A horse race between the block maxima method and the peak–over–threshold approach

Axel Bücher
Heinrich-Heine-Universität Düsseldorf
and

Chen Zhou
Erasmus University Rotterdam and De Nederlandsche Bank

Abstract. Classical extreme value statistics consists of two fundamental approaches: the block maxima (BM) method and the peak-over-threshold (POT) approach. It seems to be general consensus among researchers in the field that the POT approach makes use of extreme observations more efficiently than the BM method. We shed light on this discussion from three different perspectives. First, based on recent theoretical results for the BM method, we provide a theoretical comparison in i.i.d. scenarios. We argue that the data generating process may favour either one or the other approach. Second, if the underlying data possesses serial dependence, we argue that the choice of a method should be primarily guided by the ultimate statistical interest: for instance, POT is preferable for quantile estimation, while BM is preferable for return level estimation. Finally, we discuss the two approaches for multivariate observations and identify various open ends for future research.

Key words and phrases: extreme value statistics, extreme value index, extremal index, stationary time series.

1. INTRODUCTION

Extreme-Value Statistics can be regarded as the art of extrapolation outside the sample. Based on a finite sample from some distribution $F$, typical quantities of interest are quantiles whose levels are larger than the largest observation or probabilities of rare
BÜCHER AND ZHOU

2

events which have not occurred yet in the observed sample. Estimating such objects typically relies on the following fundamental domain-of-attraction condition: there exists a constant $\gamma \in \mathbb{R}$ and sequences $a_r > 0$ and $b_r$, $r \in \mathbb{N}$, such that

$$\lim_{r \to \infty} F^r(a_r x + b_r) = \exp \left\{ -(1 + \gamma x)^{-1/\gamma} \right\} \text{ for all } 1 + \gamma x > 0.$$  

In that case, $\gamma$ is called the extreme value index. The limit appears unnecessarily specific, but it is in fact the only possible non-degenerate limit of the expression on the left-hand side. An equivalent representation of the domain of attraction condition (1.1) is as follows: there exists a positive function $\sigma = \sigma(t)$ such that

$$\lim_{t \uparrow x^*} \frac{1 - F(t + \sigma(t)x)}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma} \text{ for all } 1 + \gamma x > 0,$$

where $x^*$ denotes the right end-point of the support of $F$, see Balkema and de Haan (1974). The two sequences in (1.1) are related to the function $\sigma$ as follows:

$$a_r = \sigma(b_r)$$

and

$$b_r = \frac{U(r)}{U(1 - 1/r)} = \frac{1}{U(1 - F)}.$$ 

Consider for instance the consequences of the previous two displays for high quantiles of $F$. By (1.1), for all $p$ sufficiently small, $F^r(1 - p)$ can be approximated (by first solving the equation $(1 - p) = \exp\{- (1 + \gamma x)^{-1/\gamma}\}$ for $x$ and then rescaling the solution) as follows:

$$F^r(1 - p) \approx b_r + a_r \left\{ -r \log(1 - p) \right\}^{-\gamma} - 1 \approx b_r + a_r \left( \frac{rp}{\gamma} \right)^{-\gamma} - 1.$$  

Hence, by the plug-in-principle, a suitable choice of $r$ and suitable estimators of $a_r, b_r$ and $\gamma$ immediately suggest estimators for high quantiles.

Similarly, by (1.2), for all $p$ sufficiently small,

$$F^r(1 - p) \approx t + \sigma(t) \left\{ \frac{p}{1 - F(t)} \right\}^{-\gamma} - 1.$$ 

Again, by the plug-in-principle, a suitable choice of $t$ and suitable estimators of $\sigma(t)$, $\gamma$ and $1 - F(t)$ immediately leads to estimators for high quantiles. Here, $t$ is typically chosen as a large order statistic $t = X_{n-k,n}$ and $1 - F(t)$ is replaced by $k/n$.

In practice, estimators for the parameters in these two approaches typically follow their corresponding basic principles: the block maxima (BM) method motivated by (1.1) and the peak-over-threshold (POT) approach motivated by (1.2). Let $X_1, X_2, \ldots, X_n$ be a sample of observations drawn from $F$, and for the moment assume that the observations are independent. Then (1.1) gives rise to the BM method (Gumbel, 1958): for some block size $r \in \{1, \ldots, n\}$, divide the data into $k = [n/r]$ blocks of length $r$ (and a possibly remaining block of smaller size which has to be discarded). By independence, each block has cdf $F^r$. By (1.1), for large block sizes $r$, the sample of block maxima can then be
regarded as an approximate i.i.d. sample from the three-parametric generalized extreme-value (GEV) distribution $G_{\gamma,b,a}$ with location parameter $b = b_r$, scale parameter $a = a_r$ and shape parameter $\gamma$, defined by its cdf

$$G_{\gamma,b,a}^{GEV}(x) := \exp\left\{ -\left(1 + \gamma \frac{x - b}{a}\right)^{-1/\gamma}\right\}1\left(1 + \gamma \frac{x - b}{a} > 0\right).$$

The three parameters can be estimated by maximum-likelihood or moment-matching, among others. Irrespective of the particular estimation principle, any estimator defined in terms of the sample of block maxima will be referred to as an estimator based on the BM method. An illustration of the sample of block maxima is provided in the upper panel of Figure 1.

Often, an available data-sample consists of block maxima only, for example, annual maxima of a river level. Then a practitioner may only rely on the block maxima method. If the underlying observations are available, then (1.2) gives rise to the competing POT approach (Pickands, 1975): for sufficiently large $t$ in (1.2), we obtain that, for any $x > 0$,

$$\Pr(X > t + x \mid X > t) = \frac{\Pr(X > t + x)}{\Pr(X > t)} \approx \left(1 + \gamma \frac{x}{\sigma}\right)^{-1/\gamma} = 1 - G_{\gamma,\sigma}^{GP}(x),$$

where the right-hand side defines the two-parametric generalized Pareto (GP) distribution with scale parameter $\sigma := \sigma(t)$ and shape parameter $\gamma$. In practice, $t$ is typically chosen as the $(n-k)$-th order statistic $X_{n-k:n}$ for some intermediate value $k$ (hence, $X_{n-k:n}$ is the $(1-1/r)$-sample quantile with $r = n/k$). Then, one may regard the sample $X_{n-k+1:n} - X_{n-k:n}, \ldots, X_{n:n} - X_{n-k:n}$ as observations from the two-parametric GP distribution. The parameters can hence be estimated by moment matching, and even by maximum-likelihood since the sample of order statistics can actually be regarded as independent (see, e.g., Lemma 3.4.1 in de Haan and Ferreira, 2006). In general, any estimator defined in terms of all observations exceeding some (random) threshold will be referred to as an estimator based on the POT approach. An illustration of the POT-sample is provided in the lower panel of Figure 1. The vanilla estimator within this class is the Hill estimator (Hill, 1975) in the case $\gamma > 0$.

The goal of the present paper is an in-depth comparison of the two approaches, in particular in terms of recent solid theoretical advances on asymptotic theory for the BM method, but also with a view on time series data and multivariate observations. The discussion will mostly be of reviewing nature, but some new insights will be presented as well. The next paragraphs summarize our contribution in a chronological order.

1. **Efficiency comparison in i.i.d. scenarios.** It seems to be general consensus among researchers in extreme value statistics that the POT approach produces more efficient estimators than the BM method. The main heuristic reason is that all large observations are used for the calculation of POT estimators, while BM estimators may
miss some large observations falling into the same block. This heuristics was confirmed by simulation studies in Caires (2009), see also the additional references mentioned in Ferreira and de Haan (2015). Due to some recent advances (see also Section 9.3 in Reiss, 1989, for some early results in the case $\gamma > 0$) on theoretic properties of BM estimators (Dombry, 2015; Ferreira and de Haan, 2015; Bücher and Segers, 2014, 2018b; Dombry and Ferreira, 2019), the two approaches may actually be compared on solid theoretical grounds. For a certain type of cdfs, such a discussion has been carried out in Ferreira and de Haan (2015) and Dombry and Ferreira (2019); their findings are summarized and extended in Section 2 of this paper. We show that, depending on the data generating process, the convergence rate of the two methods may be different, with no general winner being identifiable. In case the rates are the same, BM estimators typically have a smaller variance, but a larger bias than their POT-competitors.

2. POT and BM applied to time series. The above discussion motivating the POT approach and the BM method was based on an i.i.d. assumption on the underlying sample. This assumption is actually quite restrictive since it excludes many common environmental or financial applications, where the underlying sample is typically a (stationary) time series. In this setting, it seems to be general consensus that the block maxima method still ‘works’ because the block maxima are (1) still approximately GEV-distributed (Leadbetter, 1974) and (2) distant from each other and thus bear low
serial dependence. Consequently, the sample of block maxima may still be regarded as an approximate i.i.d. sample from the three-parametric GEV-distribution. This heuristics is confirmed by recent theoretical results in Bücher and Segers (2018b, 2014). Nevertheless, as discussed in Section 3 below, an obstacle occurs: the location and scale parameters attached to block maxima of a time series will typically be different from those of an i.i.d. series from the same stationary distribution $F$, whence estimators for quantities that depend on the stationary distribution only will possibly be inconsistent. The missing link is provided by the extremal index (Leadbetter, 1983), a parameter in $[0, 1]$ capturing the tendency of the extreme observations of a stationary time series to occur in clusters. The discussion will be worked out on the example of high quantile estimation: based on suitable estimators for the extremal index, see Section 3 below, (1.3) can in fact be modified to obtain consistent BM estimators of large quantiles. On the other hand, estimators based on the POT approach for characteristics of the stationary distribution remain consistent. This however comes at the cost of an increased variance of the estimators due to potential clustering of extremes, see Hsing, 1991; Drees, 2000; Rootzén, 2009, among many others. Should the ultimate interest be in return level or return periods estimation, the picture is reversed: the BM method is consistent without the need of estimating the extremal index, while POT estimators typically require estimates of the extremal index. More details are provided in Section 3.

3. Extensions to multivariate observations and stochastic processes. The previous discussion focussed on the univariate case. Section 4 briefly discusses multivariate extensions. On the theoretical side, while there are many results available for the POT approach, there is clearly a supply issue regarding the BM method: almost all statistical theory is formulated under the assumption that the block maxima genuinely follow a multivariate extreme value distribution, thereby ignoring a potential bias and rendering a fair theoretical comparison impossible for the moment (to the best of our knowledge, the only available results on the BM method are provided in Bücher and Segers, 2014). Instead, we provide a review on some of the existing theoretical results using these two approaches, and identify the open ends that may eventually lead to results allowing for an in-depth theoretical comparison in the future. Not surprisingly, a fair comparison is even more difficult when considering extreme value analysis for stochastic processes. Most of the existing statistical methods are based on max-stable process models, i.e., on limit models arising for maxima taken over i.i.d. stochastic processes. The respective statistical theory is again mostly formulated under the assumption that the observations are genuine observations from the max-stable model, whence the statistical methods can (in most cases) be generically attributed to the BM method. As for multivariate models, potential bias issues are mostly ignored. By contrast to multivariate models, however, very little is known for the POT approach to processes. A comparison is hence not feasible for the moment, and we limit ourselves to a brief review of existing results in Section 5.

Finally, we end the paper by a section summarizing possible open research questions, Section 6, and by a short conclusion, Section 7.
2. EFFICIENCY COMPARISON FOR UNIVARIATE I.I.D. OBSERVATIONS

The efficiency of POT and BM estimators can be compared in terms of their asymptotic bias and variance. In this section, we carry out such a comparison for estimators based on fitting the GPD distribution to the sample of exceedances (for simplicity referred to as POT subsequently) and for estimators based on fitting the GEV distribution to the sample of block maxima. It is important to note that our comparison does not include other POT-type estimators as, e.g., the Hill estimator based on fitting the Pareto distribution to the sample of excess ratios. Furthermore, for the following reasons, we mostly focus on the estimation of the extreme value index $\gamma$ only. First, among the relevant statistical problems in extremes, carrying out a solid theoretical comparison seems to be easiest for the plain problem of estimating $\gamma$, thanks to the recent theoretical advances on the BM method mentioned in the introduction. Second, an accurate estimation of $\gamma$ is often crucial for the estimation of other tail related characteristics such as high quantiles, in particular for $\gamma > 0$. The practically relevant problem of estimating high quantiles will be studied through a Monte Carlo simulation experiment in Section 2.3.

In both the POT approach and the BM method, a key tuning parameter is the intermediate sequence $k = k(n)$, which corresponds to either the number of blocks in the BM method, or the number of upper order statistics in the POT approach. For most data generating processes, consistency of respective estimators can be guaranteed if $k$ is chosen in such a way that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. Here, the small fraction $k/n$ reflects the fact that the inference is based on observations in the tail only. Typically, the variance of respective estimators is of order $1/k$, while the bias depends on how well the distribution of block maxima or threshold exceedances is approximated by the GEV or GP distribution, respectively. Choosing $k$ in an optimal way, i.e., such that the asymptotic mean squared error is minimal (see Section 2.1 below for details), one may derive what is commonly referred to as the optimal rate of convergence of a given estimator. Depending on the model, the optimal choice of $k$ may result in a faster rate for the BM method or the POT approach, as will be discussed next.

It is instructive to consider two extreme examples first (where the condition $k/n \to 0$ as $n \to \infty$ may in fact be discarded): if $F$ is the standard Fréchet-distribution, then block maxima of size $r = 1$ are already GEV-distributed. In other words a sample of $k = n$ block maxima of size $r = 1$ can be used for estimation via the BM method. The rate of convergence is thus $1/\sqrt{n}$ and the POT approach fails to achieve this rate. On the other hand, if $F$ is the standard Pareto distribution, then all $k = n$ largest order statistics can be used for the estimation via the POT approach. The rate of convergence is $1/\sqrt{n}$ for the POT approach, which is not achievable via the BM method.

Apart from these two (or similar) extreme cases, the optimal choice of $k$ depends on second order conditions quantifying the speed of convergence in the domain of attraction condition. These are often (though not always) formulated in terms of the two quantile

\footnote{For instance, the weak convergence result of Theorem 4.3.1 in de Haan and Ferreira (2006) shows that, if $\gamma > 0$, the asymptotic distribution of an appropriately standardized estimation error of a certain high quantile estimator (motivated by (1.4)) is exactly the same as that of the standardized estimation error of the estimator $\hat{\gamma}$ used for estimating the high quantile.}
functions

\[ U(x) = \left( \frac{1}{1 - F} \right)^{\downarrow} (x) \quad \text{and} \quad V(x) = \left( -\log F \right)^{\downarrow} (x) \]

for the POT approach and the BM method, respectively. The definition of the functions \( U \) and \( V \) reflects the fact that the two equivalent expressions of the domain of attraction condition (1.2) and (1.1) are limit relations for \( 1 - F \) and \( -\log F \), respectively. Further note that the domain of attraction condition (1.1) is equivalent to the fact that there exists a positive function \( a_{\text{POT}} \) such that, for all \( x > 0 \),

\[
\lim_{t \to \infty} \frac{U(tx) - U(t)}{a_{\text{POT}}(t)} = \int_1^x s^{\gamma - 1} \, ds, \tag{2.1}
\]

see Theorems 1.1.6 and 1.2.1 in de Haan and Ferreira (2006). The function \( a_{\text{POT}} \) is related to the sequence \((a_r)\) appearing in (1.1) via \( a_{\text{POT}}(r) = a_{[r]} \).

In parallel, (1.1) is also equivalent to the fact there exists a positive function \( a_{\text{BM}} \) such that, for all \( x > 0 \),

\[
\lim_{t \to \infty} \frac{V(tx) - V(t)}{a_{\text{BM}}(t)} = \int_1^x s^{\gamma - 1} \, ds. \tag{2.2}
\]

The bias of certain POT or BM estimators is determined by the speed of convergence in the latter two limit relations, which can be captured by suitable second order conditions.

For \( \gamma \in \mathbb{R}, \rho \leq 0 \) and \( x > 0 \), let

\[ h_\gamma(x) = \int_1^x s^{\gamma - 1} \, ds, \quad H_{\gamma, \rho}(x) = \int_1^x s^{\gamma - 1} \int_1^s u^{\rho - 1} \, du \, ds. \]

**Definition 2.1 (Second order conditions).** Let \( F \) be a cdf satisfying the domain-of-attraction condition (1.1) for some \( \gamma \in \mathbb{R} \). Consider the following two assumptions.

\( \text{(SO)}_U \) Suppose that there exists \( \rho_{\text{POT}} \leq 0 \), a positive function \( a_{\text{POT}} \) and a positive or negative function \( A_{\text{POT}} \) with \( \lim_{t \to \infty} A_{\text{POT}}(t) = 0 \), such that, for all \( x > 0 \),

\[
\lim_{t \to \infty} \frac{1}{A_{\text{POT}}(t)} \left( \frac{U(tx) - U(t)}{a_{\text{POT}}(t)} - h_\gamma(x) \right) = H_{\gamma, \rho_{\text{POT}}}(x). \]

\( \text{(SO)}_V \) Suppose that there exists \( \rho_{\text{BM}} \leq 0 \), a positive function \( a_{\text{BM}} \) and a positive or negative function \( A_{\text{BM}} \) with \( \lim_{t \to \infty} A_{\text{BM}}(t) = 0 \), such that, for all \( x > 0 \),

\[
\lim_{t \to \infty} \frac{1}{A_{\text{BM}}(t)} \left( \frac{V(tx) - V(t)}{a_{\text{BM}}(t)} - h_\gamma(x) \right) = H_{\gamma, \rho_{\text{BM}}}(x). \]

The functions \(|A_{\text{BM}}|\) and \(|A_{\text{POT}}|\) are then necessarily regularly varying with index \( \rho_{\text{BM}} \) and \( \rho_{\text{POT}} \), respectively. The limit function \( H_{\gamma, \rho} \) might appear unnecessarily specific, but in fact it is not, see de Haan and Stadtmüller (1996) or Section B.3 in de Haan and Ferreira (2006). If the speed of convergence in (2.1) or (2.2) is faster than any power function, we set the respective second order parameter as minus infinity. For example, for \( F = G_{\gamma, \sigma}^{\text{GP}} \) from the GP family, we have \( \{U(tx) - U(t)\}/(\sigma t^\gamma) = h_\gamma(x) \), i.e.
\[ \rho_{\text{POT}} = -\infty \] in this case. Likewise, any \( F = G_{\gamma,\sigma,\mu}^{\text{GEV}} \) from the GEV distribution satisfies \( \{ V(tx) - V(t) \}/(\sigma t^\gamma) = h_\gamma(x) \), which prompts us to define \( \rho_{\text{BM}} = -\infty \).

It is important to note that \( \rho_{\text{BM}} \) and \( \rho_{\text{POT}} \) can be vastly different. A general result can be found in Drees et al. (2003), Corollary A.1: under an additional condition\(^3\), which only concerns the cases \( \gamma = 1 \), \( \rho_{\text{BM}} = -1 \) or \( \rho_{\text{POT}} = -1 \), the two coefficients are equal within the range \([-1, 0]\). Otherwise, if \((\text{SO})_V\) holds with \( \rho_{\text{BM}} < -1 \), then \((\text{SO})_U\) holds with \( \rho_{\text{POT}} = -1 \); if \((\text{SO})_U\) holds with \( \rho_{\text{POT}} < -1 \), then \((\text{SO})_V\) holds with \( \rho_{\text{BM}} = -1 \). Some values of the parameters for various types of distributions are collected in Table 1.\(^4\) Notice that we have \( \rho_{\text{POT}} < \rho_{\text{BM}} \) for the five models in the first category (if we consider \( \tau/\alpha < 1 \) in the HW-model, \( \lambda < 1 \) in the Burr distribution and \( 1 \neq \nu < 2 \) in the \( \text{POT} \) distribution). For the five models in the second category, we have \( \rho_{\text{POT}} = \rho_{\text{BM}} \), while for the last three models, \( \rho_{\text{POT}} > \rho_{\text{BM}} \) if we consider \( \beta < 1 \) in the model \( F(x) = \exp(-1 + x^\beta) \) and \( \tau/\alpha < 1 \) in the model \( F(x) = \exp(-1 + x^{-\tau})/2 \). Finally, note that a model on the positive real line with parameters \( (\gamma, \rho_{\text{POT}}, \rho_{\text{BM}}) \) with \( \gamma > 0 \) can be transferred into a model with parameters \( (-\gamma, \rho_{\text{POT}}, \rho_{\text{BM}}) \) by considering the transformation \( g(x) = -1/x \).

Let us now consider asymptotic theory for the estimation of the extreme value index \( \gamma \). Perhaps surprisingly, asymptotic theory for the BM method has hitherto mostly ignored the fact that block maxima are only approximately GEV distributed (see, e.g., Prescott and Walden, 1980; Hosking et al., 1985; Bücker and Segers, 2017, among others). With the notable exception of Reiss (1989), Section 9.3, case \( \gamma > 0 \) (which went unnoticed in most of the recent work described below), only recent theoretical studies in Ferreira and de Haan (2015) and Dombry and Ferreira (2019) for the probability weighted moment (PWM) and the maximum likelihood (ML) estimator, respectively, take the approxi-

\(^3\)The precise condition is as follows: if \((\text{SO})_U\) is met and if \( \lim_{t \to -\infty} 2t A_{\text{POT}}(t) = c \in [ -\infty, \infty ] \setminus \{ 1 - \gamma \} \) then \((\text{SO})_V\) is met with \( \rho_{\text{BM}} = \max(\rho_{\text{POT}}, -1) \); and vice versa for \((\text{SO})_V\Rightarrow(\text{SO})_U\).

\(^4\)We remark that for the the \( t_1 \)-distribution, we obtain \( \rho_{\text{BM}} = \rho_{\text{POT}} = -2 \). This is a special example for which Corollary 4.1 in Drees et al. (2003) is not applicable: \( 2t A(t) \) converges to \( 0 = 1 - \gamma \).

\begin{table}
\begin{center}
\begin{tabular}{|c|c|c|c|}
\hline
Distribution & \( \gamma \) & \( \rho_{\text{POT}} \) & \( \rho_{\text{BM}} \) \\
\hline
\( \text{GP}(\gamma, \sigma) \) & \( \gamma \) & \( -\infty \) & \( -1 \) \\
Arcsin & \( -2 \) & \( -2 \) & \( -1 \) \\
\( \text{HW}(\alpha, \tau) \) & \( 1/\alpha \) & \( -\tau/\alpha \) & \( \max(-\tau/\alpha, -1) \) \\
\( \text{Burr}(\eta, \tau, \lambda) \) & \( 1/(\lambda \tau) \) & \( -1/\lambda \) & \( \max(-1/\lambda, -1) \) \\
\( t_{\nu}, \nu \neq 1 \) & \( 1/\nu \) & \( -2/\nu \) & \( \max(-2/\nu, -1) \) \\
\hline
Cauchy\((= t_1)\) & 1 & -2 & -2 \\
\( \text{LogGamma}(\alpha, \beta), \alpha \neq 1 \) & \( \beta \) & 0 & 0 \\
\( \text{Weibull}(\lambda, \beta), \beta \neq 1 \) & 0 & 0 & 0 \\
\( \text{Gamma}(\alpha, \beta), \alpha \neq 1 \) & 0 & 0 & 0 \\
\( \text{Normal}(\mu, \sigma^2) \) & 0 & 0 & 0 \\
\( F(x) = \exp(- (1 + x^\beta)^{-\nu}) \) & \( 1/(\alpha \beta) \) & \( \max(-1/\beta, -1) \) & \( -1/\beta \) \\
\( F(x) = \exp(- x^{-\alpha} (1 + x^{-\tau})/2) \) & \( 1/\alpha \) & \( \max(-\tau/\alpha, -1) \) & \( -\tau/\alpha \) \\
\( \text{GEV}(\gamma, \mu, \sigma) \) & \( \gamma \) & \( -1 \) & \( -\infty \) \\
\hline
\end{tabular}
\end{center}
\caption{Extreme value index and second order parameters for various models.}
\end{table}
mation into account. Correspondingly, the asymptotic bias can be explicitly analyzed, relying on the second order condition (SO)\(_V\) above. On the other hand, solid theoretical studies regarding the POT approach have a much longer history, see de Haan and Ferreira (2006) for a comprehensive overview. For the sake of theoretical comparability with the BM method, we will subsequently exemplarily deal with the ML estimator and the PWM estimator, for which Theorems 3.4.2 and 3.6.1 in de Haan and Ferreira (2006) provide the respective asymptotic theory under the assumption that (SO)\(_U\) is met (the results rely on Drees, 1998; Drees et al., 2004).

Summarizing the above mentioned results, for both methods (POT and BM), the ML-estimators are consistent for \(\gamma > -1\) and asymptotically normal for \(\gamma > \frac{-1}{2}\), while PWM-estimators are consistent for \(\gamma < 1\) and asymptotically normal for \(\gamma < \frac{1}{2}\).

Asymptotic theory is formulated under the conditions that \(k = k_n\) satisfies \(k \to \infty\) and \(k/n \to 0\) (POT approach) or \(r = r_n\) satisfies \(r \to \infty\) and \(r/n \to 0\) (BM method), as \(n \to \infty\). Further define \(k_n = n/r_n\) for the BM method. Then under the respective second order conditions (SO)\(_U\) and (SO)\(_V\) formulated above, the asymptotic results can be summarized as

\[
\hat{\gamma} \overset{d}{\approx} N\left(\gamma + A_m(n/k)b, \frac{1}{k}\sigma^2\right), \quad m \in \{\text{BM, POT}\},
\]

where \(\hat{\gamma}\) is one of the four estimators, and where the asymptotic bias \(b\) and the asymptotic variance \(\sigma^2\) depend on the specific estimator, the second order index \(\rho_m\) and \(\gamma\). In particular, the rate of convergence of the bias \(A_m(n/k)\) crucially depends on the second order index \(\rho_m\).

In the next two subsections, we first discuss the rate of convergence for estimators based on the POT approach and the BM method and then the asymptotic mean squared error in case the rates are the same. Then we perform an Monte Carlo experiments to illustrate the theoretical comparison in a finite-sample situation. Finally, in the last subsection, we discuss the choice of \(k\), i.e., the number of large order statistics in the POT approach or the number of blocks in the BM method.

2.1 Rate of convergence

As is commonly done, we consider the rate of convergence of the root mean squared error at a fixed model \(F\). It is instructive to first elaborate on the case \(A_m(t) \asymp t^{\rho_m}\) with \(\rho_m \in (-\infty, 0)\). In this case, the fastest rate of convergence, which is obtained by solving the minimization problem

\[
\min_{k=1, \ldots, n} \text{AMSE}(\hat{\gamma}) = \min_{k=1, \ldots, n} \frac{\sigma^2}{k} + A_m^2(n/k)b^2,
\]

is achieved when squared bias and variance are of the same order, that is, when

\[
A_m^2\left(\frac{n}{k}\right) \asymp \left(\frac{n}{k}\right)^{2\rho_m} \asymp \frac{1}{k}.
\]

Solving for \(k\) yields \(k \asymp n^{-2\rho_m/(1-2\rho_m)}\), which implies

\[
\text{Rate of Convergence of } \hat{\gamma} = n^{\rho_m/(1-2\rho_m)}.
\]
irrespective of \( m \in \{ \text{BM, POT} \} \). For the POT approach, this result is known to hold for many other estimators of \( \gamma \) that are shift invariant (including some that are not based on threshold exceedances); see de Haan and Ferreira (2006), though not for every estimator, see Table 3.1 in that reference.

Since \( \rho_{\text{BM}} \) and \( \rho_{\text{POT}} \) might not be the same, the rate of convergence may be different for the BM method and the POT approach. Table 2 provides a summary of which method results in a better rate. The case where the rates are the same is discussed in more detail in Section 2.2 below.

<table>
<thead>
<tr>
<th>2nd Order Parameters</th>
<th>Rate POT</th>
<th>Rate BM</th>
<th>Better rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho = \rho_{\text{BM}} = \rho_{\text{POT}} \in [-1, 0) )</td>
<td>( n^{\rho/(1-2\rho)} )</td>
<td>( n^{\rho/(1-2\rho)} )</td>
<td>-</td>
</tr>
<tr>
<td>( \rho_{\text{BM}} = -1, \rho_{\text{POT}} &lt; -1 )</td>
<td>( n^{{\rho_{\text{POT}}}/(1-2\rho_{\text{POT}})} )</td>
<td>( n^{-1/3} )</td>
<td>POT</td>
</tr>
<tr>
<td>( \rho_{\text{POT}} = -1, \rho_{\text{BM}} &lt; -1 )</td>
<td>( n^{-1/3} )</td>
<td>( n^{\rho_{\text{BM}}/(1-2\rho_{\text{BM}})} )</td>
<td>BM</td>
</tr>
</tbody>
</table>

Table 2

Best attainable rates of convergence for the BM and POT approaches based on ML and PWM at a fixed distribution \( F \) satisfying \( A_m(t) \propto t^m \) with \( \rho_m < 0 \), for typical relationships between \( \rho_{\text{BM}} \) and \( \rho_{\text{POT}} \).

Let us finally mention that the specific assumption on the function \( A_m \) made above (i.e., \( A_m(t) \propto t^m \) with \( \rho_m \in (-\infty, 0) \)) is not essential, see the argumentation on pages 79–80 in de Haan and Ferreira (2006). Note however that the minimization problem in (2.3) is not necessarily solved by balancing variance and squared bias anymore. Moreover, for \( \rho_m = -\infty \), the convergence rate is ‘faster than \( n^{-1/2+\varepsilon} \) for any \( \varepsilon > 0 \’\), and, depending on the underlying distribution, in fact could even achieve \( n^{-1/2} \) (see also Remark 3.2.6 in de Haan and Ferreira, 2006).

In the case \( \rho_{\text{BM}} = \rho_{\text{POT}} = 0 \), as, e.g., for the normal distribution, the rate of convergence of \( \hat{\gamma} \) has to be slower than any power function of \( n \). This is the case where extreme value statistics suffers major problems of accuracy. Nevertheless, the above arguments show that the problem is essentially the same for the POT approach and the BM method.

As mentioned above, many existing estimators achieve the same rate of convergence, at any fixed distribution \( F \). In fact, under specific assumptions, this rate of convergence is also the best attainable rate of convergence for estimating \( \gamma \) by any method of choice (i.e., the minimax rate), uniformly over some suitable neighborhood of the limiting distribution, see Hall and Welsh (1984) for the case \( \gamma > 0 \). To see this, we reformulate the results in the latter paper in such a way that it fits into the context of the second order conditions. Hall and Welsh (1984) consider the class of distributions \( D_{\text{Pareto}} = D_{\text{Pareto}}(\gamma_0, C_0, \varepsilon, \rho, A) \) on the positive real line (with fixed \( \gamma_0, C_0, \varepsilon, A > 0 \) and \( \rho < 0 \)) which have a Lebesgue density of the form

\[
f(x) = C x^{-1/\gamma-1} \{1 + r(x)\} \quad \text{with} \quad |r(x)| \leq Ax^{\rho/\gamma} \quad \forall x > 0,
\]

for some nonnegative constants \( \gamma \) and \( C \) with \( |\gamma^{-1} - \gamma_0^{-1}| < \varepsilon \) and \( |C - C_0| < \varepsilon \). Note that for \( C = 1 \) and \( r \equiv 0 \) we retrieve the Pareto(1/\( \gamma_0 \))-distribution, so that \( D_{\text{Pareto}} \) actually defines a neighborhood of the Pareto(1/\( \gamma_0 \)) distribution.
Next, suppose $\hat{\gamma}_n$ is an estimator of $\gamma$ based on $n$ i.i.d. observations such that

$$\liminf_{n \to \infty} \inf_{f \in \mathcal{D}_{\text{Pareto}}} \Pr_f(|1/\hat{\gamma}_n - 1/\gamma| \leq a_n) = 1,$$

for all $A > 0$. Then, by Theorem 2 in Hall and Welsh (1984),

$$\liminf_{n \to \infty} n^{-\rho/(1-2\rho)} a_n = \infty,$$

whence $n^{\rho/(1-2\rho)}$ may be interpreted as the minimax rate of convergence (for the neighborhoods $\mathcal{D}_{\text{Pareto}}$ with $A > 0$). Notice that this is a rather broad statement: although the class $\mathcal{D}_{\text{Pareto}}$ is a neighborhood of the Pareto distribution, the estimator $\hat{\gamma}_n$ can be any estimator, being derived from the POT or the BM methods or even other methods. In their Theorem 3, Hall and Welsh (1984) show that the Hill estimator achieves the minimax rate in a certain uniform sense. A similar statement is in fact true for the maximum likelihood estimator based on the BM method provided $\rho \in [-1, 0)$ (see Theorem 9.3.1 in Reiss, 1989), thereby providing an example that the minimax rate is also “achievable” by an estimator from the BM method. However, for $\rho < -1$, Theorem 9.3.1 in Reiss (1989) shows that the maximum likelihood estimator based on the BM method only achieves the rate of convergence $n^{-1/3}$ for some fixed $f \in \mathcal{D}_{\text{Pareto}}(\gamma_0, C_0, \varepsilon, \rho, A)$, which is slower than the minimax rate.

Recall again the pointwise (in $F$) investigation from the beginning of this section: for the POT approach, the rate of convergence for many estimators of $\gamma$ is $n^{\rho_{\text{POT}}/(1-2\rho_{\text{POT}})}$, with $\rho_{\text{POT}}$ defined in the second order condition (SO)$_U$. The following lemma shows that $\rho \geq \rho_{\text{POT}}$. The proof is given in the supplementary material.

**Lemma 2.2.** Suppose a distribution function $F$ has a density $f \in \mathcal{D}_{\text{Pareto}}(\gamma_0, C_0, \varepsilon, \rho, A)$ for some $\gamma_0, C_0, \varepsilon, A > 0$ and $\rho < 0$. Suppose the corresponding quantile function $U$ satisfies the second order condition (SO)$_U$ with a second order index $\rho_{\text{POT}}$. Then $\rho_{\text{POT}} \leq \rho$.

Hence, we conclude that for any fixed $F$ from the intersection of $\mathcal{D}_{\text{Pareto}}(\gamma_0, C_0, \varepsilon, \rho, A)$ with all distributions satisfying the second order condition (SO)$_U$, the convergence rate of most existing estimators for $\gamma$ based on the POT approach (and in case $\rho_{\text{BM}} \geq -1$ also for the BM method) is at least the minimax rate provided $k$ is chosen optimally, which often is a daunting task. In the interior of the above set, the rate may even be faster.

We conjecture that a result similar to Theorem 2 in Hall and Welsh (1984) holds true for distributions in the neighborhood of a Fréchet distribution. Consider a parallel class $\mathcal{D}_{\text{Fréchet}} = \mathcal{D}_{\text{Fréchet}}(\gamma_0, C_0, \varepsilon, \rho, A)$ defined as follows: any distribution in $\mathcal{D}_{\text{Fréchet}}$ has a density

$$f(x) = \frac{C}{\gamma} x^{-1/\gamma - 1} \exp(-Cx^{-1/\gamma}) \{1 + r(x)\} \quad \text{with} \quad |r(x)| \leq Ax^{\rho/\gamma} \quad \forall x > 0,$$

for some nonnegative constants $\gamma$ and $C$ with $|\gamma^{-1} - \gamma_0^{-1}| \leq \varepsilon$ and $|C - C_0| \leq \varepsilon$. Note that for $C = 1$ and $r \equiv 0$ we retrieve the Fréchet$(1/\gamma_0)$-distribution, so that $\mathcal{D}_{\text{Fréchet}}$
actually defines a neighborhood of the Fréchet\((1/\gamma_0)\) distribution. We conjecture that a similar statement as in (2.4) and (2.5) is met, with \( \mathcal{D}_{\text{Pareto}} \) replaced by \( \mathcal{D}_{\text{Fréchet}} \), i.e., that the minimax rate of convergence is \( n^{\rho/(1-2\rho)} \). The proof of this conjecture is still an open problem for research.

In addition, we can show a result similar to Lemma 2.2: if a distribution function is in the class \( \mathcal{D}_{\text{Fréchet}} \) and satisfies the second order condition \( (SO)_V \), then \( \rho \geq \rho_{BM} \). The proof is again given in the supplementary material. Therefore, if the aforementioned conjecture regarding the minimax rate of convergence holds, the ML and PWM estimators for \( \gamma \) based on the BM method achieve the minimax rate of convergence for any fixed \( F \) in a neighborhood of the Fréchet distribution. We may further conjecture that this is not the case for the respective POT versions if \( \rho_{POT} < -1 \).

2.2 Asymptotic mean squared error

As discussed in the previous subsection, if \( \rho_{POT} \neq \rho_{BM} \), the approach corresponding to a lower \( \rho \) generically yields estimators for \( \gamma \) with a faster attainable rate of convergence than the other approach. In this subsection, we consider the case \( \rho_{POT} = \rho_{BM} \). Then both approaches, at their best attainable rate of convergence, will yield estimators of \( \gamma \) with the same speed of convergence. Hence, the efficiency comparison should be made at the level of asymptotic mean squared error (AMSE)\(^5\) or, more precisely, its two sub-components: asymptotic bias and asymptotic variance. Notice that the asymptotic bias and variance depends on the specific estimator used, whence the comparison can only be performed based on some preselected estimators.

A detailed analysis of the PWM and the ML estimators under the BM method and the POT approach has been carried out in Ferreira and de Haan (2015) and Dombry and Ferreira (2019), for the case \( \rho_{BM} = \rho_{POT} \in [-1, 0] \) and \( \gamma \in (-0.5, 0.5) \). The results are as follows: when using the same value for \( k \), being either the number of large order statistics in the POT approach or the number of blocks in the BM method, the BM version of either ML or PWM leads to a lower asymptotic variance compared to the corresponding POT version, for all \( \gamma \in (-0.5, 0.5) \). On the other hand, the (absolute) asymptotic bias is smaller for the POT versions of the two estimators, for all \( (\gamma, \rho) \in (-0.5, 0.5) \times [-1, 0] \).

When comparing the optimal AMSE (where optimal refers to the fact that the parameter \( k \) is chosen in such a way that the AMSE for the specific estimator is minimized), it turns out that, for the ML estimator, the POT approach yields a smaller optimal AMSE. For the PWM estimator, the BM method is preferable for most combinations of \( (\gamma, \rho) \). When comparing all four estimators, the combination ML-POT has the overall smallest optimal AMSE.

2.3 Illustrative Monte Carlo experiments

An illustrative Monte Carlo simulation study was performed to analyze the theoretical asymptotic results in a finite-sample situation. The eighteen considered models are summarized in Table 3.

\(^5\)The AMSE is defined as the squared asymptotic bias plus the asymptotic variance, instead of the limit of the mean squared error. It is often used to assess the performance of an estimator, see, e.g. Danielsson et al. (2001) for the Hill estimator.
Two estimators were investigated: the maximum likelihood estimator (MLE) for fitting a GEV-distribution to the sample of block maxima of size \( r \) (implemented in Stephenson, 2018, function \( \text{fgew} \)), and the MLE for fitting a GP-distribution to the sample of the \( k \) largest order statistics (implemented in Stephenson, 2018, function \( \text{fpot} \)). Based on the discussion in the previous sections, for the six models with \( \rho_{\text{POT}} < \rho_{\text{BM}} \) one may expect a better performance of the POT-MLE, while for the six models with \( \rho_{\text{POT}} > \rho_{\text{BM}} \) the BM-MLE should yield better results. For the six models with \( \rho_{\text{POT}} = \rho_{\text{BM}} \), given that the rate of convergence for the two approaches will be the same, the superiority is determined by the balance between asymptotic bias and variance. We thus do not have a clear prior guess on the superiority of one of the two methods.

Two targets were investigated: the extreme value index \( \gamma \) and the 0.999-quantile of the distribution under consideration. The latter was estimated by either the right-hand side of (1.3) with \((a, b, \gamma)\) replaced by the GEV-MLE based on block maxima of size \( r \), or by the right-hand side of (1.4) with \( t = X_{n-k}, 1 - F(t) = k/n \) and \((\sigma(t), \gamma)\) replaced by the GPD-MLE based on exceedances of \( X_{n-k} \).

Further parameters of the simulation experiment were chosen as follows: the sample size was fixed to \( n = 1000 \). The BM-approach was investigated for block length choices \( r \) from \( \{1, 2, \ldots, 10, 12, \ldots, 20, 25, \ldots, 50\} \) (if \( r \) is not a divisor of \( n = 1000 \), then a small block at the end of the observation period was discarded), and the number of exceedances \( k \) for the POT-approach \( k \) was chosen from \( \{20, 40, \ldots, 500, 550, \ldots, 800, 900, 1000\} \). \( N = 3000 \) repetitions have been performed.

The results are summarized in Figures 2 and 3: for each model, the MSE of the estimator is plotted against the effective sample size \( k \) (i.e., \( k = n/r \) for the BM method). For the estimation of \( \gamma \), the results almost perfectly match the expected behavior described.
above, with the following exceptions: for model S0BM ($\gamma = 0, \rho_{BM} = -2, \rho_{POT} = -1$), the POT-estimator slightly outperforms the BM-estimator. For the six models with $\rho_{POT} = \rho_{BM}$, the BM-approach performs slightly better in four cases and slightly worse in two cases. For the estimation of the 0.999-quantile, mostly the POT approach performs better, with the only (substantial) exceptions being the GEV-models ‘Fréchet’, ‘Gumbel’ and ‘RWeibull’ and the $|t_4|$- and S0-distribution. For the normal distribution and the distribution of $1/(1 + |t_4|)$, the performance is remarkably similar.

2.4 Threshold and block length choice

Both the POT approach and the BM method require a practical selection for the intermediate sequence $k = k_n$ in a sample of size $n$. In the POT approach, the choice of $k$ problem can be interpreted as the choice of the threshold above which the POT approximation in (1.5) is regarded as sufficiently accurate. Similarly, in the BM method, $k$ is related to $r = n/k$, which is the size of the block of which the GEV approximation to the block maximum is regarded as sufficiently accurate.

The theoretical conditions that $k \to \infty$ and $k/n \to 0$, as $n \to \infty$ are useless in guiding
Fig 3. Mean-squared-error for the estimation of $F^{-1}(0.999)$ in the eighteen models from Table 3 for various choices of the effective sample size $k$ (i.e., $k = n/r$ for BM and $k = k$ for POT).

the practical choice. Practically, often a plot between the estimates based on various $k$ against the values of $k$ is made for resolving this problem. The ultimate choice is then made by taking a $k$ from the first stable region in this plot. Nevertheless, the estimators are often rather sensitive to the choice of $k$.

For the POT approach based on the sample of excess ratios (see Footnote 1), there exist a few attempts on resolving the choice of $k$ issue in a formal manner. For example, one solution is to find the optimal $k$ that minimizes the asymptotic MSE; see, e.g., Danielsson et al. (2001), Drees and Kaufmann (1998) and Guillou and Hall (2001) for the Hill estimator ($\gamma > 0$) and Draisma et al. (1999) for the moment estimator ($\gamma \in \mathbb{R}$).

As an indirect solution to the problem, one may also rely on bias corrections, which typically allows for a much larger choice of $k$, see, e.g., Gomes et al. (2008). After the bias correction, the plot of estimates against various values of $k$ usually shows a stable behavior and the estimates are less sensitive to the choice of $k$. For an extensive review on bias corrections, see Beirlant et al. (2012).

\footnote{For the Hill estimator, such a plot is commonly called the “Hill plot”; see Drees et al. (2000).}
Compared to the extensive studies on the threshold choice and on bias corrections for the POT approach, there is, to the best of our knowledge, no existing literature addressing these issues for the univariate BM method. This may partly be explained by the fact that block sizes are often given by the problem at hand, for instance, block sizes corresponding to year. Nevertheless, based on the recent solid theoretical advances on the BM method, the foundations are laid to explore these issues in a rigorous manner in the future (see also Zou et al., 2020 for bias corrections for the BM method in the multivariate time series case discussed below).

3. POT AND BM FOR UNIVARIATE STATIONARY TIME SERIES

In many practical applications, the discussion from the previous section is not quite helpful: the underlying data sample is not i.i.d., but in fact a stretch of a possibly non-stationary time series. Often, by either restricting attention to a proper time horizon or by some suitable transformation, the time series can at least be assumed to be stationary. Throughout this section, we make the following generic assumption: $(X_t)_{t \in \mathbb{Z}}$ is a strictly stationary univariate time series, and the stationary cdf $F$ satisfies the domain-of-attraction condition (1.1). It is important to note that the parameters $\gamma, a_r$ and $b_r$ only depend on the stationary cdf $F$, and that for instance (1.3) expressing high quantiles of $F$ through these parameters continues to hold for time series. Let us begin by passing over the arguments from Section 1 that eventually led to the BM method and the POT approach.

3.1 The POT approach for time series

Recall that the POT approach is based on the sample of large order statistics $X_{\text{POT}} = \{X_{n-k,n}, \ldots, X_{n,n}\}$. The main motivation that lead us to consider this sample was the marginal limit relation (1.2). Bearing in mind that, under mild extra conditions on the serial dependence (ergodicity, mixing conditions, . . . ), empirical moments are consistent for their theoretical counterparts, it is thus still reasonable to estimate the respective parameters by any form of moment matching, e.g., by PWM. The asymptotic variance of such estimators will however be different from the i.i.d. case in general (a consequence of central limit theorems for time series under mixing conditions).

Consider the ML-method: unlike for i.i.d. data, the sample $X_{\text{POT}}$ cannot be regarded as independent anymore, whence it is in general impossible to derive the (approximate) generalized Pareto likelihood of $X_{\text{POT}}$. As a circumvent, one may ‘do as if’ the likelihood arising in the i.i.d. case is also the likelihood for the time series case (quasi-maximum likelihood), and use essentially the same ML-estimators as for the i.i.d. case. Then, since the latter estimator is in fact also depending on empirical moments only, we still obtain proper asymptotic properties such as consistency and asymptotic normality.

Respective theory can be found in Hsing (1991); Resnick and Stărică (1998) for the Hill estimator and in Drees (2000) for a large class of estimators, including PWM and ML.

---

7For example, for financial applications, the stationarity assumption can often be approximately guaranteed by restricting attention to a time horizon during which few macroeconomic conditions had changed. Similarly, for environmental applications, this can be achieved by restricting attention to observations falling into, say, the summer months.
Most of the estimators have the same bias as in the i.i.d. case, whereas their asymptotic variances depend on the serial dependence structure and are usually higher than those obtained in the i.i.d. case. Since the asymptotic bias shares the same explicit form, bias correction can also be performed in the same way as in the i.i.d. case; see, e.g., de Haan et al. (2016) and Chavez-Demoulin and Guillou (2018).

3.2 The BM method for time series

Recall that the BM method is based on the sample of disjoint block maxima \( X_{\text{BM}} = \{M_{1,r}, \ldots, M_{k,r}\} \), where \( M_{j,r} \) denotes the maximum within the \( j \)th block of observations of size \( r \). The main motivation in Section 1 that lead us to consider this sample as approximately GEV-distributed was the relation

\[
\text{Pr}(M_{1,r} \leq a_r x + b_r) = F^r(a_r x + b_r) \approx G^{\text{GEV}}_{\gamma,0,1}(x),
\]

for large \( r \). The first equality is not true for time series, whence more sophisticated arguments must be found for the BM method to work for time series. In fact, it can be shown that if \( F \) satisfies (1.1), if \( \text{Pr}(M_{1,r} \leq a_r x + b_r) \) is convergent for some \( x \) and if mild mixing conditions on the serial dependence (known as \( D(u_n) \)-conditions) are met, then there exists a constant \( \theta \in [0, 1] \) such that

\[
\lim_{r \to \infty} \text{Pr}(M_{1,r} \leq a_r x + b_r) = (G^{\text{GEV}}_{\gamma,0,1}(x))^\theta
\]

for all \( x \in \mathbb{R} \) (Leadbetter, 1983). The constant \( \theta \) is called the extremal index and can be interpreted as capturing the tendency of the time series that extremal observations occur in clusters. If \( \theta > 0 \), then letting

\[
\tilde{a}_r = a_r \theta^\gamma, \quad \tilde{b}_r = b_r - a_r \frac{1 - \theta^\gamma}{\gamma}
\]

we immediately obtain that

\[
\lim_{r \to \infty} \text{Pr}(M_{1,r} \leq \tilde{a}_r x + \tilde{b}_r) = G^{\text{GEV}}_{\gamma,0,1}(x)
\]

for all \( x \in \mathbb{R} \). Hence, the sample \( X_{\text{BM}} \) is approximately GEV-distributed with parameter \((\tilde{a}_r, \tilde{b}_r, \gamma)\), which can then be estimated by any method of choice. It is important to note that, unless \( \theta = 1 \), \((\tilde{a}_r, \tilde{b}_r)\) are different from \((a_r, b_r)\). Consequently, additional steps must be taken for estimating quantiles of \( F \) via (1.3), see also Section 3.3.1 below. Via (3.1), it is possible to transform between \((a_r, b_r)\) and \((\tilde{a}_r, \tilde{b}_r)\) if the extremal index \( \theta \) is known or estimated. Regarding the estimation of the extremal index, a large variety of estimators has been proposed, which may itself be grouped into four categories: 1) BM-like estimators based on “blocking” techniques (Northrop, 2015; Berghaus and Bücher, 2018), 2) POT-like estimators that rely on threshold exceedances (Ferro and Segers, 2003; Süveges, 2007), 3) estimators that use both principles simultaneously (Hsing, 1993; Robert, 2009; Robert et al., 2009) and 4) estimators which, next to choosing a threshold sequence, require the choice of a run-length parameter (Smith and Weissman, 1994; Weissman and Novak, 1998).
Provided the block length parameter is sufficiently large, the distance between the time points at which the maxima within two successive blocks are attained is likely to be quite large as well, whence the sample $X_{BM}$ can be regarded as approximately independent. As a matter of fact, the literature on statistical theory for the BM method is mostly based on the assumption that $X_{BM}$ is a genuine i.i.d. sample from the GEV-family (see, e.g., Prescott and Walden, 1980; Hosking et al., 1985; Bücher and Segers, 2017, among others). Two approximation errors are thereby completely ignored: the cdf is only approximately GEV, and the sample is only approximately independent. Solid theoretical results taking these errors into account are rare: Bücher and Segers (2018b) treat the ML-estimator in the heavy-tailed case ($\gamma > 0$). The main conclusions are: the sample can safely be regarded as independent, but a bias term may appear which, similar as in Section 2, depends on the speed of convergence in (3.2). Bücher and Segers (2018a) improve upon that estimator by using sliding/overlapping blocks instead of disjoint blocks (the $i$th sliding block maximum is equal to $\max(X_i, X_{i+1}, \ldots, X_{i+r-1})$, see Figure 4 for an illustration). The asymptotic variance of the estimator decreases, while the bias stays the same. Moreover, the resulting plots of the estimates against values of $k$ are much smoother, guiding a simpler choice for the block length parameter.

### 3.3 Comparison between the two methods

Let us summarize the main conceptual differences between the BM method and the POT approach for time series. First of all, POT and BM estimate ‘the same’ extreme value index $\gamma$, but possibly different scaling sequence $\tilde{a}_r, \tilde{b}_r$ and $a_r, b_r$. Second, the sample $X_{BM}$ can be regarded as asymptotically independent (asymptotic variances of estimators are the same as if the sample was i.i.d.), while $X_{POT}$ is serially dependent, possibly increasing asymptotic variances of estimators compared to the i.i.d. case.

Due to the lack of a general theoretical result on the BM method, a theoretical comparison on which method is more efficient along the lines of Section 2 seems out of reach for the moment. In particular, a relationship between the respective second order conditions controlling the bias is yet to be found. However, some insight into the merits and pitfalls of two approaches can be gained by considering the problem of estimating high quantiles and return levels.
3.3.1 Estimating high quantiles

Recall that high quantiles of the stationary distribution can be expressed in terms of \( \alpha_r, \beta_r \), and \( \gamma \), see (1.3). As a consequence, based on the plug-in principle, the POT approach immediately yields estimators for high quantiles. On the other hand, the BM method cannot be used straight-forwardly, as it commonly only provides estimators of \( \tilde{\alpha}_r, \tilde{\beta}_r \), and \( \tilde{\gamma} \). Via (3.1), the latter estimators may be transferred into estimators of \( \alpha_r, \beta_r \), and \( \gamma \) using an additional estimator of the extremal index \( \theta \). It is important to note that the latter estimators typically depend on the choice of one or two additional parameters, and that they are often quite variable. By contrast, the POT approach therefore seems more suitable when estimating high quantiles or, more generally, parameters that only depend on the stationary distribution (such as probabilities of rare events). Recall though that estimators based on the POT approach usually suffer from a higher asymptotic variance due to the serial dependence.

3.3.2 Estimating return levels

Let \( F_r(x) = \Pr(M_1 \leq x) \). For \( T \geq 1 \), the \( T \)-return level of the sequence of block maxima is defined as the \( 1 - 1/T \) quantile of \( F_r \), that is,

\[
RL(T, r) = F_r^{-1}(1 - 1/T) = \inf\{x \in \mathbb{R} : F_r(x) \geq 1 - 1/T\}.
\]

Since block maxima are asymptotically independent, it will take on average \( T \) blocks of size \( r \) until the first such block whose maximum exceeds \( RL(T, r) \). Now, since \( F_r \) is approximately equal to the GEV-cdf with parameters \( \gamma, \tilde{\beta}_r, \tilde{\alpha}_r \) for large \( r \) by (3.2), we obtain that

\[
RL(T, r) \approx \tilde{\beta}_r + \tilde{\alpha}_r \left\{ -r \log(1 - 1/T) \right\}^{-\gamma} - 1 \approx \tilde{\beta}_r + \tilde{\alpha}_r \left( r/T \right)^{-\gamma} - 1.
\]

In comparison to the estimation of high-quantiles, see (1.3), we have now expressed the object of interest in terms of the sequences \( \tilde{\alpha}_r \), \( \tilde{\beta}_r \), and the extreme-value index \( \gamma \). Following the discussion in the previous section, it is now the BM method which yields simpler estimators that do not require additional estimation of the extremal index. By contrast, the POT approach only results in estimators of \((\alpha_r, \beta_r)\) and \( \gamma \), and therefore requires a transformation to \((\tilde{\alpha}_r, \tilde{\beta}_r)\) (via (3.1)) based on an estimate of the extremal index \( \theta \).

4. POT AND BM FOR MULTIVARIATE OBSERVATIONS

Due to the lack of asymptotic results on the multivariate BM method which take the approximation error into account, a deep comparison between the POT approach and the BM method is not feasible yet. Within this section we try to identify the open ends that may eventually lead to such results in the future.

Let \( F \) be a \( d \)-dimensional cdf. The basic assumption of multivariate extreme-value theory, generalizing (1.1), is as follows: suppose that there exists a non-degenerate cdf \( G \) and sequences \((a_{r,j})_{r \in \mathbb{N}}, (b_{r,j})_{r \in \mathbb{N}}, j = 1, \ldots, d \), with \( a_{r,j} > 0 \) such that

\[
\lim_{r \to \infty} \Pr\left( \frac{\max_{i=1}^r X_{i,1} - b_{r,1}}{a_{r,1}} \leq x_1, \ldots, \frac{\max_{i=1}^r X_{i,d} - b_{r,d}}{a_{r,d}} \leq x_d \right) = G(x_1, \ldots, x_d)
\]
for any $x_1,\ldots,x_d \in \mathbb{R}$, where $X_i = (X_{i,1},\ldots,X_{i,d})'$, $i \in \mathbb{N}$, is an i.i.d. sequence from $F$, and where the marginal distributions $G_j$ of $G$, $j = 1,\ldots,d$, are GEV-distributions with location parameter 0, scale parameter 1 and shape parameter $\gamma_j \in \mathbb{R}$ (location 0 and scale 1 can always be reached by adapting the sequences $a_{r,j}$ are $b_{r,j}$ if necessary). The dependence between the coordinates of $G$ can be described in various equivalent ways (see, e.g., Resnick, 1987; Beirlant et al., 2004; de Haan and Ferreira, 2006): by the stable tail dependence function $L$ (Huang, 1992), by the exponent measure $\mu$ (Balkema and Resnick, 1977), by the Pickands dependence function $A$ (Pickands, 1981), by the tail copula $\Lambda$ (Schmidt and Stadtmüller, 2006), by the spectral measure $\Phi$ (de Haan and Resnick, 1977), by the madogram $\nu$ (Naveau et al., 2009), or by other less popular objects. All these objects are in one-to-one correspondence, and for each of them a large variety of estimators has been proposed, both in a nonparametric way and under the assumption that the objects are parametrized by an Euclidean parameter.

In this paper, we will mainly focus on nonparametric estimation. As in the univariate case, the estimators may again be grouped into POT and BM based estimators, see Sections 4.1 and 4.2 below. Often, estimation of the marginal parameters and of the dependence structure is treated successively. It is important to note that standard errors for estimators of the dependence structure may then be influenced by standard errors for the marginal estimation, a point which is often ignored in the literature on statistics for multivariate extremes. In fact, a phenomenon well-known in statistics for copulas (Genest and Segers, 2010) may show up: possibly completely ignoring additional information about the marginal cdfs, estimators for the dependence structure may have a lower asymptotic variance if marginal cdfs are estimated nonparametrically; see Bücher (2014) for a discussion of the empirical stable tail dependence function from Section 4.1 below, and Genest and Segers (2009) for estimation of Pickands dependence function based on i.i.d. data from a bivariate extreme value distribution, Section 4.2 below.

4.1 The POT approach in the multivariate case

Suppose $X_1,\ldots,X_n$, with $X_i = (X_{i,1},\ldots,X_{i,d})'$, is an i.i.d. sample from $F$. Recall that the univariate POT approach was based on the observations $X_{\text{POT}} = \{X_{n-k,n},\ldots,X_{n,n}\}$, which may be rewritten as $X_{\text{POT}} = \{X_i \mid \text{rank}(X_i \text{ among } X_1,\ldots,X_n) \geq n - k\}$. Thus, a possible generalization to multivariate observations consists of defining

$$X_{\text{POT}} = \{X_i \mid \text{rank}(X_{i,j} \text{ among } X_{1,j},\ldots,X_{n,j}) \geq n - k \text{ for some } j = 1,\ldots,d\},$$

that is, $X_{\text{POT}}$ comprises all observations for which at least one coordinate is large. Any estimator defined in terms of these observations may be called an estimator based on the multivariate POT approach.

As an example, consider the estimation of the so-called stable tail dependence function $L$, which is defined as

$$L(x) = \lim_{t \downarrow 0} \frac{1}{t} \Pr(F_1(X_1) > 1 - tx_1 \text{ or } \ldots \text{ or } F_d(X_d) > 1 - tx_d),$$

where $x = (x_1,\ldots,x_d)' \in [0,1]^d$; a limit that necessarily exists under (4.1), but may also exist for marginals $F_j$ not in any domain-of-attraction. The function $L$ can be estimated
by its empirical counterpart, defined as

\[ \hat{L}(x_1, \ldots, x_d) = \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}(\hat{F}_{n,1}(X_{i,1}) > 1 - \frac{k}{n} x_1 \text{ or } \ldots \text{ or } \hat{F}_{n,d}(X_{i,d}) > 1 - \frac{k}{n} x_d), \]

where \( \hat{F}_{n,j} \) denotes the empirical cdf based on the observations \( X_{1,j}, \ldots, X_{n,j} \); see, e.g., Huang (1992). Since \( x \in [0, 1]^d \), the estimator in fact only depends on the sample \( X_{\text{pot}} \).

Suppose the following natural second order condition quantifying the speed of convergence in (4.2) is met: there exists a positive or negative function \( A \) and a real-valued function \( g \neq 0 \) such that

\[ \lim_{t \to \infty} \frac{t \Pr(F_1(X_1) > 1 - \frac{x_1}{t} \text{ or } \ldots \text{ or } F_d(X_d) > 1 - \frac{x_d}{t} \} - L(x_1, \ldots, x_d)}{A(t)} = g(x) \]

uniformly in \( x \in [0, 1]^d \). Then, under additional smoothness conditions on \( L \), it can be shown that \( \hat{L} \) is consistent and asymptotically Gaussian in terms of functional weak convergence, the variance being of order \( 1/k \) and the bias being of order \( A(n/k) \), provided that \( k = k_n \to \infty \) and \( k/n \to 0 \) as \( n \to \infty \); see, e.g., Huang (1992); Einmahl et al. (2012), among others. Following the discussion in Section 2, if we additionally assume that \( A(t) \asymp t^\rho \) for some \( \rho \in (-\infty, 0) \), the best attainable convergence rate, achieved when squared bias and variance are balanced, is

\[ \text{Rate of Convergence of } \hat{L}(x) = n^{\rho/(1-2\rho)}. \]

This convergence rate is in fact optimal under additional conditions on the data-generating process, see Drees and Huang (1998). Also note that \( \hat{L} \) suffers from an asymptotic bias as in the univariate case, and that corresponding bias corrections for the bivariate case have been proposed in Fougères et al. (2015) and Beirlant et al. (2016).

As in the univariate case, the literature on further theoretical foundations for the multivariate POT approach is vast, see, e.g., Einmahl et al. (2001); Einmahl and Segers (2009) for nonparametric estimation of the spectral measure, Drees and de Haan (2015) for estimation of failure probabilities, or de Haan et al. (2008); Einmahl et al. (2012) for parametric estimators, among many others.

An alternative way using the multivariate POT approach is to consider the multivariate Generalized Pareto distribution (MGPD). Rootzén and Tajvidi (2006) shows that the domain of attraction condition (4.1) is equivalent to the following limit relation: there exists an increasing \( d \)-dimensional curve \( u(t) = (u_1(t), \ldots, u_d(t))^t \) for \( t \in [1, +\infty) \) and a \( d \)-dimensional function \( \sigma(t) = (\sigma_1(t), \ldots, \sigma_d(t))^t > 0 \) such that \( u(1) = 0 \) and as \( t \to \infty \), \( F(u(t)) \to 1 \),

\[ \Pr\left( \frac{X_i - u_i(t)}{\sigma_i(t)} \leq x_i \bigg| X_i \geq u_i(t) \text{ for some } i = 1, \ldots, d \right) \to H(x_1, \ldots, x_d), \]

for all \( x = (x_1, \ldots, x_d)^t \in (0, +\infty)^d \). The limit distribution \( H(x) \) is called the MGPD and is related to the stable distribution function \( L \) but also involves marginal information such as the marginal extreme value indices.
The theoretical properties of MGPD have been studied in Falk and Guillou (2008), Rootzén et al. (2018a) and Rootzén et al. (2018b), among others. There is also an emerging literature focusing on statistical analysis for the MGPD model, see, e.g., Thibaud and Opitz (2015), Huser et al. (2016) and de Fondeville and Davison (2018). Nevertheless, most of the statistical studies focus only on a parametric form of the MGPD model. Nonparametric estimation similar to the estimator $\hat{L}$ is yet open for future research.

4.2 The BM method in the multivariate case

Again suppose $X_1, \ldots, X_n$ is an i.i.d. sample from $F$. Let $r$ denote a block size, and $k = \lfloor n/r \rfloor$ the number of blocks. For $\ell = 1, \ldots, k$, let $M_{\ell,r} = (M_{\ell,1,r}, \ldots, M_{\ell,1,r})'$ denote the vector of componentwise block-maxima in the $\ell$th block of observations of size $r$ (it is worthwhile to note that $M_{\ell,r}$ may be different from any $X_i$). Any estimator defined in terms of the sample $X_{\text{BM}} = (M_{1,r}, \ldots, M_{k,r})$ is called an estimator based on the BM method.

Just as for the univariate BM method, asymptotic theory is usually formulated under the assumption that $M_1, \ldots, M_k$ is a genuine i.i.d. sample from the limiting distribution $G$; a potential bias is completely ignored. Moreover, estimation of the marginal parameters is often disentangled from estimation of the dependence structure, with theory for the latter either developed under the assumption that marginals are completely known (which usually leads to wrong asymptotic variances), or under the assumption that marginals are estimated nonparametrically. See, for instance, Pickands (1981); Capéraà et al. (1997); Zhang et al. (2008); Genest and Segers (2009); Gudendorf and Segers (2012) for nonparametric estimators and Genest et al. (1995); Dombry et al. (2016) for parametric ones, among many others.

To the best of our knowledge, references that takes the approximation error induced by the assumption of observing data from a genuine extreme-value model into account are rare. Bücher and Segers (2014) consider estimation of the Pickands dependence function $A$ based on the BM-method. Not only the bias is treated carefully there, but also the underlying observations $X_1, \ldots, X_n$ may possess serial dependence in form of a stationary time series. Just like in the univariate case described above, the best attainable convergence rate of the estimator again depends on a second order condition. A recent follow-up paper, Zou et al. (2020), discusses the use of sliding block maxima and bias corrections based on suitable aggregation over multiple block sizes.

4.3 Comparison between the two methods

As already discussed in the univariate case, a comparison between the two methods may be carried out by investigating the relationship between respective second order conditions. In analogy to the univariate result in Drees et al. (2003), such a comparison has been recently worked out for the serially independent case in Bücher et al. (2019). More precisely, the POT second order condition in (4.4) has been compared to the following block maxima second order condition:
There exists a positive or negative function $B$ and a function $h \neq 0$ such that

$$
\lim_{r \to \infty} \frac{\Pr(F_1(X_1) \leq u_1^{1/r}, \ldots, F_d(X_d) \leq u_d^{1/r} - C_\infty(u)}{B(t)} = h(u)
$$

uniformly in $u \in [\delta, 1]^d$ and for each $\delta > 0$. Here, $C_\infty(u) = \exp\{-L(-\log u_1, \ldots, -\log(u_d))\}$ denotes the extreme-value copula associated with the stable tail dependence function $L$ in (4.2).

Note that this is a suitable condition for investigating certain block maxima estimators, such as those in Bücher and Segers (2014). Both $A$ in (4.4) and $B$ in (4.5) are regularly varying with a corresponding second order parameter, say $\rho_{\text{POT}}$ and $\rho_{\text{BM}}$, respectively. The two second order parameters are shown to satisfy a similar relationship as in the univariate case: equality in the range $\rho_{\text{POT}} = \rho_{\text{BM}} \in (-1, 0]$, and one of the two possibly being lower than $-1$ while the other is equal to $-1$. Again, the lower of the two parameters typically results in a more efficient estimation of the respective method.

The superiority of one of the two estimators may also be illustrated by finite-sample simulation results, see Figure 5 taken from Bücher et al. (2019). The upper row of the figure corresponds to a block maxima estimator (details can be found in the cited reference), while the lower row corresponds to the empirical stable tail dependence function in (4.3), evaluated at $(1/2, 1/2)$ (bivariate case). The three columns correspond to models where $\rho_{\text{BM}} = -1 = \rho_{\text{POT}}$ (Model $\psi_1$), $\rho_{\text{BM}} = -2 < -1 = \rho_{\text{POT}}$ (Model $\psi_2$) and $\rho_{\text{BM}} = -1 > -2 = \rho_{\text{POT}}$ (Model $\psi_3$). As expected from the parameters of the models, both estimators perform similarly for the first model, while the BM estimator outperforms the POT estimator in Model 2 and vice versa for Model 3.

### 4.4 Multivariate time series

Moving from i.i.d. multivariate observations to multivariate strictly stationary time series induces similar phenomena as in the univariate case, whence we keep the discussion quite short. Under suitable conditions on the serial dependence, estimators based on the POT approach are still consistent and asymptotically normal, though with a possibly different asymptotic variance (this can for instance be deduced from Drees and Rootzén, 2010). Regarding the BM method, the same heuristics as in the univariate case apply: block maxima may safely be assumed as independent and as following a multivariate extreme value distribution (Bücher and Segers, 2014). The estimators based on the BM method are then also consistent and asymptotically normal with a potential bias. Similar to the discussion on the location and scale parameters in the univariate case, the objects that are estimated by POT and BM may be different but are linked by the multivariate extremal index (Nandagopalan, 1994, see also Section 10.5 in Beirlant et al., 2004).

Hence, following the discussion in Section 3.3, it seems preferable to estimate quantities that only depend on the tail of the stationary distribution by the POT approach, while tail quantities similar to the univariate return levels (that also depend on the serial dependence) are preferably estimated by the BM method. As in the univariate case, a detailed theoretical comparison does not seem to be feasible.
5. POT AND BM FOR STOCHASTIC PROCESSES

The BM method for stochastic processes is based on modeling by max-stable processes, i.e., on limit models arising for block maxima taken over i.i.d. stochastic processes. Recent research has focussed on the structure and characteristics of max-stable processes, see, e.g., de Haan (1984), Giné et al. (1990) and Kabluchko et al. (2009); on simulating from max-stable processes, see, e.g., Dombry et al. (2013), Dieker and Mikosch (2015), Dombry et al. (2016) and Oesting et al. (2018); and on statistical inference based on max-stable processes, see, e.g., Coles and Tawn (1996), Buishand et al. (2008), Padoan et al. (2010) and Huser and Davison (2014).

As mentioned in the introduction, there is a clear supply issue regarding the POT approach to stochastic process models. Early studies such as Einmahl and Lin (2006) consider the estimation of marginal parameters only, or consider nonparametric estimation of the dependence structure (de Haan and Lin, 2003), however with only weak consistency established. Recent development on Generalized Pareto Processes allow for considering parametric estimation for the dependence structure, see, e.g., Ferreira and de Haan (2014), Thibaud and Opitz (2015) and Huser and Wadsworth (2019). Given the imbalanced nature, we skip a deeper review on the POT approach and the BM method for extremes regarding stochastic processes.

6. OPEN PROBLEMS

Throughout this paper, we have already identified a number of open research problems, mostly related to an honest verification of the BM method. Within the following list, we
recapitulate those issues and add several further possible research questions:

- Asymptotic theory on further estimators based on the block maxima method, if possible allowing for a comparison between the imposed second order condition and those from the POT approach.
- Solid theoretical comparisons between BM and POT estimators for other univariate tail related quantities, such as high quantiles and tail probabilities.
- In case the BM method yields to faster attainable rates of convergence than the POT approach (Section 2.1): are the obtained rates optimal?
- Derive a test for which approach is preferably for a given data set \( H_0 : \rho_{BM} \leq \rho_{POT} \), or similar).
- Block length choice and bias reduction for BM. Choice of \( k \) and bias reduction for fitting the GPD to the sample of threshold exceedances using POT.
- More results on the sliding block maxima method (e.g., on the univariate non-heavy tailed case).
- A comparison of return level/quantile estimation based on POT and BM, possibly incorporating an estimator for the extremal index.
- Extension to stochastic processes (max-stable processes and generalized Pareto processes): theoretical results on statistical methodology are still rare, and a comparison between POT and BM is not feasible yet.

7. CONCLUSION

There is no winner.

ACKNOWLEDGEMENTS

Axel Bücher’s research has been supported by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation, which is gratefully acknowledged. The authors are grateful to two referees and an associate editor for their constructive comments on an earlier version of this article which lead to a substantial improvements. They are particularly grateful for pointing out that Section 9.3 in Reiss (1989) provides some early results on the BM-method that do take care of the approximation error in the GEV approximation to block maxima.

REFERENCES


