all $1 + \gamma x > 0$,

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m - b_m}{a_m} \leq x\right) = \lim_{m\to\infty} F^m(a_m x + b_m) = G_\gamma(x), \tag{2}$$

where $G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right)$ is called the generalized extreme value (GEV) distribution. The limit appears unnecessarily specific, but it is in fact the only possible non-degenerate limit of the expression on the left-hand side if a_m and b_m are properly chosen (Fisher and Tippett (1928), Gnedenko (1943)). If (2) holds, then F is said to be in the domain of attraction of G_γ, denoted as $F \in \mathcal{D}(G_\gamma)$.

The shape parameter of the GEV distribution, γ, is called the extreme value index. The GEV distribution subsumes three types of distributions depending on the sign of γ; the Weibull distribution ($\gamma < 0$), the Gumbel distribution ($\gamma = 0$) and the Fréchet distribution ($\gamma > 0$, Fréchet (1927)). If $F \in \mathcal{D}(G_\gamma)$ with $\gamma > 0$, then F is called a heavy-tailed distribution.

2.1.1 The Fréchet case

In this research, we only consider distributions in the Fréchet domain of attraction ($\gamma > 0$) because financial data usual exhibits heavy tails (Jansen & De Vries, 1991). The domain of attraction condition for the Fréchet case can be represented in different ways. We state the following two theorems where $U := (1/(1-F))^{\leftarrow}$ denotes the left-continuous inverse function of $1/(1-F)$.

Theorem 2.1 (Corollary 1.2.10, de Haan and Ferreira (2007))**.** *Suppose that $F \in \mathcal{D}(G_\gamma)$ with $\gamma > 0$. Then there exists a positive number γ such that*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma, \tag{3}$$

for $x > 0$.

Theorem 2.2 (Corollary 1.2.4, de Haan and Ferreira (2007))**.** *Suppose that $F \in \mathcal{D}(G_\gamma)$ with $\gamma > 0$. Then for $x > 0$,*

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m}{a_m} \leq x\right) = \lim_{m\to\infty} F^m(a_m x) = \exp\left(-x^{-1/\gamma}\right), \tag{4}$$

with $a_m = U(m)$.

In statistical analysis, the peak-over-threshold approach (Section 3) is motivated by (3) and the block maxima method (Section 4) is motivated by (4).

2.2 Serial dependence case

Let $\{X_1, X_2, ...\}$ be a strictly stationary time series with a common distribution function F. That is, for any $n \in \mathbb{N}$ and $h, i_1, ..., i_n \in \mathbb{Z}$, the distribution of $\{X_{i_1+h}, ..., X_{i_n+h}\}$ is the same as the distribution of $\{X_{i_1}, ..., X_{i_n}\}$ (Bücher & Segers, 2018b). Under serial dependence, the first equality in (4) may not hold. Hence, more sophisticated arguments must be found to establish limiting theory similar to the i.i.d. case. We state the following theorem.

Theorem 2.3 (Leadbetter (1983), Bücher and Zhou (2018))**.** *Suppose that* $F \in \mathcal{D}(G_\gamma)$ *with* $\gamma > 0$*. Assume that mild conditions on the serial dependence structure (known as* $D(u_n)$*-conditions) are met. Then there exists a constant* $\theta \in [0, 1]$ *such that*

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m}{a_m} \le x\right) = \exp\left(-\theta x^{-1/\gamma}\right), \tag{5}$$

for all $x \in \mathbb{R}$ *and with* $a_m = U(m)$.

The constant θ is called the extremal index and can be interpreted as a summary measure for the strength of serial dependence between extremes (Bücher & Segers, 2018a). For an i.i.d. process $\theta = 1$ and if $\theta \to 0$, there is increasing dependence between the extremes of the process.

By using (4) and (5), we obtain that as $m \to \infty$,

$$\mathbb{P}\left(\frac{M_m}{a_m} \le x\right) \sim \mathbb{P}\left(\frac{\tilde{M}_{\lfloor m\theta \rfloor}}{a_m} \le x\right), \tag{6}$$

for all $x \in \mathbb{R}$ and where $\tilde{M}_{\lfloor m\theta \rfloor} = \max(\tilde{X}_1, ..., \tilde{X}_{\lfloor m\theta \rfloor})$ with $\tilde{X}_i$, $i \in \mathbb{N}$, an associated i.i.d. series with the same distribution function F.[1] Intuitively, the maximum of m observations from the stationary series with extremal index θ behaves like the maximum of $\lfloor m\theta \rfloor < m$ observations from the associated i.i.d. series (McNeil, 1998). Hence, the serial dependence between large observations reduces the effective sample size by the factor θ (Drees, 2003).

If $\theta > 0$, then letting $\tilde{a}_m = a_m \theta^\gamma$ and by using (5), we obtain that

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{M_m}{\tilde{a}_m} \le x\right) = \exp\left(-\theta(\theta^\gamma x)^{-1/\gamma}\right) = \exp\left(-x^{-1/\gamma}\right), \tag{7}$$

for every $x \in \mathbb{R}$. Notice that, unless $\theta = 1$, $\tilde{a}_m$ is different from a_m.

[1] Here $c_m \sim d_m$ means $c_m/d_m \to 1$ as $m \to \infty$, and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

3 Peak-over-threshold approach

We discuss the peak-over-threshold approach for estimating the return level. Both under a simplified model without serial dependence (Section 3.1) and under a general model with serial dependence (Section 3.2), we can express the return level as a high quantile. The POT approach provides an extrapolation method to estimate a high quantile and consequently leads to an estimator for the return level.

For both the i.i.d. case and the serial dependence case, we start by motivating the estimator for the return level. Then we present the conditions and auxiliary lemmas necessary for the asymptotic theory. Finally, we state and prove the main theorem on the asymptotic normality of the estimator for the return level.

3.1 I.i.d. case

Let $\{X_1, X_2, ...\}$ be an i.i.d. sequence of random variables with a common distribution function F. We assume that this distribution function belongs to the Fréchet domain of attraction ($\gamma > 0$).

3.1.1 Motivating the estimator

The return level, $R_{m,y}$, refers to a quantile of $F_m(x) = \mathbb{P}(M_m \leq x)$, where $M_m = \max(X_1, ..., X_m)$. From (1), we deduce that

$$F(R_{m,y}) = (1 - 1/y)^{1/m}. \tag{8}$$

Hence, the return level is equal to the $1 - (1 - 1/y)^{1/m}$-th quantile of F,

$$R_{m,y} = x_{1-(1-1/y)^{1/m}}, \tag{9}$$

where $x_\alpha = F^{-1}(1 - \alpha)$ denotes the α-th quantile.

Recall the domain of attraction condition in (3) and notice that $U(n) = x_{1/n}$. The limit relation (3) determines how a high quantile, say $U(tx)$, can be extrapolated from an intermediate quantile $U(t)$ (De Haan et al., 2016). By using (3), we obtain

$$\begin{aligned} x_\alpha = F^{-1}(1-\alpha) &\approx F^{-1}\left(1 - \frac{k}{n}\right)\left(\frac{k}{n\alpha}\right)^\gamma \\ &\approx X_{n-k,n}\left(\frac{k}{n\alpha}\right)^{\hat{\gamma}} =: \hat{x}_\alpha, \end{aligned} \tag{10}$$

where $\hat{\gamma}$ denotes a suitable estimator of the extreme value index γ and $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ are the order statistics of $\{X_1, ..., X_n\}$. Notice that $\hat{x}_\alpha$ is exactly a Weisman-type estimator (Weissman, 1978). To justify the first approximation k/n must be small. On the other hand, k should be sufficiently large such that the $(n-k)$-th order statistic $X_{n-k,n}$ estimates the intermediate quantile $F^{-1}(1-k/n)$ well (Drees, 2003). We thus require that k is an intermediate sequence, that is, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

In the Fréchet case, Hill (1975) proposes for estimating γ the so-called Hill estimator

$$\hat{\gamma}_{\mathrm{H}} := \frac{1}{k} \sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}, \tag{11}$$

where k is an intermediate sequence. The Hill estimator is consistent under the domain of attraction condition (3) (see de Haan and Ferreira (2007), Theorem 3.2.2).

From (9), a suitable estimator for the return level is an estimator for the high quantile. By using (10), we propose the following estimator for the return level

$$\hat{R}_{m,y} := X_{n-k,n} \left(\frac{k}{n(1-(1-1/y)^{1/m})} \right)^{\hat{\gamma}_{\mathrm{H}}}, \tag{12}$$

where $\hat{\gamma}_{\mathrm{H}}$ is the Hill estimator defined in (11) and k is an intermediate sequence.

3.1.2 Conditions for asymptotic theory

We present the conditions for establishing the asymptotic normality of the return level estimator (12) in Section 3.1.4. We need a second order reinforcement of the domain of attraction condition (3) in combination with a growth restriction on the number of blocks.

Second order condition (Theorem 2.3.9, de Haan and Ferreira (2007)) Suppose that there exist a positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a real number $\rho < 0$ such that

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} = x^{\gamma} \frac{x^{\rho}-1}{\rho}, \tag{13}$$

for all $x > 0$.

The second order condition quantifies the speed of convergence in (3). The parameter ρ controls the speed of convergence for the return level estimator towards a normal distribution.

Condition on k Suppose that the intermediate sequence k satisfies

$$\lim_{n\to\infty} \sqrt{k} A\left(\frac{n}{k}\right) = 0. \tag{14}$$

The condition (14) imposes an upper bound on the speed at which k goes to infinity. As it appears below, it assumes away the asymptotic bias for the estimator for the return level.

3.1.3 Auxiliary lemmas

The following three lemmas are useful in the proof of the asymptotic normality of the estimator for the return level (Theorem 3.4).

Lemma 3.1 (Theorem 3.2.5, de Haan and Ferreira (2007))**.** *Suppose that* $\{X_1, X_2, ...\}$ *is an i.i.d. sequence of random variables with a common distribution function* F*. Assume that* F *satisfies the second order condition (13) with parameters* $\gamma > 0$ *and* $\rho \leq 0$*. Suppose that the intermediate sequence* k *satisfies (14). Then as* $n \to \infty$*,*

$$\sqrt{k}(\hat{\gamma}_H - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\right).$$

Lemma 3.2. *Suppose that* $k = \lfloor n/m \rfloor$*. Assume that both* $m \to \infty$ *and* $k \to \infty$ *as* $n \to \infty$*. Then as* $n \to \infty$*,*

$$\frac{k}{n\left(1 - (1 - 1/y)^{1/m}\right)} \sim \frac{1}{-\log(1 - 1/y)}.$$

Proof. Consider the expansion

$$(1 - 1/y)^{1/m} = \exp\left(\log(1 - 1/y)\frac{1}{m}\right).$$

Because $e^x \sim 1 + x$ as $x \to 0$, we get that as $n \to \infty$,

$$(1 - 1/y)^{1/m} \sim 1 + \frac{\log(1 - 1/y)}{m}.$$

Hence,

$$n\left(1 - (1 - 1/y)^{1/m}\right) \sim -k\log(1 - 1/y),$$

and the lemma is proved. ∎

Lemma 3.3 (Theorem 2.4.8, de Haan and Ferreira (2007))**.** *Suppose that* $\{X_1, X_2, ...\}$ *is an i.i.d. sequence of random variables with a common distribution function* F*. Let* $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ *be the order statistics of* $\{X_1, ..., X_n\}$*. Assume that* F *satisfies the second order condition (13) with parameters* $\gamma > 0$ *and* $\rho \leq 0$*. Suppose that the intermediate sequence* k *satisfies (14). Then there exists a sequence of Brownian motions* $\{W_n(s)\}_{s \geq 0}$ *such that as* $n \to \infty$*,*

$$\sup_{0 < s \leq 1} s^{\gamma + 1/2 + \varepsilon}\left|\sqrt{k}\left(\frac{X_{n - \lfloor ks \rfloor, n}}{U\left(\frac{n}{k}\right)} - s^{-\gamma}\right) - \gamma s^{-\gamma - 1}W_n(s)\right| \xrightarrow{P} 0,$$

for each $\varepsilon > 0$*.*

3.1.4 Main result

The asymptotic normality of the return level estimator (12) is given by the following theorem.

Theorem 3.4. *Suppose that $\{X_1, X_2, ...\}$ is an i.i.d. sequence of random variables with a common distribution function F. Assume that F satisfies the second order condition (13) with parameters $\gamma > 0$ and $\rho < 0$. Suppose that the intermediate sequence k satisfies (14). Then as $n \to \infty$,*

$$\sqrt{k}\left(\frac{\hat{R}_{m,y}}{R_{m,y}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \left(1 + \log^2(c_y)\right)\gamma^2\right),$$

with $c_y = -\log(1 - 1/y)$.

Proof. Consider the expansion

$$\begin{aligned}\frac{\hat{R}_{m,y}}{R_{m,y}} &= \frac{X_{n-k,n}\left(\frac{k}{n(1-(1-1/y)^{1/m})}\right)^{\hat{\gamma}_{\mathrm{H}}}}{U(\frac{n}{k})\left(\frac{k}{n(1-(1-1/y)^{1/m})}\right)^{\gamma}} \cdot \frac{U(\frac{n}{k})\left(\frac{k}{n(1-(1-1/y)^{1/m})}\right)^{\gamma}}{R_{m,y}}\\ &=: I_1 \cdot I_2.\end{aligned}$$

Rewrite I_1 as

$$\begin{aligned}I_1 &= \frac{X_{n-k,n}}{U(\frac{n}{k})} \cdot \left(\frac{k}{n(1-(1-1/y)^{1/m})}\right)^{\hat{\gamma}_{\mathrm{H}}-\gamma}\\ &=: I_{11} \cdot I_{12}.\end{aligned}$$

By taking $s = 1$ in Lemma 3.3, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{X_{n-k,n}}{U\left(\frac{n}{k}\right)} - 1\right) - \gamma W_n(1) \xrightarrow{P} 0.$$

Hence,

$$\sqrt{k}(I_{11} - 1) \xrightarrow{d} B \sim \mathcal{N}(0, \gamma^2),$$

where $B = \gamma W_n(1)$. Rewrite I_{12} as

$$I_{12} = \exp\left(\frac{\log\left(\frac{k}{n(1-(1-1/y)^{1/m})}\right)}{\sqrt{k}}\sqrt{k}(\hat{\gamma}_{\mathrm{H}} - \gamma)\right).$$

From Lemma 3.1, we get that as $n \to \infty$,

$$\sqrt{k}(\hat{\gamma}_{\mathrm{H}} - \gamma) \xrightarrow{d} \Gamma \sim \mathcal{N}\left(0, \gamma^2\right),$$

From the second proof of Theorem 3.2.5 in de Haan and Ferreira (2007), p. 76, and by using (14), it follows that

$$\Gamma = \gamma\left(\int_0^1 s^{-1}W_n(s)\mathrm{d}s - W_n(1)\right).$$

By using Lemma 3.2, we get that as $n \to \infty$,

$$\log\left(\frac{k}{n\left(1-(1-1/y)^{1/m}\right)}\right) \to -\log\left(-\log(1-1/y)\right) = -\log c_y.$$

According to the Cramér's delta method (Cramér, 1946) and Slutsky's theorem (Slutsky, 1925), we get that as $n \to \infty$,

$$\sqrt{k}(I_{12}-1) \xrightarrow{d} -\log c_y \Gamma \sim \mathcal{N}\left(0, \log^2(c_y)\gamma^2\right).$$

We compute the covariance between Γ and B as follows

$$\begin{aligned}
\operatorname{Cov}(\Gamma, B) &= \operatorname{Cov}\left(\gamma\left(\int_0^1 s^{-1} W_n(s)\mathrm{d}s - W_n(1)\right), \gamma W_n(1)\right)\\
&= \gamma^2\left(\operatorname{Cov}\left(\int_0^1 s^{-1} W_n(s)\mathrm{d}s, W_n(1)\right) - \operatorname{Cov}\left(W_n(1), W_n(1)\right)\right)\\
&= \gamma^2\left(\int_0^1 s^{-1}\mathbb{E}\left(W_n(s)W_n(1)\right)\mathrm{d}s - 1\right)\\
&= \gamma^2\left(\int_0^1 s^{-1}\min(s,1)\mathrm{d}s - 1\right)\\
&= 0.
\end{aligned}$$

Hence, and according to the Cramér's delta method, we get that as $n \to \infty$,

$$\sqrt{k}\left(I_1 - 1\right) \xrightarrow{d} B - \log(c_y)\Gamma \sim \mathcal{N}\left(0, \left(1+\log^2(c_y)\right)\gamma^2\right).$$

For I_2, consider the expansion

$$\sqrt{k}A\left(\frac{n}{k}\right)\frac{I_2 - 1}{A\left(\frac{n}{k}\right)}.$$

From Theorem 2.3.9 in de Haan and Ferreira (2007), we get that as $n \to \infty$,

$$\frac{I_2 - 1}{A\left(\frac{n}{k}\right)} \to -\frac{1}{\rho}. \tag{15}$$

Combining (14) with (15), we get that as $n \to \infty$,

$$\sqrt{k}(I_2 - 1) \to 0.$$

The theorem is proved by combining the limit properties of the two terms in the expansion. According to Cramér's delta method, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{\hat{R}_{m,y}}{R_{m,y}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \left(1+\log^2(c_y)\right)\gamma^2\right).$$

□

3.2 Serial dependence case

Let $\{X_1, X_2, ...\}$ be a strictly stationary time series with a common distribution function F. We assume that this distribution function belongs to the Fréchet domain of attraction ($\gamma > 0$).

3.2.1 Motivating the estimator

Under serial dependence, the identity (8) may not hold. This section shows that we are still able to express the return level as a high quantile. However, the probability level is related to the extremal index (McNeil, 1998).

By using (1) and (6), we obtain as $m \to \infty$,

$$F(R_{m,y}) \sim (1 - 1/y)^{1/(m\theta)}.$$

Hence, the return level is approximately equal to the $1-(1-1/y)^{1/(m\theta)}$-quantile of F,

$$R_{m,y} \approx x_{1-(1-1/y)^{1/(m\theta)}}. \tag{16}$$

For the estimation of the return level, we need an accurate estimator for the extremal index θ.

Regarding the estimation of the extremal index, a large variety of estimators has been studied in the literature (see Bücher and Zhou (2018), Section 3.2, for an overview). We use the estimator proposed in Berghaus and Bücher (2018) which is defined as follows. Suppose that we observe a stretch of length n from the time series $\{X_1, X_2, ...\}$. Divide the sample into k blocks of length m. For simplicity, we assume that $n = k \cdot m$ (otherwise, the final block would consist of less than m observations and should be omitted). For $i = 1, ..., k$, let $M_{mi} = \max(X_{(i-1)m+1}, ..., X_{im})$ denote the maximum of the i-th block of observations. Then, the estimator for the extremal index is given as

$$\hat{\theta}^{\mathrm{B}} = \left(\frac{1}{k} \sum_{i=1}^{k} \hat{Z}_{ni} \right)^{-1}, \tag{17}$$

where $\hat{Z}_{ni} = m(1 - \hat{F}_n(M_{mi}))$ and $\hat{F}_n(x) = \frac{1}{n} \sum_{s=1}^{n} \mathbf{1}_{\{X_s \le x\}}$ denotes the empirical distribution function of $\{X_1, ..., X_n\}$.

From (16), an estimator for the high quantile is still a suitable estimator for the return level. By using (10), we propose the following estimator for the return level

$$\hat{R}_{m,y} := X_{n-k,n} \left(\frac{k}{n(1 - (1 - 1/y)^{1/(m\hat{\theta}^{\mathrm{B}})})} \right)^{\hat{\gamma}_{\mathrm{H}}}, \tag{18}$$

where $\hat{\gamma}_{\mathrm{H}}$ is the Hill estimator defined in (11), $\hat{\theta}^{\mathrm{B}}$ is the estimator for the extremal index defined in (17) and k is an intermediate sequence.

3.2.2 Conditions for asymptotic theory

We present the regularity conditions for establishing the asymptotic normality of the return level estimator (18) in Section 3.2.4. The serial dependence structure follows from the so-called β-mixing conditions. The β-mixing conditions have been introduced by Rootzén (1995), Drees (2000), Drees (2003) and Rootzén (2009) as follows. The sequence $\{X_1, X_2, ...\}$ is called β-mixing (or absolutely regular) if as $\ell \to \infty$,

$$\beta(\ell) := \sup_{j\in\mathbb{N}} \mathbb{E}\left(\sup_{E\in\mathcal{B}_{j+\ell+1}^{\infty}} \left| \mathbb{P}(E|\mathcal{B}_1^j) - \mathbb{P}(E) \right| \right) \to 0,$$

where $\mathcal{B}_1^j$ and $\mathcal{B}_{j+\ell+1}^{\infty}$ denote the σ-fields generated by $(X_i)_{1\le i\le j}$ and $(X_i)_{j+\ell+1\le i}$, respectively. The coefficients $\beta(\ell)$ are called the β-mixing constants of the sequence.

The following first three conditions are condition (C1), (11) and (13) in Drees (2003). The last two conditions are a slightly adaption of conditions (C4) and (C5) in Drees (2003). In fact, they imply that conditions (C4) and (C5) are correct (see De Haan et al. (2016), Section 3).

Regularity conditions Suppose there exist a constant $\varepsilon > 0$, a function $c(\cdot,\cdot)$, an intermediate sequence k, a sequence l_n, $n \in \mathbb{N}$, a positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a real number $\rho < 0$ such that

Condition 3.1.

$$\lim_{n\to\infty} \frac{\beta(l_n)}{l_n} n + l_n k^{-1/2} \log^2 k = 0;$$

Condition 3.2.

$$\lim_{n\to\infty} \frac{n}{l_n k} \mathrm{Cov}\left(\sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-kx/n)\}}, \sum_{i=1}^{l_n} 1_{\{X_i > F^{-1}(1-ky/n)\}} \right) = c(x,y),$$

for all $0 < x, y \le 1 + \varepsilon$;

Condition 3.3. For some constant $D > 0$,

$$\frac{n}{l_n k} \mathbb{E}\left(\left(\sum_{i=1}^{l_n} 1_{\{F^{-1}(1-ky/n) < X_i \le F^{-1}(1-kx/n)\}} \right)^4 \right) \le D(y - x),$$

for all $0 < x, y \le 1 + \varepsilon$ and $n \in \mathbb{N}$;

Condition 3.4.

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

for all $x > 0$;

Condition 3.5.

$$\lim_{n\to\infty} \sqrt{k} A\left(\frac{n}{k}\right) = 0;$$

We discuss these conditions one by one. The β-mixing constants measure the influence of the past on future events. Intuitively, the condition 3.1 states that this influence disappears sufficiently fast as past and future are separated by a time interval of increasing length (Drees, 2003). Drees (2000) shows that condition 3.1 is fulfilled if the original time series $\{X_1, X_2, ...\}$ is geometrically β-mixing, that is, $\beta(\ell) = O(\eta^\ell)$ for some $\eta \in (0,1)$. In that case, one may take $l_n = -2\log n/\log\eta$ (De Haan et al., 2016). Condition 3.2 is satisfied if all vectors (X_1, X_{1+m}) belong to the domain of attraction of a bivariate extreme value distribution (see Drees (2003), Remark 2.1). In that case, for any sequence k, one may take a sequence l_n such that $l_n = o(n/k)$. The limit function $c(x,y)$ depends only on the tail dependence structure of (X_1, X_{1+m}) for $m \in \mathbb{N}$ (De Haan et al., 2016). These two sufficient versions of conditions 3.1 and 3.2 hold for the ARMA models, see Section 5.2 below. Furthermore, condition 3.3 has been verified for these models as well (De Haan et al., 2016). Condition 3.4 is the second order condition (13). Condition 3.5 imposes an upper bound on the speed at which k goes to infinity to assume away the asymptotic bias for the estimator for the return level.

3.2.3 Auxiliary lemmas

The following four lemmas are useful in the proof of the asymptotic normality of the estimator for the return level (Theorem 3.9).

Lemma 3.5 (Drees (1998a), Drees (1998b), Theorem 2.2 in Drees (2003))**.** *Suppose that $\{X_1, X_2, ...\}$ is a stationary β-mixing time series with common distribution function F. Suppose that the regularity conditions 3.1-3.5 hold. Then as $n \to \infty$,*

$$\sqrt{k}(\hat{\gamma}_H - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \sigma_H^2\right),$$

where

$$\sigma_H^2 = \gamma^2 \int_{(0,1]}\int_{(0,1]} (st)^{-(\gamma+1)} c(s,t)\nu_H(\mathrm{d}s)\nu_{\mathrm{H}}(\mathrm{d}t),$$

where $c(\cdot,\cdot)$ is defined in the regularity condition 3.2 and with signed measure

$$\nu_H(\mathrm{d}t) = t^\gamma \mathrm{d}t - \varepsilon_1(\mathrm{d}t),$$

where ε_1 denotes the Dirac measure with mass 1 at 1.

Lemma 3.6. *Suppose that $k = \lfloor n/m \rfloor$. Assume that both $m \to \infty$ and $k \to \infty$ as $n \to \infty$. Then as $n \to \infty$,*

$$\frac{k}{n\left(1 - (1-1/y)^{1/(m\theta)}\right)} \sim \frac{\theta}{-\log(1-1/y)}.$$

Proof. The proof is similar to that of Lemma 3.2. □

Lemma 3.7 (Theorem 2.1, Drees (2003))**.** *Suppose that $\{X_1, X_2, ...\}$ is a stationary β-mixing time series with common distribution function F. Let $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ be the order statistics of $\{X_1, ..., X_n\}$. Suppose that the regularity conditions 3.1-3.5 hold. Then there exists a sequence of centred Gaussian processes $\{e_n(s)\}_{s\geq 0}$ with covariance function c defined in the regularity condition 3.2 such that as $n \to \infty$,*

$$\sup_{0<s\leq 1} s^{\gamma+3/4+\varepsilon} \left| \sqrt{k}\left(\frac{X_{n-\lfloor ks\rfloor,n}}{U\left(\frac{n}{k}\right)} - s^{-\gamma}\right) - \gamma s^{-\gamma-1} e_n(s)\right| \xrightarrow{P} 0,$$

for each $\varepsilon > 0$.

The asymptotic normality of the estimator for the extremal index (17) has been established with some extra conditions. The following lemma is proved in Theorem 3.2 in Berghaus and Bücher (2018).

Lemma 3.8. *Suppose that Condition 2.1 and (2.2) in Berghaus and Bücher (2018) are met. Then as $n \to \infty$,*

$$\sqrt{k}(\hat{\theta^{B}} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \theta^4\sigma_B^2\right),$$

where σ_B^2 is a complicated expression.

3.2.4 Main result

The asymptotic normality of the return level estimator (18) is given by the following theorem.

Theorem 3.9. *Suppose that $\{X_1, X_2, ...\}$ is a stationary β-mixing time series with common distribution function F. Assume that Condition 2.1 and (2.2) in Berghaus and Bücher (2018) are met with an intermediate sequence k_1 such that $k/k_1 \to 0$. Assume in addition that as $n \to \infty$,*

$$A\left(m\right)\left(\frac{x_{1-(1-1/y)^{1/(m\theta)}}}{R_{m,y}} - 1\right) \to 0. \tag{19}$$

Suppose that the regularity conditions 3.1-3.5 hold. Then as $n \to \infty$,

$$\sqrt{k}\left(\frac{\hat{R}_{m,y}}{R_{m,y}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\theta)'\Sigma_\gamma\beta(y,\theta)\right),$$

where $\beta(y,\theta) = \left(\log\left(\frac{\theta}{c_y}\right), 1\right)'$ with $c_y = -\log(1-1/y)$ and

$$\Sigma_\gamma = \begin{pmatrix} \sigma_H^2 & \gamma^2\left(\int_0^1 s^{-1}c(s,1)\mathrm{d}s - c(1,1)\right) \\ \gamma^2\left(\int_0^1 s^{-1}c(s,1)\mathrm{d}s - c(1,1)\right) & \gamma^2 c(1,1) \end{pmatrix},$$

with σ_H^2 and $c(\cdot,\cdot)$ as defined in Lemma 3.5 and in the regularity condition 3.2 respectively.

Proof. Consider the expansion

$$\begin{aligned}\frac{\hat{R}_{m,y}}{R_{m,y}} &= \frac{X_{n-k,n}}{U\left(\frac{n}{k}\right)} \cdot \frac{\left(\frac{k}{n(1-(1-1/y)^{1/(m\hat{\theta}^{\mathrm{B}})})}\right)^{\hat{\gamma}_{\mathrm{H}}}}{\left(\frac{k}{n(1-(1-1/y)^{1/(m\theta)})}\right)^{\gamma}} \cdot \frac{U\left(\frac{n}{k}\right)\left(\frac{k}{n(1-(1-1/y)^{1/(m\theta)})}\right)^{\gamma}}{x_{1-(1-1/y)^{1/(m\theta)}}} \\ &\quad \cdot \frac{x_{1-(1-1/y)^{1/(m\theta)}}}{R_{m,y}} \\ &=: I_1 \cdot I_2 \cdot I_3 \cdot I_4.\end{aligned}$$

By taking $s = 1$ in Lemma 3.7, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{X_{n-k,n}}{U\left(\frac{n}{k}\right)} - 1\right) - \gamma e_n(1) \xrightarrow{P} 0.$$

Hence,

$$\sqrt{k}(I_1 - 1) \xrightarrow{d} B \sim \mathcal{N}\left(0, \gamma^2 c(1,1)\right),$$

where $B = \gamma e_n(1)$ and with c as defined in the regularity condition 3.2.

For I_2, consider the expansion

$$\begin{aligned}I_2 &= \frac{\left(\frac{k}{n(1-(1-1/y)^{1/(m\hat{\theta}^{\mathrm{B}})})}\right)^{\hat{\gamma}_{\mathrm{H}}}}{\left(\frac{k}{n(1-(1-1/y)^{1/(m\hat{\theta}^{\mathrm{B}})})}\right)^{\gamma}} \cdot \frac{\left(\frac{k}{n(1-(1-1/y)^{1/(m\hat{\theta}^{\mathrm{B}})})}\right)^{\gamma}}{\left(\frac{k}{n(1-(1-1/y)^{1/(m\theta)})}\right)^{\gamma}} \\ &= \left(\frac{k}{n(1-(1-1/y)^{1/(m\hat{\theta}^{\mathrm{B}})})}\right)^{\hat{\gamma}_{\mathrm{H}}-\gamma} \cdot \left(\frac{1-(1-1/y)^{1/(m\theta)}}{1-(1-1/y)^{1/(m\hat{\theta}^{\mathrm{B}})}}\right)^{\gamma} \\ &=: I_{21} \cdot I_{22}.\end{aligned}$$

Rewrite I_{21} as

$$I_{21} = \exp\left(\frac{\log\left(\frac{k}{n(1-(1-1/y)^{1/(m\hat{\theta}^{\mathrm{B}})})}\right)}{\sqrt{k}}\sqrt{k}(\hat{\gamma}_{\mathrm{H}} - \gamma)\right)$$

From Lemma 3.5, we get that as $n \to \infty$,

$$\sqrt{k}(\hat{\gamma}_{\mathrm{H}} - \gamma) \xrightarrow{d} \Gamma \sim \mathcal{N}\left(0, \sigma^2_{\mathrm{H}}\right),$$

with σ^2_{H} as defined in Lemma 3.5. From the proof of Theorem 2.2 in Drees (2003), p. 626, it follows that

$$\Gamma = \gamma \int_0^1 s^{-\gamma-1} e_n(s) \nu_{\mathrm{H}}(\mathrm{d}s),$$

with ν_{H} as defined in Lemma 3.5. From Lemma 3.6 and Lemma 3.8, we get that as $n \to \infty$,

$$\log\left(\frac{k}{n\left(1-(1-1/y)^{1/(m\hat{\theta}^{\mathrm{B}})}\right)}\right) \to \log\left(\frac{\theta}{-\log(1-1/y)}\right) = \log\left(\frac{\theta}{c_y}\right).$$

According to the Cramér's delta method (Cramér, 1946) and Slutsky's theorem (Slutsky, 1925), we get that as $n \to \infty$,

$$\sqrt{k}(I_{21} - 1) \xrightarrow{d} \log\left(\frac{\theta}{c_y}\right)\Gamma \sim \mathcal{N}\left(0, \log^2\left(\frac{\theta}{c_y}\right)\gamma^2\right).$$

By using Lemma 3.6, we get that as $n \to \infty$,

$$\frac{1 - (1 - 1/y)^{1/(m\theta)}}{1 - (1 - 1/y)^{1/(m\hat{\theta}^{\mathrm{B}})}} \sim \frac{\hat{\theta}^{\mathrm{B}}}{\theta}.$$

From Lemma 3.8 with k_1, we get that as $n \to \infty$,

$$\sqrt{k_1}\,(I_{22} - 1) = O_{\mathbb{P}}(1),$$

Because $\frac{k}{k_1} \to 0$, we get that as $n \to \infty$,

$$\sqrt{k}\,(I_{22} - 1) = o_{\mathbb{P}}(1).$$

By combining the limit properties of I_{21} and I_{22} and according to Cramèr's delta method, we get that as $n \to \infty$,

$$\sqrt{k}(I_2 - 1) \xrightarrow{d} \log\left(\frac{\theta}{c_y}\right)\Gamma \sim \mathcal{N}\left(0, \log^2\left(\frac{\theta}{c_y}\right)\gamma^2\right).$$

We compute the covariance between Γ and B as follows

$$\begin{aligned}
\mathrm{Cov}\,(\Gamma, B) &= \mathrm{Cov}\left(\gamma\int_0^1 s^{-\gamma-1}e_n(s)\,(s^{\gamma}\mathrm{d}s - \varepsilon_1(\mathrm{d}s))\,, \gamma e_n(1)\right)\\
&= \gamma^2\ \mathrm{Cov}\left(\int_0^1 s^{-1}e_n(s)\mathrm{d}s - e_n(1), e_n(1)\right)\\
&= \gamma^2\left(\mathrm{Cov}\left(\int_0^1 s^{-1}e_n(s)\mathrm{d}s, e_n(1)\right) - \mathrm{Cov}\,(e_n(1), e_n(1))\right)\\
&= \gamma^2\left(\int_0^1 s^{-1}c\,(s,1)\,\mathrm{d}s - c(1,1)\right).
\end{aligned}$$

Hence, and according to the Cramér's delta method, we get that as $n \to \infty$,

$$\sqrt{k}(I_1 I_2 - 1) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\theta)'\Sigma\beta(y,\theta)\right).$$

For I_3, consider the expansion

$$\sqrt{k}A\left(\frac{n}{k}\right)\frac{I_3 - 1}{A\left(\frac{n}{k}\right)}.$$

From Theorem 2.3.9 in de Haan and Ferreira (2007), we get that as $n \to \infty$,

$$\frac{I_3 - 1}{A\left(\frac{n}{k}\right)} \to -\frac{1}{\rho}, \tag{20}$$

as $n \to \infty$. Combining the regularity condition 3.5 with (20), we get that as $n \to \infty$,

$$\sqrt{k}(I_3 - 1) \to 0.$$

From again the regularity condition 3.5 and (19), we get that as $n \to \infty$,

$$\sqrt{k} A\left(\frac{n}{k}\right)(I_4 - 1) \to 0.$$

The theorem is proved by combining the limit properties of the four terms in the expansion. According to Cramèr's delta method, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{\hat{R}_{m,y}}{R_{m,y}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\theta)' \Sigma_\gamma \beta(y,\theta)\right).$$

□

Remark 3.10. In typical high quantile estimation, normally it is assumed that

$$\lim_{n\to\infty} \frac{1}{\sqrt{k}} \log(n\alpha) = 0, \quad \lim_{n\to\infty} n\alpha/k = 0. \tag{21}$$

where α is the tail probability of the quantile x_α to be estimated; see, for example, (15) in Drees (2003) and Theorem 4.2 in De Haan et al. (2016). With assumption (21) only the limit of the estimator for the extreme value index contributes to the final asymptotic limit of the estimator for the high quantile. In our case, by Lemma 3.6, one can actually verify that (21) does not hold. Therefore, both the limit of the order statistic and the Hill estimator contribute to the final asymptotic limit of the estimator for the return level. This situation occurs for both the i.i.d. case and the serial dependence case.

4 Block maxima method

We discuss the block maxima method for estimating the return level. Let $\{X_1, X_2, ..., X_n\}$ be a sample of observations. For a given block size $m \in \{1, ..., n\}$, divide the data into $k = \lfloor n/m \rfloor$ blocks of length m (and a possibly remaining block of smaller size which has to be omitted). Denote the maximum within the i-th disjoint block of observations of size m as

$$M_{i,m} = \max(X_{(i-1)m+1}, ..., X_{im}),$$

for $i = 1, ..., k$. By the domain of attraction condition (4), for large block sizes m, the sample of block maxima $\{M_{1,m}, ..., M_{k,m}\}$ is approximately Fréchet distributed. Let G_{γ,a_m} denote the Fréchet distribution with shape parameter γ and scale parameter a_m, defined by its cumulative distribution function

$$G_{\gamma,a_m}(x) = \exp\left(-(x/a_m)^{-1/\gamma}\right), \tag{22}$$

for $x > 0$. We apply the maximum likelihood procedure to fit the Fréchet distribution (Section 4.1).

Both under a simplified model without serial dependence (Section 4.3) and under a general model with serial dependence (Section 4.4), we can approximate the return level by the same quantile of the Fréchet distribution. Consequently, the estimator for the return level will be the same in both cases (Section 4.2). The quantile function of the Fréchet distribution is given by

$$G_{\gamma,a_m}^{-1}(x) = a_m(-\log(x))^{-\gamma}, \tag{23}$$

for $x > 0$.

4.1 Maximum likelihood estimation

We derive the maximum likelihood estimators for the parameters of the Fréchet distribution, G_{γ,a_m}, given in (22). Let $\boldsymbol{x} = (x_1, ..., x_k) \in (0,\infty)^k$ be a sample vector to which the Fréchet distribution is to be fitted. The maximum likelihood estimator is defined as

$$(\hat{\gamma}_{\mathrm{ML}}, \hat{a}_m) = \underset{(\gamma,a_m)\in(0,\infty)^2}{\arg\max} \sum_{i=1}^{k} \ell_{\gamma,a_m}(x_i),$$

where ℓ_{γ,a_m} is the log-likelihood function equal to

$$\ell_{\gamma,a_m}(x) = -\log(\gamma) - \log(a_m) - (x/a_m)^{-1/\gamma} - (1/\gamma + 1)\log(x/a_m),$$

for $x > 0$. By Lemma 2.1 in Bücher and Segers (2018b), the maximizer exists and is unique as soon as the scalars $x_1, ..., x_k$ are not all identical. By Theorem 2.3 in Bücher and Segers (2018b), the maximum likelihood estimator is consistent in the sense that under mild conditions as $n \to \infty$,

$$(\hat{\gamma}_{\mathrm{ML}}, \hat{a}_m/a_m) \xrightarrow{P} (\gamma, 1). \tag{24}$$

The maximum likelihood estimators for the extreme value index and the scale, $\hat{\gamma}_{\mathrm{ML}}$ and $\hat{a}_m$, are obtained by solving the likelihood equations. The likelihood equations are

$$\sum_{i=1}^{k} \left(-\frac{1}{a_m} - \frac{1}{\gamma} x_i^{-1/\gamma} a_m^{1/\gamma-1} + \left(\frac{1}{\gamma} + 1\right) \frac{1}{a_m} \right) = 0, \tag{25}$$

$$\sum_{i=1}^{k} \left(-\frac{1}{\gamma} - (x_i/a_m)^{-1/\gamma} \log(x_i/a_m) \frac{1}{\gamma^2} + \frac{1}{\gamma^2} \log(x_i/a_m) \right) = 0. \tag{26}$$

Equation (25) can be simplified to

$$\frac{1}{a_m} \frac{1}{\gamma} \left(k - \sum_{i=1}^{k} (x_i/a_m)^{-1/\gamma} \right) = 0,$$

which implies that

$$k - a_m^{1/\gamma} \sum_{i=1}^{k} x_i^{-1/\gamma} = 0. \tag{27}$$

The maximum likelihood estimator for the scale parameter, $\hat{a}_m$, is equal to the explicit solution of (27) once $\hat{\gamma}_{\mathrm{ML}}$ is solved,

$$\hat{a}_m = \left(\frac{1}{k} \sum_{i=1}^{k} x_i^{-1/\hat{\gamma}_{\mathrm{ML}}} \right)^{-\hat{\gamma}_{\mathrm{ML}}}.$$

The likelihood equation (26) can be simplified to

$$\frac{1}{k} \sum_{i=1}^{k} \log(x_i) = \gamma + a_m^{1/\gamma} \frac{1}{k} \sum_{i=1}^{k} x_i^{-1/\gamma} \log(x_i) + \log(a_m) \left(1 - a_m^{1/\gamma} \frac{1}{k} \sum_{i=1}^{k} x_i^{-1/\gamma} \right). \tag{28}$$

By using (27), the last term in (28) becomes zero and we get that

$$\gamma + \frac{k}{\sum_{i=1}^{k} x_i^{-1/\gamma}} \frac{1}{k} \sum_{i=1}^{k} x_i^{-1/\gamma} \log(x_i) - \frac{1}{k} \sum_{i=1}^{k} \log(x_i) = 0. \tag{29}$$

The maximum likelihood estimator for the extreme value index, $\hat{\gamma}_{\mathrm{ML}}$, is equal to the unique zero of (29).

4.2 Motivating the estimator

From (1), (7) and (22), we deduce that as $m \to \infty$,

$$\mathbb{P}(M_m \leq R_{m,y}) = 1 - 1/y \approx \exp\left(-(R_{m,y}/\tilde{a}_m)^{-1/\gamma} \right) = G_{\gamma,\tilde{a}_m}(R_{m,y}),$$

where $\tilde{a}_m = a_m \theta^\gamma$ with $\theta \in (0,1]$ the extremal index. Hence, both under independence ($\theta = 1$) and under serial dependence, the return level is approximately equal to the $1/y$-th quantile of the Fréchet distribution.

By using (23), we propose the following estimator for the return level

$$\hat{R}_{m,y} := \hat{a}_m c_y^{-\hat{\gamma}_{\mathrm{ML}}}, \tag{30}$$

with $c_y = -\log(1 - 1/y)$ and where $(\hat{\gamma}_{\mathrm{ML}}, \hat{a}_m)$ is the maximum likelihood estimator studied in Section 4.1.

For both the i.i.d. case (Section 4.3) and the serial dependence case (Section 4.4), the structure is similar to the POT approach in Section 3. First, we present the conditions and the auxiliary lemma necessary for the asymptotic theory. Then we state and prove the main theorem on the asymptotic normality of the estimator for the return level.

4.3 I.i.d. case

Let $\{X_1, X_2, ...\}$ be an i.i.d. sequence of random variables with a common distribution function F. We assume that this distribution function belongs to the Fréchet domain of attraction ($\gamma > 0$).

4.3.1 Conditions for asymptotic theory

We present the conditions for establishing the asymptotic normality of the return level estimator (30) in Section 4.3.3. Similar to the POT approach in Section 3.1.2, we need a second order reinforcement of the domain of attraction condition (4) in combination with a growth restriction on the number of blocks.

Second order condition Let $V := (1/(-\log F))^{\leftarrow}$ be the left-continuous inverse function of $1/(-\log F)$. Suppose that there exist a positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a real number $\rho < 0$ such that

$$\lim_{t\to\infty} \frac{\frac{V(tx)}{V(t)} - x^{\gamma}}{A(t)} = x^{\gamma}\frac{x^{\rho}-1}{\rho}, \tag{31}$$

for all $x > 0$.

The second order condition captures the speed of convergence in the domain of attraction condition (4). By (4.3) in Bücher and Segers (2018b), p. 10, the scaling constants in (4) may be chosen as any sequence a_m, $m \in \mathbb{N}$ that satisfies

$$\lim_{m\to\infty} m(-\log F(a_m)) = 1. \tag{32}$$

Hence, in the i.i.d. case, we choose $a_m = V(m)$.

Condition on k Suppose that the number of blocks k satisfies

$$\lim_{n\to\infty} \sqrt{k}A\left(\frac{n}{k}\right) = 0. \tag{33}$$

The condition (33) imposes an upper bound on the speed at which k goes to infinity. As it appears below, it assumes away the asymptotic bias for the estimator for the return level.

4.3.2 Auxiliary lemma

The following lemma is useful in the proof of the asymptotic normality of the estimator for the return level (Theorem 4.2).

Lemma 4.1 (Theorem 4.2 and Lemma B.3, Bücher and Segers (2018b))**.** *Suppose that $\{X_1, X_2, ...\}$ is an i.i.d. sequence of random variables with a common distribution function F. Assume that F satisfies the second order condition (31)* *with parameters $\gamma > 0$ and $\rho < 0$. Assume that both $m \to \infty$ and $k \to \infty$ as $n \to \infty$. Suppose that k satisfies (33). Then as $n \to \infty$,*

$$\sqrt{k}\begin{pmatrix} \frac{1}{\hat{\gamma}_{ML}} - \frac{1}{\gamma} \\ \frac{\hat{a}_m}{V(m)} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\gamma\right),$$

where

$$\Sigma_\gamma = \frac{6}{\pi^2}\begin{pmatrix} \gamma^{-2} & (g-1) \\ (g-1) & \gamma^2\left((1-g)^2 + \pi^2/6\right) \end{pmatrix},$$

with $g = 0.5772...$ the Euler-Mascheroni constant.

4.3.3 Main result

The asymptotic normality of the return level estimator (30) is given by the following theorem.

Theorem 4.2. *Suppose that $\{X_1, X_2, ...\}$ is an i.i.d. sequence of random variables with a common distribution function F. Assume that F satisfies the second order condition (31) with parameters $\gamma > 0$ and $\rho < 0$. Assume that both $m \to \infty$ and $k \to \infty$ as $n \to \infty$. Suppose that k satisfies (33). Then as $n \to \infty$,*

$$\sqrt{k}\left(\frac{\hat{R}_{m,y}}{R_{m,y}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\gamma)'\Sigma_\gamma\beta(y,\gamma)\right),$$

where $\beta(y,\gamma) = (\gamma^2 \log(c_y), 1)'$ and with Σ_γ as defined in Lemma 4.1.

Proof. Consider the expansion

$$\begin{aligned}\frac{\hat{R}_{m,y}}{R_{m,y}} &= \frac{\hat{a}_m c_y^{-\hat{\gamma}_{\mathrm{ML}}}}{V(m)c_y^{-\gamma}} \cdot \frac{V(m)c_y^{-\gamma}}{R_{m,y}} \\ &=: I_1 \cdot I_2.\end{aligned}$$

For I_1, consider the expansion

$$\begin{aligned}\sqrt{k}\,(I_1 - 1) &= \sqrt{k}\left(\frac{\hat{a}_m}{V(m)} - 1\right)\left(c_y^{\gamma-\hat{\gamma}_{\mathrm{ML}}} - 1\right) + \sqrt{k}(c_y^{\gamma-\hat{\gamma}_{\mathrm{ML}}} - 1) + \sqrt{k}\left(\frac{\hat{a}_m}{V(m)} - 1\right) \\ &=: I_{11} + I_{12} + I_{13}.\end{aligned}$$

From (24), we get that $I_{11} = o_{\mathbb{P}}(1)$ as $n \to \infty$. Denote the limit of $\sqrt{k}\left(\frac{1}{\hat{\gamma}_{\mathrm{ML}}} - \frac{1}{\gamma}\right)$ as Γ. Then according to Cramér's delta method (Cramér, 1946), we get that as $n \to \infty$,

$$I_{12} \xrightarrow{d} \gamma^2 \log(c_y)\Gamma.$$

By applying Lemma 4.1, we get that as $n \to \infty$,

$$I_{12} + I_{13} = \left(\gamma^2 \log(c_y), 1\right) \cdot \sqrt{k}\begin{pmatrix} \frac{1}{\hat{\gamma}_{\mathrm{ML}}} - \frac{1}{\gamma} \\ \frac{\hat{a}_m}{V(m)} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\gamma)'\Sigma_\gamma\beta(y,\gamma)\right).$$

By combining the limit properties of I_{11}, I_{12} and I_{13}, we get that as $n \to \infty$,

$$\sqrt{k}\,(I_1 - 1) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\gamma)'\Sigma_\gamma \beta(y,\gamma)\right).$$

Lastly, we deal with the term I_2. The second order condition (31) implies that as $n \to \infty$,

$$\frac{\frac{V(mx)}{V(m)} - x^\gamma}{A(m)} \to x^\gamma \frac{x^\rho - 1}{\rho}.$$

By using Vervaat's lemma (see de Haan and Ferreira (2007), Lemma A.0.2), we get that as $n \to \infty$,

$$\frac{\frac{1}{-m \log F(V(m)x)} - x^{1/\gamma}}{A(m)} \to -x^{1/\gamma} \frac{x^{\rho/\gamma} - 1}{\rho\gamma}.$$

According to Cramér's delta method, we get that as $n \to \infty$,

$$\frac{-m \log F\left(V(m)x\right) - x^{-1/\gamma}}{A(m)} \to x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\rho\gamma}. \tag{34}$$

We can find a x_0 such that $x = 1 - 1/y \geq x_0$. Then the right side of (34) is uniformly bounded for all $x > x_0$. By using (33), we get that as $n \to \infty$,

$$\sqrt{k}\left(-m \log F(V(m)x) - x^{-1/\gamma}\right) \to 0.$$

By using again the Cramér's delta method, we get that as $n \to \infty$,

$$\sqrt{k}\left(F^m(V(m)x) - \exp\left(-x^{-1/\gamma}\right)\right) \to 0.$$

By again applying Vervaat's lemma and then set $x = 1 - 1/y$, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{R_{m,y}}{V(m)} - c_y^{-\gamma}\right) \to 0.$$

Hence, we get that as $n \to \infty$,

$$\sqrt{k}(I_2 - 1) \to 0.$$

The theorem is proved by combining the limit properties of the two terms in the expansion. According to Cramèr's delta method, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{\hat{R}_{m,y}}{R_{m,y}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\gamma)'\Sigma_\gamma \beta(y,\gamma)\right).$$

$\square$

4.4 Serial dependence case

Let $\{X_1, X_2, ...\}$ be a strictly stationary time series with a common distribution function F. We assume that this distribution function belongs to the Fréchet domain of attraction ($\gamma > 0$). Additionally, we assume that the series possesses a positive extremal index ($\theta \in (0, 1)$). Hence, the limit relation (7) holds with $\tilde{a}_m = a_m \theta^\gamma$.

4.4.1 Conditions for asymptotic theory

We present the regularity conditions for establishing the asymptotic normality of the return level estimator (30) in Section 4.4.3. The serial dependence structure follows from the so-called α-mixing conditions. The α-mixing conditions have been introduced by Rosenblatt (1956) as follows. The sequence $\{X_1, X_2, ...\}$ is called α-mixing (or strongly mixing) if as $\ell \to \infty$,

$$\alpha(\ell) := \sup\left(|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{B}_{-\infty}^0, B \in \mathcal{B}_\ell^\infty\right) \to 0,$$

where $\mathcal{B}_{-\infty}^0$ and $\mathcal{B}_\ell^\infty$ denote the σ-fields generated by $(X_i)_{i\le 0}$ and $(X_i)_{i\ge \ell}$, respectively. The coefficients $\alpha(\ell)$ are called the α-mixing constants of the sequence.

The following four conditions are Conditions 3.2, 3.3, 3.4 and 3.5 in Bücher and Segers (2018b).

Regularity conditions Suppose that $\lim_{\ell\to\infty} \alpha(\ell) = 0$ and there exist a constant $\varepsilon > 0$, some $\nu > 2/\varepsilon$ and a constant $d > 0$ such that

Condition 4.1. For every $c \in (0, \infty)$,

$$\lim_{n\to\infty} \mathbb{P}\left(\min\left(M_{1,m}, ..., M_{k,m}\right) \le c\right) = 0;$$

Condition 4.2.

$$\lim_{n\to\infty} k^{1+\varepsilon}\alpha(m) = 0;$$

Condition 4.3.

$$\limsup_{n\to\infty} \mathbb{E}(g_{\nu,\gamma}((M_m \vee 1)/\tilde{a}_m)) < \infty,$$

where $g_{\nu,\gamma}(x) = \left(x^{-1/\gamma} 1_{\{x\le e\}} + \log(x) 1_{\{x>e\}}\right)^{2+\nu}$;

Condition 4.4. For $j = 1, 2, 3$

$$\lim_{n\to\infty} \sqrt{k}\left(\mathbb{E}(f_j((M_m \vee d)/\tilde{a}_m)) - \int_0^\infty f_j(x) f_\gamma(x) \mathrm{d}x\right) = 0,$$

where

$$f_1(x) = x^{-1/\gamma}\log(x), \qquad f_2(x) = x^{-1/\gamma}, \qquad f_3(x) = \log(x).$$

and

$$f_\gamma(x) = \frac{1}{\gamma}\exp\left(-x^{-1/\gamma}\right) x^{-1/\gamma-1}. \tag{35}$$

We discuss these conditions one by one. Condition 4.1 ensures that all block maxima diverge (Bücher & Segers, 2018b). To control the serial dependence within the time series, condition 4.2 requires that the α-mixing coefficients decay sufficiently fast (Bücher & Segers, 2018b). Furthermore, this condition ensures that the block sizes m are sufficiently large such that maxima over large disjoint blocks are asymptotically independent. Condition 4.3 is an asymptotic bound on certain moments of the block maxima (Bücher & Segers, 2018b). Condition 4.4 guarantees that the asymptotic distribution in Theorem 4.4 is centred and the return level estimator is asymptotically unbiased.

4.4.2 Auxiliary lemma

The following lemma is useful in the proof of the asymptotic normality of the estimator for the return level (Theorem 4.4).

Lemma 4.3 (Theorem 3.6, Bücher and Segers (2018b))**.** *Suppose that $\{X_1, X_2, ...\}$ is a strictly stationary α-mixing time series with common distribution function F. Assume that $F \in \mathcal{D}(G_\gamma)$ with $\gamma > 0$ and that the series has extremal index $\theta \in (0,1)$. Assume that both $m \to \infty$ and $k \to \infty$ as $n \to \infty$. Suppose that the regularity conditions 4.1-4.4 hold. Then as $n \to \infty$,*

$$\sqrt{k}\begin{pmatrix} \frac{1}{\hat{\gamma}_{ML}} - \frac{1}{\gamma} \\ \frac{\hat{a}_m}{\tilde{a}_m} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_\gamma\right),$$

where $\tilde{a}_m = a_m \theta^\gamma$ and with Σ_γ as defined in Lemma 4.1.

4.4.3 Main result

The asymptotic normality of the return level estimator (30) is given by the following theorem.

Theorem 4.4. *Suppose that $\{X_1, X_2, ...\}$ is a strictly stationary α-mixing time series with common distribution function F. Assume that $F \in \mathcal{D}(G_\gamma)$ with $\gamma > 0$ and that the series has extremal index $\theta \in (0,1)$. Assume that both $m \to \infty$ and $k \to \infty$ as $n \to \infty$. Suppose that there exists a positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ such that*

$$\mathbb{P}\left(\frac{M_m}{\tilde{a}_m} \leq x\right) - \exp\left(-x^{-1/\gamma}\right) = O\left(A\left(m\right)\right), \tag{36}$$

where $\tilde{a}_m = a_m \theta^\gamma$. Assume in addition that k satisfies

$$\lim_{n\to\infty} \sqrt{k} A\left(\frac{n}{k}\right) = 0. \tag{37}$$

Suppose that the regularity conditions 4.1-4.4 hold. Then as $n \to \infty$,

$$\sqrt{k}\left(\frac{\hat{R}_{m,y}}{R_{m,y}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\gamma)' \Sigma_\gamma \beta(y,\gamma)\right),$$

where $\beta(y,\gamma) = (\gamma^2 \log(c_y), 1)'$ and with Σ_γ as defined in Lemma 4.1.

Proof. Consider the expansion

$$\frac{\hat{R}_{m,y}}{R_{m,y}} = \frac{\hat{a}_m c_y^{-\hat{\gamma}_{\mathrm{ML}}}}{\tilde{a}_m c_y^{-\gamma}} \cdot \frac{\tilde{a}_m c_y^{-\gamma}}{R_{m,y}}$$
$$=: I_1 \cdot I_2.$$

Similar to the first part of the proof of Theorem 4.2, we get that as $n \to \infty$,

$$\sqrt{k}\,(I_1 - 1) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\gamma)' \Sigma_\gamma \beta(y,\gamma)\right).$$

From (36) and (37), we get that as $n \to \infty$,

$$\sqrt{k}\left(\mathbb{P}\left(\frac{M_m}{\tilde{a}_m} \le x\right) - \exp\left(-x^{-1/\gamma}\right)\right) \to 0.$$

By applying Vervaat's lemma (see de Haan and Ferreira (2007), Lemma A.0.2) and then set $x = 1 - 1/y$, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{R_{m,y}}{\tilde{a}_m} - c_y^{-\gamma}\right) \to 0.$$

Hence, we get that as $n \to \infty$,

$$\sqrt{k}(I_2 - 1) \to 0.$$

The theorem is proved by combining the limit properties of the two terms in the expansion. According to Cramèr's delta method, we get that as $n \to \infty$,

$$\sqrt{k}\left(\frac{\hat{R}_{m,y}}{R_{m,y}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \beta(y,\gamma)' \Sigma_\gamma \beta(y,\gamma)\right).$$

□

Remark 4.5. Both under independence and serial dependence, both the limits of the maximum likelihood estimators for the parameters of the Fréchet distribution, $\hat{a}_m$ and $\hat{\gamma}_{\mathrm{ML}}$, contribute to the final asymptotic limit of the estimator for the return level. This situation is similar to the POT approach where both the limit of the order statistic and the limit of the Hill estimator contribute to the final asymptotic limit, as mentioned in Remark 3.10.

4.4.4 Discussion on conditions

We finish this section with some remarks on the conditions used in the serial dependence case. Conditions (36) and (37) together almost imply the regularity condition 4.4. In fact, if we assume a stronger version of (36), the condition 4.4 is implied as shown in the following lemma.

Lemma 4.6. *Let $H_m(x) = \mathbb{P}\left(\frac{M_m}{\tilde{a}_m} \le x\right)$ and $G(x) = G_{\gamma,1}(x)$. Suppose that there exist a positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a function $\tilde{G}$ such that for any constant $d > 0$ and $x \ge d/\tilde{a}_m$,*

$$\left|\frac{H_m(x) - G(x)}{A(m)}\right| \le \tilde{G}(x) \tag{38}$$

and

$$\int_0^\infty \tilde{G}(x) x^c \mathrm{d}x < \infty, \tag{39}$$

for all $c \le -1$. Assume in addition that k satisfies (37). Then the regularity condition 4.4 holds.

Proof. For $j = 1, 2, 3$, write

$$\begin{aligned}
A(f_j) &:= \int_0^\infty f_j\left(x \vee \frac{d}{\tilde{a}_m}\right) \mathrm{d}H_m(x) - \int_0^\infty f_j(x)\mathrm{d}G(x) \\
&= f_j\left(\frac{d}{\tilde{a}_m}\right) H_m\left(\frac{d}{\tilde{a}_m}\right) - \int_0^{d/\tilde{a}_m} f_j(x)\mathrm{d}G(x) + \int_{d/\tilde{a}_m}^\infty f_j(x)\mathrm{d}H_m(x) - \int_{d/\tilde{a}_m}^\infty f_j(x)\mathrm{d}G(x) \\
&=: A_1(f_j) - A_2(f_j) + A_3(f_j) - A_4(f_j).
\end{aligned}$$

By partial integration,

$$\begin{aligned}
A_2(f_j) &= \Big[f_j(x)G(x)\Big]_0^{d/\tilde{a}_m} - \int_0^{d/\tilde{a}_m} G(x)\mathrm{d}f_j(x) \\
&= f_j\left(\frac{d}{\tilde{a}_m}\right) G\left(\frac{d}{\tilde{a}_m}\right) - \int_0^{d/\tilde{a}_m} G(x)\mathrm{d}f_j(x).
\end{aligned}$$

For $A_3(f_j)$ and $A_4(f_j)$, we first consider $j = 1, 2$ with $f_1(x) = x^{-1/\gamma}\log(x)$ and $f_2(x) = x^{-1/\gamma}$. By partial integration, these two integrals can be calculated as follows

$$\begin{aligned}
A_3(f_j) &= \Big[f_j(x)H_m(x)\Big]_{d/\tilde{a}_m}^\infty - \int_{d/\tilde{a}_m}^\infty H_m(x)\mathrm{d}f_j(x) \\
&= -f_j\left(\frac{d}{\tilde{a}_m}\right) H_m\left(\frac{d}{\tilde{a}_m}\right) - \int_{d/\tilde{a}_m}^\infty H_m(x)\mathrm{d}f_j(x),
\end{aligned}$$

$$\begin{aligned}
A_4(f_j) &= \Big[f_j(x)G(x)\Big]_{d/\tilde{a}_m}^\infty - \int_{d/\tilde{a}_m}^\infty G(x)\mathrm{d}f_j(x) \\
&= -f_j\left(\frac{d}{\tilde{a}_m}\right) G\left(\frac{d}{\tilde{a}_m}\right) - \int_{d/\tilde{a}_m}^\infty G(x)\mathrm{d}f_j(x).
\end{aligned}$$

Combining the expressions for $A_1(f_j)$-$A_4(f_j)$, we get that for $j = 1, 2$,

$$A(f_j) = \int_0^{d/\tilde{a}_m} G(x)\mathrm{d}f_j(x) - \int_{d/\tilde{a}_m}^\infty \left(H_m(x) - G(x)\right)\mathrm{d}f_j(x). \tag{40}$$

Secondly, we consider $j=3$ with $f_3(x)=\log(x)$. By again partial integration, $A_3(f_3)$ and $A_4(f_3)$ can be calculated as follows

$$
\begin{aligned}
A_3(f_3) &= -\int_{d/\tilde{a}_m}^{\infty} f_3(x)\mathrm{d}(1-H_m(x)) \\
&= -\Big[f_3(x)(1-H_m(x))\Big]_{d/\tilde{a}_m}^{\infty} + \int_{d/\tilde{a}_m}^{\infty}(1-H_m(x))\mathrm{d}f_3(x) \\
&= f_3\left(\frac{d}{\tilde{a}_m}\right)\left(1-H_m\left(\frac{d}{\tilde{a}_m}\right)\right) + \int_{d/\tilde{a}_m}^{\infty}(1-H_m(x))\mathrm{d}f_3(x),
\end{aligned}
$$

$$
\begin{aligned}
A_4(f_3) &= -\int_{d/\tilde{a}_m}^{\infty} f_3(x)\mathrm{d}(1-G(x)) \\
&= -\Big[f_3(x)(1-G(x))\Big]_{d/\tilde{a}_m}^{\infty} + \int_{d/\tilde{a}_m}^{\infty}(1-G(x))\mathrm{d}f_3(x) \\
&= f_3\left(\frac{d}{\tilde{a}_m}\right)\left(1-G\left(\frac{d}{\tilde{a}_m}\right)\right) + \int_{d/\tilde{a}_m}^{\infty}(1-G(x))\mathrm{d}f_3(x).
\end{aligned}
$$

Combining the expressions for $A_1(f_j)$-$A_4(f_j)$, we get that

$$
\begin{aligned}
A(f_3) &= \int_0^{d/\tilde{a}_m} G(x)\mathrm{d}f_3(x) + \int_{d/\tilde{a}_m}^{\infty}\left((1-H_m(x))-(1-G(x))\right)\mathrm{d}f_3(x) \\
&= \int_0^{d/\tilde{a}_m} G(x)\mathrm{d}f_3(x) - \int_{d/\tilde{a}_m}^{\infty}\left(H_m(x)-G(x)\right)\mathrm{d}f_3(x),
\end{aligned}
$$

which is equivalent to (40).

Hence, for $j=1,2,3$, we get that

$$
\sqrt{k}\cdot A(f_j) = \sqrt{k}\int_0^{d/\tilde{a}_m} G(x)\mathrm{d}f_j(x) - \sqrt{k}\int_{d/\tilde{a}_m}^{\infty}\left(H_m(x)-G(x)\right)\mathrm{d}f_j(x). \quad (41)
$$

From the proof of Lemma 5.1 in Bücher and Segers (2018b), p. 33, it follows that the first integral in (41) converges to zero as $n\to\infty$.

From (38), it follows that

$$
\begin{aligned}
\left|\sqrt{k}\int_{d/\tilde{a}_m}^{\infty}\left(H_m(x)-G(x)\right)f_j(x)\mathrm{d}x\right| &\le \sqrt{k}A\left(\frac{n}{k}\right)\int_{d/\tilde{a}_m}^{\infty}\left|\frac{H_m(x)-G(x)}{A\left(\frac{n}{k}\right)}\right| f_j(x)\mathrm{d}x \\
&\le \sqrt{k}A\left(\frac{n}{k}\right)\int_0^{\infty}\tilde{G}(x)f_j(x)\mathrm{d}x. \quad (42)
\end{aligned}
$$

For $j=1,2$ it holds that, up to a factor, $f_j'(x)\le x^{-1/\gamma-1\pm\delta}$ for some $\delta<1/\gamma$. The symbol $\pm$ means taking the suitable sign according to whether x is higher or lower than 1. If x is above 1, we take $+\delta$ and if x is below 1 we take $-\delta$. Hence, we get that

$$
\int_0^{\infty}\tilde{G}(x)\mathrm{d}f_j(x) \le \int_0^{\infty}\tilde{G}(x)x^{-1/\gamma-1\pm\delta}\mathrm{d}x < \infty. \quad (43)
$$

Notice that we use (39) with $c = -1/\gamma - 1 \pm \delta \le -1$. For $j = 3$, by using again (39) with $c = -1$, we get that

$$\int_0^\infty \tilde{G}(x)\mathrm{d}f_3(x) = \int_0^\infty \tilde{G}(x)x^{-1}\mathrm{d}x < \infty. \tag{44}$$

From (37), (42), (43) and (44), it follows that for $j = 1, 2, 3$, the second integral in (41) converges to zero as $n \to \infty$.

Hence, we conclude that for $j = 1, 2, 3$, $\lim_{n\to\infty} \sqrt{k} \cdot A(f_j) = 0$ and the lemma is proved. $\square$

The essential part to prove the regularity condition 4.4, is that conditions (38) and (39) hold. Then we can prove that the second integral in (41) converges to zero as $n \to \infty$. For the i.i.d. case, the second order condition (31) guarantees that conditions (38) and (39) are satisfied as shown in Example 4.7. Bücher and Segers (2018b) also prove that under i.i.d. the regularity condition 4.4 is satisfied (see pp. 25-28 in their paper). However, our alternative proof of Lemma 4.6 is easier to understand and shows the essence of the proof, namely that conditions (38) and (39) need to hold.

Example 4.7. Under i.i.d, it holds that $\tilde{a}_m = V(m)$. By using (32) and (34), we get that as $n \to \infty$,

$$\frac{x^{1/\gamma}\frac{-\log F(V(m)x)}{-\log F(V(m))} - 1}{A(m)} \to \frac{x^{\rho/\gamma} - 1}{\rho\gamma}.$$

By applying Theorem B.2.18 in de Haan and Ferreira (2007), accredited to Drees (1998a), we get that for all c, $\varepsilon > 0$,

$$\left| \frac{x^{1/\gamma}\frac{-\log F(V(m)x)}{-\log F(V(m))} - 1}{A_0(m)} - \frac{x^{\rho/\gamma} - 1}{\rho\gamma} \right| \le c x^{\rho/\gamma \pm \varepsilon},$$

for $V(m)x > d$. Or, equivalently,

$$\left| \frac{x^{1/\gamma}\frac{-\log F(V(m)x)}{-\log F(V(m))} - 1}{A_0(m)} \right| \le c_1 x^{\rho/\gamma \pm \varepsilon} + c_2,$$

which implies

$$\left| \frac{\log H_m(x) + x^{-1/\gamma}}{A_0(m)} \right| \le x^{-1/\gamma}\left(c_1 x^{\rho/\gamma \pm \varepsilon} + c_2\right)$$

and

$$\left| \frac{H_m(x) - G(x)}{A_0(m)} \right| \le G(x)x^{-1/\gamma}\left(\tilde{c}_1 x^{\rho/\gamma \pm \varepsilon} + \tilde{c}_2\right),$$

for $V(m)x > d$. In the i.i.d. case, we get that condition (38) holds with

$$\begin{aligned}\tilde{G}(x) &= \exp\left(-x^{-1/\gamma}\right) x^{-1/\gamma} \left(\tilde{c}_1 x^{\rho/\gamma \pm \varepsilon} + \tilde{c}_2\right) \\ &= x f_\gamma(x) \left(\tilde{c}_1 x^{\rho/\gamma \pm \varepsilon} + \tilde{c}_2\right),\end{aligned}$$

with $f_\gamma(x) = G'(x)$, see also (35) in the regularity condition 4.4. Hence, we get that

$$\int_0^\infty \tilde{G}(x) x^c \mathrm{d}x = \int_0^\infty x^{c+1} f_\gamma(x) \left(\tilde{c}_1 x^{\rho/\gamma \pm \varepsilon} + \tilde{c}_2\right) \mathrm{d}x.$$

Let Z be a Fréchet distributed random variable. The rth moment of the Fréchet distribution is finite if $r < 1/\gamma$ (Zayed & Butt, 2017). Hence, we get that for all $c \leq -1$,

$$\int_0^\infty \tilde{G}(x) x^c \mathrm{d}x = \tilde{c}_1 \mathbb{E}\left(Z^{c+1+\rho/\gamma \pm \varepsilon}\right) + \tilde{c}_2 \mathbb{E}\left(Z^{c+1}\right).$$

Notice that by choosing ε small $c + 1 + \rho/\gamma \pm \varepsilon < 1/\gamma$ and $c + 1 < 1/\gamma$ to ensure that these moments of the Fréchet distribution are finite and thus (39) holds.

5 Comparison between the POT approach and the BM method

We provide a theoretical comparison between the POT approach and the BM method. For i.i.d. observations (Section 5.1), the autoregressive model (Section 5.2.1) and the moving average model (Section 5.2.2), we derive explicit expressions of the asymptotic variances for the return level estimators based on the two methods and make a comparison. To invoke the asymptotic theories, we verify that the (regularity) conditions hold for the models.

A key tuning parameter in the statistical analysis is the intermediate sequence k, which is defined as either the number of large order statistics in the POT approach, or the number of blocks in the BM method. We use the same k level for both the POT approach and the BM method. Hence, the comparison is made at the level of asymptotic variance.

5.1 I.i.d. case

We compare the POT approach with the BM method under the assumption that the sample of observations $\{X_1, X_2, ...\}$ is i.i.d. with a common distribution function F. We assume that this distribution function belongs to the Fréchet domain of attraction ($\gamma > 0$) and that F satisfies the second order conditions (13) and (31) for the POT approach and the BM method, respectively. The second order auxiliary functions $|A_{\mathrm{POT}}|$ and $|A_{\mathrm{BM}}|$ are necessarily regularly varying with index ρ_{POT} and ρ_{BM}, respectively (Bücher & Zhou, 2018). For the theoretical comparison, we consider the case $\rho = \rho_{\mathrm{POT}} = \rho_{\mathrm{BM}} \in [-1, 0)$ such that the comparison is made at the level of asymptotic variance. We refer to

Section 2 in Bücher and Zhou (2018) for a detailed comparison of the second order conditions. Furthermore, we assume that as $n \to \infty$, $k \to \infty$, $k/n \to 0$, $m \to \infty$, and $\sqrt{k} A_z\left(\frac{n}{k}\right) \to 0$ with $z \in \{\text{POT}, \text{BM}\}$.

Let $\tilde{\sigma}^2_{\text{POT}}$ denote the asymptotic variance of the estimator for the return level based on the POT approach under independence. From Theorem 3.4, we get that as $n \to \infty$,

$$\tilde{\sigma}^2_{\text{POT}} = \gamma^2 \left(1 + \log^2(c_y)\right),$$

with $c_y = -\log(1 - 1/y)$. Let σ^2_{BM} denote the asymptotic variance of the estimator for the return level based on the BM method. From Theorem 4.2, we get that as $n \to \infty$,

$$\sigma^2_{\text{BM}} = \gamma^2 \left(\frac{6}{\pi^2}\left(\log^2(c_y) + 2\log(c_y)(g-1) + (1-g)^2\right) + 1\right),$$

with $c_y = -\log(1 - 1/y)$ and $g = 0.5772...$ the Euler-Mascheroni constant.

For $y > 1$, we compare the asymptotic variances for the two methods by calculating the following ratio

$$\frac{\tilde{\sigma}^2_{\text{POT}}}{\sigma^2_{\text{BM}}} = \frac{1 + \log^2(c_y)}{\frac{6}{\pi^2}\left(\log^2(c_y) + 2\log(c_y)(g-1) + (1-g)^2\right) + 1}, \tag{45}$$

with $c_y = -\log(1 - 1/y)$ and $g = 0.5772...$ the Euler-Mascheroni constant.

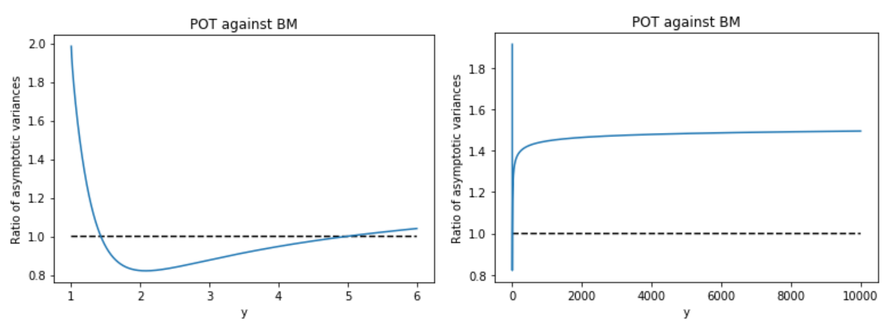


Figure 1: Ratio between the asymptotic variances for the POT approach and the BM method under independence, for $y \in (1, 6]$ (left) and $y \in (1, 10000]$ (right). The dashed line corresponds to the ratio equal to one.

In Figure 1, we plot this ratio against the parameter y. From the left figure, we observe that only for $y \in (1.43, 4.98)$ the ratio in (45) is less than one and consequently σ^2_{BM} is higher than $\tilde{\sigma}^2_{\text{POT}}$. Otherwise, σ^2_{BM} is lower than $\tilde{\sigma}^2_{\text{POT}}$ and therefore the BM method outperforms the POT approach in terms of having a lower asymptotic variance. Notice that the difference between the two asymptotic variances is not very large as $\tilde{\sigma}^2_{\text{POT}}/\sigma^2_{\text{BM}} \to \pi^2/6$ as $y \to \infty$, i.e. $\tilde{\sigma}^2_{\text{POT}} < \frac{\pi^2}{6} \cdot \sigma^2_{\text{BM}}$ for all $y > 1.08$.

5.2 Serial dependence case

We compare the POT approach with the BM method under the AR(1) model (Section 5.2.1) and under the MA(1) model (Section 5.2.2). In the POT framework, we model the serial dependence by the β-mixing condition and the regularity conditions 3.1-3.5. In order to establish the asymptotic normality of the estimator for the extremal index, we need Condition 2.1 and (2.2) in Berghaus and Bücher (2018) to hold. In the BM framework, we model the serial dependence by the α-mixing condition and the regularity conditions 4.1-4.4. Notice that β-mixing implies α-mixing (see Bradley et al. (2005), Section 2.1).

Condition 2.1 in Berghaus and Bücher (2018) consists of seven separate conditions which we will not discuss in detail. However, we make a few remarks. Condition 2.1 (iii) and regularity condition 4.2 are both conditions on the α-mixing constant. In fact, Condition 2.1 (iii) is more restrictive than regularity condition 4.2. Condition 2.1 (iv) is comparable with regularity condition 3.3 which in a slightly different form concerns only the tail. Hence, regularity condition 3.3 is more restrictive than Condition 2.1 (iv). Condition 2.1 (v) is similar to regularity condition 4.1 and both conditions ensure that all block maxima diverge. In fact, Condition 2.1 (v) is more restrictive than regularity condition 4.1.

Both the AR(1) model and the MA(1) model satisfy the regularity conditions 3.1-3.5 (see Drees (2003), Section 3.2). We assume that Condition 2.1 (i)-(iii), (v)-(vii) and (2.2) in Berghaus and Bücher (2018) hold for both models. From Section 5.1 in Bücher and Segers (2018b), we quote: "For many stationary time series models, the distribution of the sample maximum is a difficult object to work with. This is true even for linear time series models, since the maximum operator is non-linear. In such cases, checking the conditions of Section 3 may be hard or even impossible task[2]." For that reason, we assume that the regularity conditions 4.3 and 4.4 hold for both the AR(1) model and the MA(1) model. Notice that we have taken into account that some conditions can be replaced by others, as discussed above.

Let σ^2_{POT} denote the asymptotic variance of the estimator for the return level based on the POT approach under serial dependence. From Theorem 3.9, we get that as $n \to \infty$,

$$\sigma^2_{\text{POT}} = \log^2\left(\frac{\theta}{c_y}\right)\sigma^2_{\text{H}} + 2\gamma^2 \log\left(\frac{\theta}{c_y}\right)\left(\int_0^1 s^{-1}c(s,1)\mathrm{d}s - c(1,1)\right) + \gamma^2 c(1,1), \tag{46}$$

where θ is the extremal index and with $c_y = -\log(1-1/y)$, σ^2_{H} as defined in Lemma 3.5 and $c(\cdot,\cdot)$ as defined in the regularity condition 3.2. From Drees (2000) (see also Stărică (1999)), we get that under serial dependence, $\sigma^2_{\text{H}} = \gamma^2 c(1,1)$. Let σ^2_{BM} denote the asymptotic variance of the estimator for the return

[2]Regularity conditions 4.1-4.4 are Conditions 3.2-3.5 of Section 3 in Bücher and Segers (2018b), as mentioned in Section 4.4.1.

level based on the BM method. From Theorem 4.4, we get that as $n \to \infty$,

$$\sigma^2_{\text{BM}} = \gamma^2 \left(\frac{6}{\pi^2} \left(\log^2(c_y) + 2\log(c_y)(g-1) + (1-g)^2 \right) + 1 \right), \tag{47}$$

with $c_y = -\log(1 - 1/y)$ and $g = 0.5772...$ the Euler-Mascheroni constant. Notice that for the BM method the asymptotic variance is unaffected by serial dependence since, even for time series, maxima over large disjoint blocks are asymptotically independent because of the regularity condition 4.2 (Bücher & Segers, 2018b).

5.2.1 Autoregressive model

Consider the stationary solution of the AR(1) equation

$$X_i = \phi X_{i-1} + U_i, \tag{48}$$

for some $\phi \in (0, 1)$ and i.i.d. random variables U_i. Let F_u denote the distribution function of the innovations. Assume that F_u possesses a positive Lebesgue density f_u. Then the time series X_i, $i \in \mathbb{N}$, is geometrically β-mixing (see Doukhan (1994), Theorem 2.4.6). In model (48), the variables X_i are heavy-tailed if and only if the innovations have heavy tails. Hence, we assume that F_u has balanced heavy tails, that is,

$$1 - F_u(x) \sim p x^{-1/\gamma} l(x) \quad \text{and} \quad F_u(-x) \sim (1-p) x^{-1/\gamma} l(x), \tag{49}$$

as $x \to \infty$, for some slowly varying function l and $p \in (0, 1)$. From Section 5.1 of De Haan et al. (2016), we get that the regularity conditions 3.1-3.5 hold with

$$c(s,t) = s \wedge t + \sum_{m=1}^{\infty} \left(s \wedge t\phi^{m/\gamma} + t \wedge s\phi^{m/\gamma} \right). \tag{50}$$

Under serial dependence, the asymptotic variance of the estimator for the return level based on the POT approach is given by (46). From (50), we deduce that $c(1,1) = 1 + 2 \cdot \phi^{1/\gamma} / \left(1 - \phi^{1/\gamma}\right)$ (see also Drees (2003), Section 3.2). The extremal index for the AR(1) model is equal to $\theta = 1 - \phi$ (see Chernick, Hsing, and McCormick (1991), p. 843). Chernick et al. (1991) use an AR(1) model with Cauchy marginals. Indeed, the Cauchy distribution has regularly varying tails and satisfies the tail balancing condition (49) with $p = 1/2$. Combining all these parts, we get that

$$\begin{aligned} \sigma^2_{\text{POT}} = {} & \gamma^2 \left(1 + 2\frac{\phi^{1/\gamma}}{1 - \phi^{1/\gamma}} \right) \left(\log^2\left(\frac{1-\phi}{c_y} \right) - 2\log\left(\frac{1-\phi}{c_y} \right) + 1 \right) \\ & + \gamma^2 2\log\left(\frac{1-\phi}{c_y} \right) \left(1 - 2\frac{\phi^{1/\gamma}}{\phi^{1/\gamma} - 1} - \log\phi \frac{\phi^{1/\gamma}}{\gamma\left(\phi^{1/\gamma} - 1\right)^2} \right), \end{aligned}$$

with $c_y = -\log(1 - 1/y)$.

For $y > 1$, we compare σ^2_{POT} with σ^2_{BM}, which is equal to (47), by calculating the following ratio

$$\frac{\sigma^2_{\text{POT}}}{\sigma^2_{\text{BM}}} = \frac{\left(1 + 2\frac{\phi^{1/\gamma}}{1-\phi^{1/\gamma}}\right)\left(\log^2\left(\frac{1-\phi}{c_y}\right) - 2\log\left(\frac{1-\phi}{c_y}\right) + 1\right)}{\frac{6}{\pi^2}\left(\log^2(c_y) + 2\log(c_y)(g-1) + (1-g)^2\right) + 1}$$
$$+ \frac{2\log\left(\frac{1-\phi}{c_y}\right)\left(1 - 2\frac{\phi^{1/\gamma}}{\phi^{1/\gamma}-1} - \log\phi\frac{\phi^{1/\gamma}}{\gamma\left(\phi^{1/\gamma}-1\right)^2}\right)}{\frac{6}{\pi^2}\left(\log^2(c_y) + 2\log(c_y)(g-1) + (1-g)^2\right) + 1}, \tag{51}$$

with $c_y = -\log(1 - 1/y)$ and $g = 0.5772...$ the Euler-Mascheroni constant. Only for $y \geq 155000$, we obtain the ratio in (51) is always higher than one and consequently σ^2_{BM} is lower than σ^2_{POT} for all ϕ and γ. Otherwise, there are values for ϕ and γ such that σ^2_{BM} is higher than σ^2_{POT}.

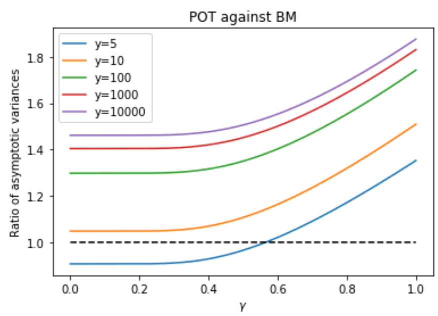
In Figure 2, we plot this ratio against the extreme value index γ for different values of the parameters ϕ and y. From Figure 2, we can draw the following conclusions:

- By comparing the distinct y values represented by the different colors, we conclude that the higher the y, the higher the ratio in (51) for all Figs. 2(a)-2(d) (except for Fig. 2(d), $y = 5$).
- For all Figs. 2(a)-2(d), the ratio in (51) has the same pattern as γ increases. The ratio starts flat for low value of γ and for γ above a certain threshold, the ratio is increasing in γ. Hence, we conclude that the more heavy-tailed, the higher the ratio in (51).
- By comparing across Figs. 2(a)-2(d), we conclude that the more dependence, the higher the ratio in (51) at $\gamma = 1$.

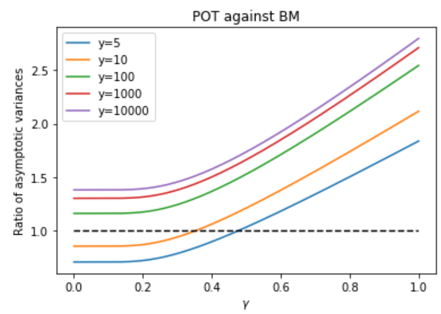
Let γ^* denote the value of γ for which σ^2_{POT} is equal to σ^2_{BM}. For $\gamma > \gamma^*$, it holds that the ratio in (51) is greater than one and consequently σ^2_{POT} is higher than σ^2_{BM}. Similarly, for $\gamma < \gamma^*$, it holds that σ^2_{POT} is lower than σ^2_{BM}. In Table 1, we report γ^* for $\phi \in \{0.1, 0.3, 0.5, 0.9\}$. From Table 1, we observe that the higher the y, the lower the γ^* (except for $\phi = 0.9$ and $y = 5$). By comparing across the rows, we note that the more dependence, the lower the γ^*.

In Figure 3, we plot the ratio in (51) against the parameter ϕ for different values of the extreme value index γ and y. Let I_ϕ denote the specific interval for ϕ such that the ratio in (51) is less than one and consequently σ^2_{BM} is higher than σ^2_{POT}. From Figure 3, we observe that only for $y = 5$ and $y = 10$ (and $y = 100$ in Fig. 3(a)) such an interval exists. In Table 4, we report I_ϕ for $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$. From the table, we observe that the higher the γ, the narrower the interval I_ϕ.

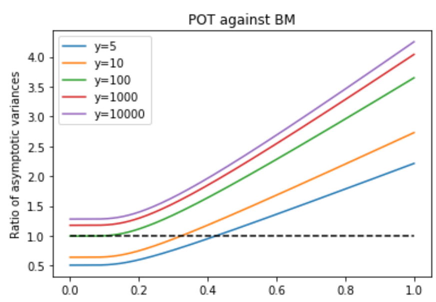
In conclusion, the BM method outperforms the POT approach in terms of having a lower asymptotic variance for both high y, high γ and high ϕ in combination with high γ. The parameter y represents the average number of blocks of size m until encountering the first such block whose maximum exceeds the return level, $R_{m,y}$. Hence, high y actually means more extreme events. Under serial dependence, the POT approach first approximates the return level by a high quantile and then offers an extrapolation method to estimate this quantile. The BM method directly approximates the return level by a quantile of the Fréchet distribution. The indirect approximation of the POT approach becomes worse in the far tail due to the inaccuracy in the approximation. Only when y is very low, the approximation is actually good enough to use it. The direct approximation of the BM method works well even in the far tail. The extreme value index γ determines the tail behaviour of a distribution and indicates the heaviness of the tail. Hence, high γ also means that observations may contain more extreme events. The parameter ϕ measures the dependence. The higher the ϕ, the stronger the dependence and the more complex the dependence structure. The indirect approximation formula for the POT approach becomes less accurate for higher ϕ. Therefore, the direct approach for the BM method is better for high ϕ in combination with high γ.



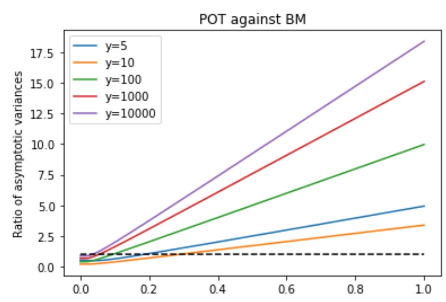
(a) $\phi = 0.1$



(b) $\phi = 0.3$

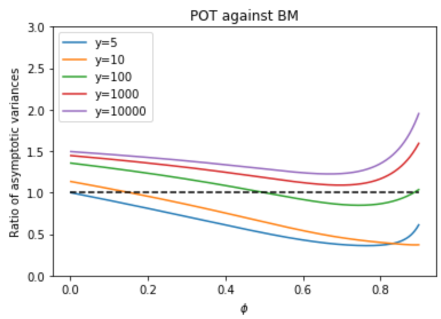


(c) $\phi = 0.5$

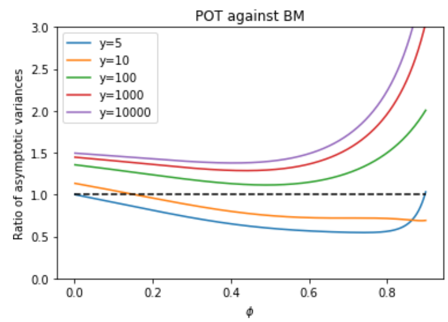


(d) $\phi = 0.9$

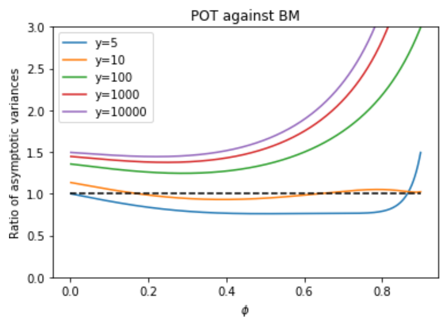
Figure 2: Ratio between the asymptotic variances for the POT approach and the BM method under AR(1). The dashed line corresponds to the ratio equal to one.



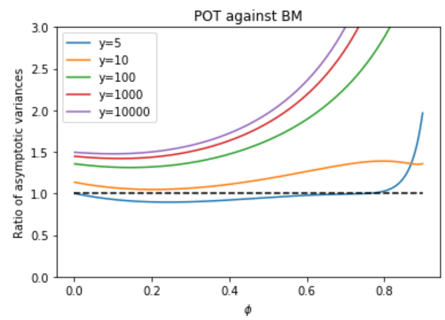
(a) $\gamma = 0.1$



(b) $\gamma = 0.2$



(c) $\gamma = 0.3$



(d) $\gamma = 0.4$

Figure 3: Ratio between the asymptotic variances for the POT approach and the BM method under AR(1). The dashed line corresponds to the ratio equal to one.

Table 1: γ^* under AR(1)

y	5	10	100	1000	10000
$\phi = 0.1$	0.57	-	-	-	-
$\phi = 0.3$	0.48	0.35	-	-	-
$\phi = 0.5$	0.43	0.32	0.098	-	-
$\phi = 0.9$	0.19	0.29	0.096	0.056	0.037

Table 2: I_ϕ under AR(1)

y	5	10
$\gamma = 0.1$	$(0, 0.900]$	$(0.153, 0.900]$
$\gamma = 0.2$	$(0, 0.897)$	$(0.154, 0.900]$
$\gamma = 0.3$	$(0, 0.865)$	$(0.174, 0.635)$
$\gamma = 0.4$	$(0, 0.759)$	-

5.2.2 Moving average model

Consider the stationary solution of the MA(1) equation

$$X_i = \phi U_{i-1} + U_i,$$

for some $\phi \in (0,1)$ and where the innovation U satisfies the same conditions as in the AR(1) model in the previous section. From Section 5.2 of De Haan et al. (2016), we get that the regularity conditions 3.1-3.5 hold with

$$c(s,t) = s \wedge t + \frac{1}{1+\phi^{1/\gamma}} \left(s \wedge t\phi^{1/\gamma} + t \wedge s\phi^{1/\gamma} \right). \tag{52}$$

Under serial dependence, the asymptotic variance of the estimator for the return level based on the POT approach is given by (46). From (52), we deduce that $c(1,1) = 1 + 2 \cdot \phi^{1/\gamma} / \left(1+\phi^{1/\gamma}\right)$. The extremal index for the MA(1) model is equal to $\theta = 1/\left(1+\phi^{1/\gamma}\right)$ (see Chernick et al. (1991), Proposition 2.1). Combining all these parts, we get that

$$\begin{aligned}\sigma^2_{\text{POT}} = \gamma^2 &\left(1 + 2\frac{\phi^{1/\gamma}}{1+\phi^{1/\gamma}}\right)\left(\log^2\left(c_y\left(1+\phi^{1/\gamma}\right)\right) + 2\log\left(c_y\left(1+\phi^{1/\gamma}\right)\right) + 1\right) \\ &- \gamma^2 2\log\left(c_y\left(1+\phi^{1/\gamma}\right)\right)\left(1 + \frac{\phi^{1/\gamma}}{1+\phi^{1/\gamma}}\left(2 - \frac{1}{\gamma}\log\phi\right)\right),\end{aligned}$$

with $c_y = -\log(1-1/y)$.

For $y > 1$, we compare σ^2_{POT} with σ^2_{BM}, which is equal to (47), by calculating the following ratio

$$\begin{aligned}\frac{\sigma^2_{\text{POT}}}{\sigma^2_{\text{BM}}} = &\frac{\left(1+2\frac{\phi^{1/\gamma}}{1+\phi^{1/\gamma}}\right)\left(\log^2\left(c_y\left(1+\phi^{1/\gamma}\right)\right) + 2\log\left(c_y\left(1+\phi^{1/\gamma}\right)\right) + 1\right)}{\frac{6}{\pi^2}\left(\log^2(c_y) + 2\log(c_y)(g-1) + (1-g)^2\right) + 1} \\ &- \frac{2\log\left(c_y\left(1+\phi^{1/\gamma}\right)\right)\left(1 + \frac{\phi^{1/\gamma}}{1+\phi^{1/\gamma}}\left(2 - \frac{1}{\gamma}\log\phi\right)\right)}{\frac{6}{\pi^2}\left(\log^2(c_y) + 2\log(c_y)(g-1) + (1-g)^2\right) + 1},\end{aligned} \tag{53}$$

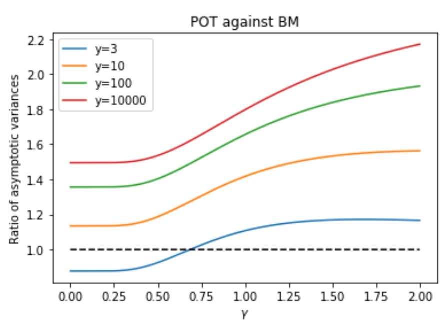
with $c_y = -\log(1-1/y)$ and $g = 0.5772...$ the Euler-Mascheroni constant. Only for $y \in (1.43, 4.98)$, we obtain there are values for ϕ and γ such that the ratio in (53) is less than one and consequently σ^2_{BM} is higher than σ^2_{POT}. Otherwise, σ^2_{BM} is lower than σ^2_{POT} for all ϕ and γ.

In Figure 4, we plot this ratio against the the extreme value index γ for different values of the parameters ϕ and y. The general feature is comparable to that observed from Figure 2. Notice that γ is between 0 and 2 instead of 0 and 1 as in Figure 2. A notable difference between Figures 2 and 4 is the convexities with respect to γ: we observe a concave (resp., convex) relation in γ under the MA(1) (resp., AR(1)) model. Hence, for high γ, the ratio between the asymptotic variances is much lower for the MA(1) model. For example, by comparing between Figure 2(d) and Figure 4(d) with $\gamma = 1$ and $y = 10000$, the ratio under the AR(1) model is equal to 18.4 and the ratio under the MA(1) model is equal to 2.5.

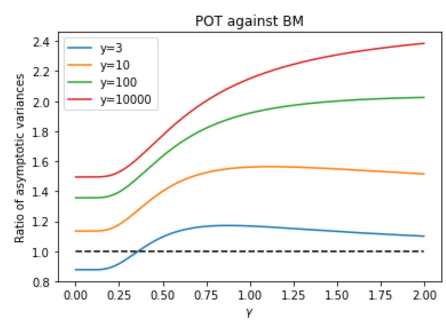
Let γ^* denote the value of γ for which σ^2_{POT} is equal to σ^2_{BM}. From Figure 4, we observe that only if $y = 3$ there is a γ^*. In Table 3, we report γ^* for different values of the parameter ϕ. By comparing across rows, we note that, similar to the AR(1) model, the more dependence, the lower the γ^*.

In Figure 5, we plot the ratio in (53) against the parameter ϕ for different values of the extreme value index γ and y. Let I_ϕ denote the specific interval for ϕ such that the ratio in (53) is less than one and consequently σ^2_{BM} is higher than σ^2_{POT}. From Figure 5, we observe that only for $y = 3$ the specific interval I_ϕ exists. In Table 4, we report I_ϕ for $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$. From the table, we observe that, similar to the AR(1) model, the more heavy tailed, the narrower the interval I_ϕ.

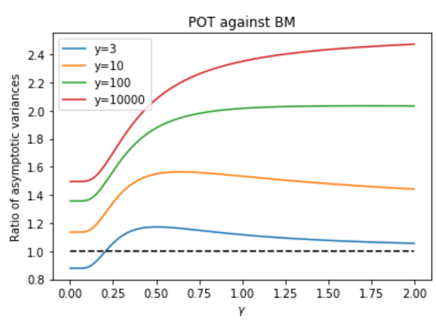
In conclusion, the BM method outperforms the POT approach in terms of having a lower asymptotic variance for $y > 4.98$. For γ and ϕ, there is no clear pattern.



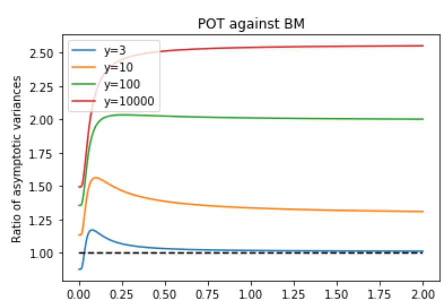
(a) $\phi = 0.1$



(b) $\phi = 0.3$

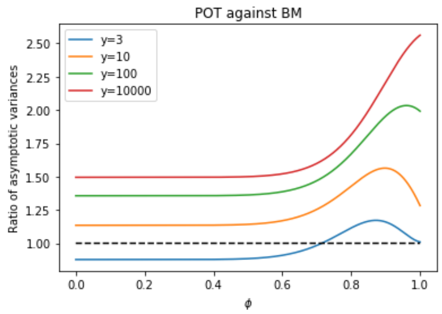


(c) $\phi = 0.5$

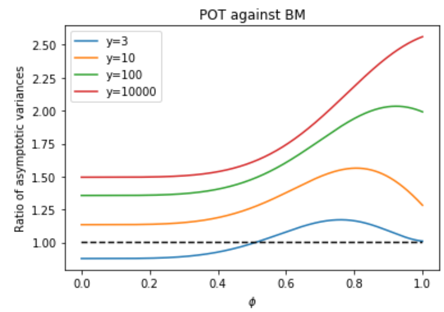


(d) $\phi = 0.9$

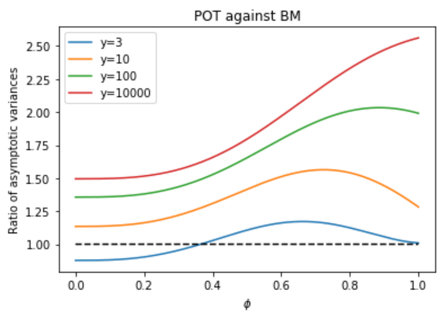
Figure 4: Ratio between the asymptotic variances for the POT approach and the BM method under MA(1). The dashed line corresponds to the ratio equal to one.



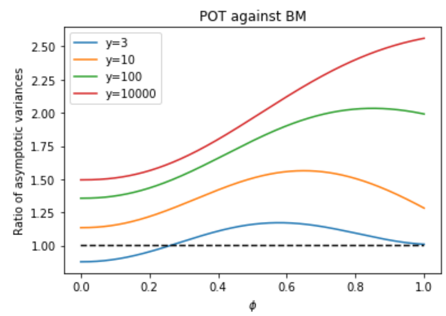
(a) $\gamma = 0.1$



(b) $\gamma = 0.2$



(c) $\gamma = 0.3$



(d) $\gamma = 0.4$

Figure 5: Ratio between the asymptotic variances for the POT approach and the BM method under MA(1). The dashed line corresponds to the ratio equal to one.

Table 3: γ^* under MA(1)

y	3
$\phi = 0.1$	0.69
$\phi = 0.3$	0.36
$\phi = 0.5$	0.21
$\phi = 0.9$	0.031

Table 4: I_ϕ under MA(1)

y	3
$\gamma = 0.1$	$(0, 0.715)$
$\gamma = 0.2$	$(0, 0.511)$
$\gamma = 0.3$	$(0, 0.365)$
$\gamma = 0.4$	$(0, 0.261)$

6 Conclusion

In this thesis, we study and compare the POT approach and the BM method for estimating the return level. We first analyze the two methods both for i.i.d. observations and for observations which exhibit serial dependence. For the i.i.d. model, the autoregressive model and the moving average model, we explicitly calculate the asymptotic variances for the two methods and make a comparison.

Under independence and under the MA(1) model, we conclude that for $y > 4.98$ the BM method theoretically outperforms the POT approach in terms of having a lower asymptotic variance for all ϕ and γ. Hence, if we consider yearly maxima, i.e. $m = 365$, and we look at the 5-year or higher return level, we should use the BM method. Under the AR(1) model, we conclude that for $y \geq 155000$ the BM method theoretically outperforms the POT approach in terms of having a lower asymptotic variance for all ϕ and γ. For $y \leq 10$ and $\gamma < 0.2$, the POT approach outperforms the BM method in terms of having a lower asymptotic variance for all $\phi \in (0.15, 0.90)$. For $y \in (10, 155000)$ and $\gamma > 0.1$, the BM method outperforms the POT approach in terms of having a lower asymptotic variance for all ϕ. Overall, we conclude that the BM method is superior for estimating the return level both under independence and under serial dependence.

This research can be improved in at least two ways. Firstly, for the POT approach under serial dependence, we need to estimate the extremal index. In the proof of the asymptotic normality of the estimator for the return level, the asymptotic distribution of the estimator for the extremal index plays a role. In the current proof, for the extremal index, we use an intermediate sequence k_1 such that as $n \to \infty$, $k/k_1 \to \infty$. With such a choise of k_1, the asymptotic distribution of the estimator for the extremal index does not have an influence on the asymptotic distribution of the estimator for the return level. In other words, we assume away the potential impact. Handling the estimator for the extremal index at the same k level is left for future research. Secondly, for the POT approach under serial dependence, we approximate the return level by a quantile. In (19), we assume that the quantile and the return level are close enough. The proof of this assumption is also left for future research.

References

Berghaus, B., & Bücher, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Annals of Statistics*, *46*(5), 2307–2335.

Bradley, R. C., et al. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, *2*, 107–144.

Bücher, A., & Segers, J. (2018a). Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, *12*(1), 1098–1125.

Bücher, A., & Segers, J. (2018b). Maximum likelihood estimation for the fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, *24*(2), 1427–1462.

Bücher, A., & Zhou, C. (2018). A horse racing between the block maxima method and the peak-over-threshold approach. ArXiv preprint, arXiv:1807.00282.

Chernick, M. R., Hsing, T., & McCormick, W. P. (1991). Calculating the extremal index for a class of stationary sequences. *Advances in applied probability*, *23*(4), 835–850.

Cramér, H. (1946). Mathematical methods of statistics. *Princeton university press*.

de Haan, L., & Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.

De Haan, L., Mercadier, C., & Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics*, *20*(2), 321–354.

Doukhan, P. (1994). *Mixing: properties and examples*. Springer-Verlag.

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

Drees, H. (1998b). A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference*, *66*(1), 95–112.

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Annals of Applied Probability*, *10*(4), 1274–1301.

Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, *9*(4), 617–657.

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annals of Société Polonaise de Mathematique*, *6*, 93-116.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, *44*, 423–453.

Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*, 1163–1174.

Jansen, D. W., & De Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. *Review of Economics and Statistics*, *73*(1), 18–24.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields*, *65*(2), 291–306.

McNeil, A. J. (1998). Calculating quantile risk measures for financial return series using extreme value theory (Tech. Rep.). ETH Zurich.

Pickands, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, *3*(1), 119–131.

Rootzén, H. (1995). *The tail empirical process for stationary sequences* (Tech. Rep.). Chalmers University of Technology.

Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications*, *119*(2), 468–490.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, *42*(1), 43.

Slutsky, E. (1925). Über stochastische asymptoten und grenzwerte. *Metron*, *5*, 3–89.

Stărică, C. (1999). On the tail empirical process of solutions of stochastic difference equations. *Preprint, Chalmers University of Gothenburg*.

von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Rev. math. Union interbalcanique*, *1*, 141–160.

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

Zayed, M., & Butt, N. S. (2017). The extended fréchet distribution: Properties and applications. *Pakistan Journal of Statistics and Operation Research*, *13*(3), 529–543.