

# Maximum likelihood estimators based on the block maxima method

CLÉMENT DOMBRY<sup>1</sup> and ANA FERREIRA<sup>2</sup>

<sup>1</sup>*Laboratoire de Mathématiques de Besançon, UMR CNRS 6623, Univ. Bourgogne Franche-Comté, 16 route de Gray, 25030 Besançon cedex, France. E-mail: [clement.dombry@univ-fcomte.fr](mailto:clement.dombry@univ-fcomte.fr)*

<sup>2</sup>*Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal. E-mail: [anafh@tecnico.ulisboa.pt](mailto:anafh@tecnico.ulisboa.pt)*

The extreme value index is a fundamental parameter in univariate Extreme Value Theory (EVT). It captures the tail behavior of a distribution and is central in the extrapolation beyond observations. Among other semi-parametric methods (such as the popular Hill estimator), the Block Maxima (BM) and Peaks-Over-Threshold (POT) methods are widely used for assessing the extreme value index and related normalizing constants. We provide asymptotic theory for the maximum likelihood estimators (MLE) based on the BM method for independent and identically distributed observations in the max-domain of attraction of some extreme value distribution. Our main result is the asymptotic normality of the MLE with a non-trivial bias depending on the extreme value index and on the so-called second-order parameter. Our approach combines asymptotic expansions of the likelihood process and of the empirical quantile process of block maxima. The results permit to complete the comparison of common semi-parametric estimators in EVT (MLE and probability weighted moment estimators based on the POT or BM methods) through their asymptotic variances, biases and optimal mean square errors.

*Keywords:* asymptotic normality; block maxima method; extreme value index; maximum likelihood estimator; peaks-over-threshold method; probability weighted moment estimator

## 1. Introduction

The Block Maxima (BM) method, also known as Annual Maxima, after Gumbel [15] is a fundamental method in Extreme Value Theory and has been widely used. The method is justified under the Maximum Domain of Attraction (MDA) condition: for an independent and identically distributed (i.i.d.) sample with distribution function  $F$ , if the linearly normalized partial maxima converge in distribution, then the limit must be a Generalized Extreme Value (GEV) distribution. In practice, one rarely exactly knows  $F$  but the MDA condition holds for most common continuous distributions. Procedures for testing MDA have been proposed by Dietrich et al. [7], Drees et al. [11], cf. also de Haan and Ferreira [6], Markovich [18] and Novak [19].

In the BM method, the initial sample is divided into blocks of the same size and the MDA condition ensures that the block maxima are approximately GEV distributed. The method is commonly used in hydrology and other environmental applications or in insurance and finance when analysing extremes – see, for example, the monographs by Embrechts et al. [13], Coles [5], Beirlant et al. [2], de Haan and Ferreira [6] and references therein.

The GEV is a three-parameter distribution, with the usual location and scale parameters, and the extreme value index being the main parameter as it characterizes the heaviness of the tail.

Several estimators have been proposed, including the classical maximum likelihood (ML) and probability weighted moments (PWM) estimators (Hosking et al. [17]). Asymptotic normality of these estimators has been established for a sample from the GEV distribution, which holds with null bias and explicit covariance matrix (Prescott and Walden [21], Hosking et al. [17], Smith [23], Bücher and Segers [3]). The underlying theory is quite difficult and technical by the fact that the support of the GEV depends on the parameter values. Differentiability in quadratic mean of the GEV model has been proven only recently by Bücher and Segers [3].

However, in applications, the sample block maxima are only approximately GEV so that the classical parametric theory suffers from model misspecification. In this paper, we intend to fill this gap for ML estimators (MLE), by showing asymptotic normality under a flexible second-order condition (a refinement of the MDA condition). Depending on the asymptotic block size, a non trivial bias may appear in the limit for which we provide an exact expression. Recently, Ferreira and de Haan [14] showed asymptotic normality of the PWM estimators under the same conditions. They derived a uniform expansion for the empirical quantile process of block maxima that is a crucial tool in our approach as well. Indeed, the MLE can be seen as a maximizer of the so-called likelihood process. By expressing the likelihood process in terms of this empirical quantile process, we are able to derive an expansion of the likelihood process that implies the asymptotic normality of the MLE. It turns out that this derivation is again quite technical by the fact that the support of the GEV depends on the parameter values. Note that the asymptotic normality for the MLE of a Fréchet distribution based on the block maxima of a stationary heavy-tailed time series has been obtained by Bücher and Segers [4]. There the issue of parameter dependent supports is avoided but time dependence has to be dealt with. Besides, the ideas underlying their proof are quite different.

The asymptotic normality result in the present paper brings novel developments to the theoretical comparison of the main semi-parametric estimation procedures in EVT. On the one hand, it permits to compare BM and Peaks-over-Threshold (POT) methods (see, e.g., Balkema and de Haan [1], Pickands [20]), the latter being another fundamental method in EVT and a competitor of BM, where the highest observations follow approximately a Generalized Pareto distribution. The POT version with random threshold, that is, taking a non-random number of higher observations corresponding to an upper intermediate order statistic as threshold, is the one more directly linked to the most common BM approach where the number of blocks is non-random. We discuss and compare the four different approaches – MLE/PWM estimators in the BM/POT approaches – based on exact theoretical formulas for asymptotic variances, biases and optimal mean square errors depending on the extreme value index and the second-order parameter. It turns out that MLE under BM has minimal asymptotic variance among all combinations MLE/PWM and BM/POT but, on the other hand it has some significant asymptotic bias. When analysing the asymptotic optimal mean square error that balances variance and bias, the most efficient combination turns out to be MLE under POT (e.g., Drees, Ferreira and de Haan [12]). It turns out that the optimal sample size is larger for POT-MLE than for BM-MLE, giving a theoretical justification to the heuristic that POT allows for a better use of the data than BM.

The outline of the paper is as follows: In Section 2, we present the main theoretical conditions and results including Theorem 2.2 giving the asymptotic normality of the MLE. In Section 3, we present a comparative study of asymptotic variances and biases, optimal asymptotic mean square errors and optimal sample sizes among all combinations MLE/PWM and BM/POT. In

Section 4, we state additional theoretical statements, including the local asymptotic normality of MLE under the fully parametric GEV model, and provide all the proofs. Finally, Appendix A gathers some formulas for the information matrix and for the bias of BM-MLE and Appendix B provides useful bounds for the derivatives of the likelihood function that are necessary for the main proofs.

## 2. Asymptotic behaviour of MLE

### 2.1. Framework and notations

The GEV distribution with index  $\gamma$  is defined by

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

and the corresponding log-likelihood by

$$g_\gamma(x) = \begin{cases} -(1 + 1/\gamma) \log(1 + \gamma x) - (1 + \gamma x)^{-1/\gamma} & \text{if } 1 + \gamma x > 0, \\ -\infty & \text{otherwise.} \end{cases} \quad (2.1)$$

For  $\gamma = 0$ , the formula is interpreted as  $g_0(x) = -x - \exp(-x)$ . The three-parameter model with index  $\gamma$ , location  $\mu$  and scale  $\sigma > 0$  is defined by the log-likelihood

$$\ell(\theta, x) = g_\gamma\left(\frac{x - \mu}{\sigma}\right) - \log \sigma, \quad \theta = (\gamma, \mu, \sigma). \quad (2.2)$$

A distribution  $F$  is said to belong to the max-domain of attraction of the extreme value distribution  $G_{\gamma_0}$ , denoted by  $F \in D(G_{\gamma_0})$ , if there exist normalizing sequences  $a_m > 0$  and  $b_m$  such that

$$\lim_{m \rightarrow +\infty} F^m(a_m x + b_m) = G_{\gamma_0}(x), \quad \text{for all } x \in \mathbb{R}.$$

When using the BM method main steps are the estimation of the extreme value index  $\gamma$  as well as the normalizing constants  $a_m$  and  $b_m$ . The set-up is the following. Consider independent and identically distributed (i.i.d.) random variables  $(X_i)_{i \geq 1}$  with common distribution function  $F \in D(G_{\gamma_0})$ . Divide the sequence  $(X_i)_{i \geq 1}$  into blocks of length  $m \geq 1$  and define the  $k$ -th block maximum by

$$M_{k,m} = \max_{(k-1)m < i \leq km} X_i, \quad k \geq 1. \quad (2.3)$$

For each  $m \geq 1$ , the variables  $(M_{k,m})_{k \geq 1}$  are i.i.d. with distribution function  $F^m$  and by the max-domain of attraction condition

$$\frac{M_{k,m} - b_m}{a_m} \xrightarrow{d} G_{\gamma_0} \quad \text{as } m \rightarrow +\infty. \quad (2.4)$$

This suggests that the distribution of  $M_{k,m}$  is *approximately* a GEV distribution with parameters  $(\gamma_0, b_m, a_m)$ . The method consists in pretending that the sample follows *exactly* the GEV distribution and in maximizing the GEV log-likelihood so as to compute the MLE. A particular feature of the method is that the model is clearly misspecified since the GEV distribution appears as the limit distribution of the block maxima as the block size  $m$  tends to  $+\infty$  while in practice we have to use a finite block size. As seen afterwards, we quantify the misspecification thanks to the so-called second-order condition that implies an asymptotic expansion of the empirical quantile process with a non trivial bias term. When plugging this expansion in the ML equations, we obtain a bias term for the likelihood process as well as for the MLE.

The (misspecified) log-likelihood of the  $k$ -sample  $(M_{1,m}, \dots, M_{k,m})$  is

$$L_{k,m}(\theta) = \sum_{i=1}^k \ell(\theta, M_{i,m}), \quad \theta = (\gamma, \mu, \sigma) \in \Theta = \mathbb{R} \times \mathbb{R} \times (0, +\infty). \tag{2.5}$$

We say that an estimator  $\hat{\theta}_n = (\hat{\gamma}_n, \hat{\mu}_n, \hat{\sigma}_n)$  is a MLE if it is a *local maximizer* of  $L_{k,m}$ . A MLE solves the so-called *score equations*

$$\begin{cases} \frac{\partial}{\partial \gamma} L_{k,m}(\gamma, \mu, \sigma) = 0, \\ \frac{\partial}{\partial \mu} L_{k,m}(\gamma, \mu, \sigma) = 0, \\ \frac{\partial}{\partial \sigma} L_{k,m}(\gamma, \mu, \sigma) = 0, \end{cases}$$

which we write shortly in vectorial notation

$$\frac{\partial L_{k,m}}{\partial \theta}(\theta) = 0. \tag{2.6}$$

Furthermore, if the score equation has a solution where Hessian matrix is positive definite, then this point is a MLE.

A main purpose of this paper is to study the existence and asymptotic normality of the MLE under the following conditions (for more technical details on regularly varying properties we refer to [6]):

- First-order condition:

$$F \in D(G_{\gamma_0}) \quad \text{with } \gamma_0 > -\frac{1}{2}.$$

Note that the also called first-order condition (2.4) is equivalent to

$$\lim_{m \rightarrow \infty} \frac{V(mx) - V(m)}{a_m} = \frac{x^{\gamma_0} - 1}{\gamma_0}, \quad x > 0,$$

with  $V = (-1/\log F)^\leftarrow$  ( $\leftarrow$  denotes the left-continuous inverse function). W.l.g., we can take  $b_m = V(m)$  in Equation (2.4), what we shall assume in the following.

- Second-order condition: for some positive function  $a$  and some positive or negative function  $A$  with  $\lim_{t \rightarrow \infty} A(t) = 0$ ,

$$\lim_{t \rightarrow \infty} \frac{\frac{V(tx) - V(t)}{a(t)} - \frac{x^{\gamma_0} - 1}{\gamma_0}}{A(t)} = \int_1^x s^{\gamma_0 - 1} \int_1^s u^{\rho - 1} du ds = H_{\gamma_0, \rho}(x), \quad x > 0, \quad (2.7)$$

with  $\gamma_0 > -\frac{1}{2}$ , where

$$H_{\gamma_0, \rho}(x) = \begin{cases} \frac{1}{\rho} \left( \frac{x^{\gamma_0 + \rho} - 1}{\gamma_0 + \rho} - \frac{x^{\gamma_0} - 1}{\gamma_0} \right), & \rho \neq 0 \neq \gamma_0, \\ \frac{1}{\gamma_0} \left( x_0^\gamma \log x - \frac{x^{\gamma_0} - 1}{\gamma_0} \right), & \rho = 0 \neq \gamma_0, \\ \frac{1}{\rho} \left( \frac{x^\rho - 1}{\rho} - \log x \right), & \rho \neq 0 = \gamma_0, \\ \frac{1}{2} (\log x)^2, & \rho = 0 = \gamma_0. \end{cases}$$

Consequently,  $\rho \leq 0$  and  $|A|$  is regularly varying with index  $\rho$ .

- Asymptotic growth for the number  $k$  of blocks and the block size  $m$ :

$$k = k_n \rightarrow \infty, \quad m = m_n \rightarrow \infty \quad \text{and} \quad \sqrt{k} A(m) \rightarrow \lambda \in \mathbb{R}, \quad \text{as } n \rightarrow \infty. \quad (2.8)$$

### 2.2. Main results

Before considering the MLE, we focus on the asymptotic properties of the likelihood and score processes. For the purpose of asymptotic, we introduce the *local parameter*  $h = (h_1, h_2, h_3) \in \mathbb{R}^3$ :

$$\begin{cases} h_1 = \sqrt{k}(\gamma - \gamma_0), \\ h_2 = \sqrt{k}(\mu - b_m)/a_m, \\ h_3 = \sqrt{k}(\sigma/a_m - 1) \end{cases} \Leftrightarrow \begin{cases} \gamma = \gamma_0 + h_1/\sqrt{k}, \\ \mu = b_m + a_m h_2/\sqrt{k}, \\ \sigma = a_m(1 + h_3/\sqrt{k}). \end{cases} \quad (2.9)$$

Set  $\theta_0 = (\gamma_0, 0, 1)$ . The *local log-likelihood process* at  $\theta_0$  is given by

$$\begin{aligned} \tilde{L}_{k,m}(h) &= L_{k,m} \left( \gamma_0 + \frac{h_1}{\sqrt{k}}, b_m + a_m \frac{h_2}{\sqrt{k}}, a_m + a_m \frac{h_3}{\sqrt{k}} \right) \\ &= \sum_{i=1}^k \ell \left( \theta_0 + \frac{h}{\sqrt{k}}, \frac{M_{i,m} - b_m}{a_m} \right) - k \log(a_m), \end{aligned} \quad (2.10)$$

and, the local score process by

$$\begin{aligned} \frac{\partial \tilde{L}_{k,m}}{\partial h}(h) &= \frac{1}{\sqrt{k}} \sum_{i=1}^k \frac{\partial \ell}{\partial \theta} \left( \theta_0 + \frac{h}{\sqrt{k}}, \frac{M_{i,m} - b_m}{a_m} \right) \\ &= \frac{1}{\sqrt{k}} \frac{\partial L_{k,m}}{\partial \theta}(\theta). \end{aligned} \tag{2.11}$$

Clearly, the score equation (2.6) rewrites in this new variable as  $\frac{\partial \tilde{L}_{k,m}}{\partial h}(h) = 0$ .

In the following,  $Q_{\gamma_0}$  denote the quantile function of the extreme value distribution  $G_{\gamma_0}$ , that is,

$$Q_{\gamma_0}(s) = \frac{(-\log s)^{-\gamma_0} - 1}{\gamma_0}, \quad s \in (0, 1). \tag{2.12}$$

**Proposition 2.1.** *Let  $X_1, X_2, \dots$  be independent and identically distributed random variables with cumulative distribution function  $F \in D(G_{\gamma_0})$  with  $\gamma_0 > -1/2$ , verifying the second-order condition (2.7). Let  $k = k_n \rightarrow \infty$  and  $m = m_n \rightarrow \infty$  be such that  $\sqrt{k}A(m) \rightarrow \lambda \in \mathbb{R}$ , with  $A$  the second-order auxiliary function from (2.7), and  $r = r_n \rightarrow \infty$  be a sequence of positive numbers such that*

$$r_n = O(k_n^\delta), \quad 0 < \delta < \min(1/2, \gamma_0 + 1/2), \text{ as } n \rightarrow \infty. \tag{2.13}$$

Let  $H_n \subset \mathbb{R}^3$  be the ball of centre 0 and radius  $r_n$ . Then, uniformly for  $h \in H_n$ ,

$$\frac{\partial^2 \tilde{L}_{k,m}}{\partial h \partial h^T}(h) = -I_{\theta_0} + o_P(1) \tag{2.14}$$

with  $I_{\theta_0}$  the Fisher information matrix

$$I_{\theta_0} = - \int_0^1 \frac{\partial^2 \ell}{\partial \theta \partial \theta^T}(\theta_0, Q_{\gamma_0}(s)) ds. \tag{2.15}$$

As a consequence, the local log-likelihood process  $\tilde{L}_{k,m}$  is strictly concave on  $H_n$  with probability tending to one.

**Remark 2.1.** The conditions in Proposition 2.1 are sufficient for consistency of MLE, see Domby [8]. In particular  $\sqrt{k}A(m) \rightarrow \lambda \in \mathbb{R}$  implies  $m/\log n \rightarrow \infty$ , the later required for consistency. When  $\gamma_0 \geq 0$ , condition (2.13) implies that (2.14) holds for  $h = o(k^{1/2-\varepsilon})$ ,  $\varepsilon > 0$ .

Our main result is the following theorem establishing the asymptotic behavior of the local likelihood process and from which the existence and asymptotic normality of the MLE will be deduced.

**Theorem 2.1.** *Let  $X_1, X_2, \dots$  be independent and identically distributed random variables with cumulative distribution function  $F \in D(G_{\gamma_0})$  with  $\gamma_0 > -1/2$ , verifying the second-order condition (2.7). Let  $k = k_n \rightarrow \infty$  and  $m = m_n \rightarrow \infty$  be such that  $\sqrt{k}A(m) \rightarrow \lambda \in \mathbb{R}$ .*

Then, the local likelihood process satisfies uniformly for  $h$  in compact sets

$$\tilde{L}_{k,m}(h) = \tilde{L}_{k,m}(0) + h^T \tilde{G}_{k,m} - \frac{1}{2} h^T I_{\theta_0} h + o_P(1), \tag{2.16}$$

$$\frac{\partial \tilde{L}_{k,m}}{\partial h}(h) = \tilde{G}_{k,m} - I_{\theta_0} h + o_P(1), \tag{2.17}$$

where

$$\tilde{G}_{k,m} = \frac{1}{\sqrt{k}} \sum_{i=1}^k \frac{\partial \ell}{\partial \theta} \left( \theta_0, \frac{M_{i,m} - b_m}{a_m} \right) \xrightarrow{d} \mathcal{N}(\lambda b, I_{\theta_0}) \tag{2.18}$$

i.e., it is asymptotically Gaussian with covariance matrix equal to the information matrix and mean-vector depending on the second-order condition through

$$b = b(\gamma_0, \rho) = \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, Q_{\gamma_0}(s)) H_{\gamma_0, \rho} \left( \frac{1}{-\log s} \right) ds. \tag{2.19}$$

**Remark 2.2.** Explicit formulas for the Fisher information matrix  $I_{\theta_0}$  have been given by Prescott and Walden [21] (see also Beirlant et al. [2], page 169). In Appendix A, it is proved to be positive definite. The vector  $b$  given by the integral representation (2.19) can also be computed explicitly, and formulas are provided in Appendix A.

**Remark 2.3.** Equation (2.8) requires that both the number of blocks  $k$  and the block size  $m$  go to infinity with a relative rate measured by the second-order scaling function  $A$  and a parameter  $\lambda$ . When  $\lambda = 0$ , the bias term disappears in (2.18); this corresponds to the situation where  $m$  grows to infinity very quickly with respect to  $k$  so that the block size is large enough and the GEV approximation (2.4) is very good.

Existence and asymptotic normality of the MLE can be deduced from Theorem 2.1, mainly by the argmax theorem. The concavity property stated in Proposition 2.1 plays an important role in the proof of existence and uniqueness.

**Theorem 2.2.** Let  $X_1, X_2, \dots$  be independent and identically distributed random variables with cumulative distribution function  $F \in D(G_{\gamma_0})$  with  $\gamma_0 > -1/2$ , verifying the second-order condition (2.7). Let  $k = k_n \rightarrow \infty$  and  $m = m_n \rightarrow \infty$  be such that  $\sqrt{k}A(m) \rightarrow \lambda \in \mathbb{R}$ , with  $A$  the second-order auxiliary function.

(a) There exists a sequence of estimators  $\hat{\theta}_n = (\hat{\gamma}_n, \hat{\mu}_n, \hat{\sigma}_n)$ ,  $n \geq 1$ , such that

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\hat{\theta}_n \text{ is a MLE}] = 1 \tag{2.20}$$

and

$$\sqrt{k} \left( \hat{\gamma}_n - \gamma_0, \frac{\hat{\mu}_n - b_m}{a_m}, \frac{\hat{\sigma}_n}{a_m} - 1 \right) \xrightarrow{d} \mathcal{N}(\lambda I_{\theta_0}^{-1} b, I_{\theta_0}^{-1}). \tag{2.21}$$

(b) If  $\widehat{\theta}_n^i = (\widehat{\gamma}_n^i, \widehat{\mu}_n^i, \widehat{\sigma}_n^i)$ ,  $i = 1, 2$ , are two sequences of estimators satisfying

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\widehat{\theta}_n^i \text{ is a MLE}] = 1$$

and

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left[\sqrt{k}\left(\widehat{\gamma}_n^i - \gamma_0, \frac{\widehat{\mu}_n^i - b_m}{a_m}, \frac{\widehat{\sigma}_n^i}{a_m} - 1\right) \in H_n\right] = 1,$$

with  $H_n$  as in Proposition 2.1, then

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\widehat{\theta}_n^1 = \widehat{\theta}_n^2] = 1.$$

**Remark 2.4.** An interesting by-product of the strict concavity stated in Proposition 2.1 is the convergence of numerical procedures for the computation of the MLE that are implemented in software. The Newton–Raphson algorithm is commonly used to solve numerically the score equation (2.6). Strict concavity of the objective function on a large neighbourhood of the solution ensures convergence of the algorithm with probability tending to one as soon as the initial value  $\theta = (\gamma, \mu, \sigma)$  belongs to this neighbourhood.

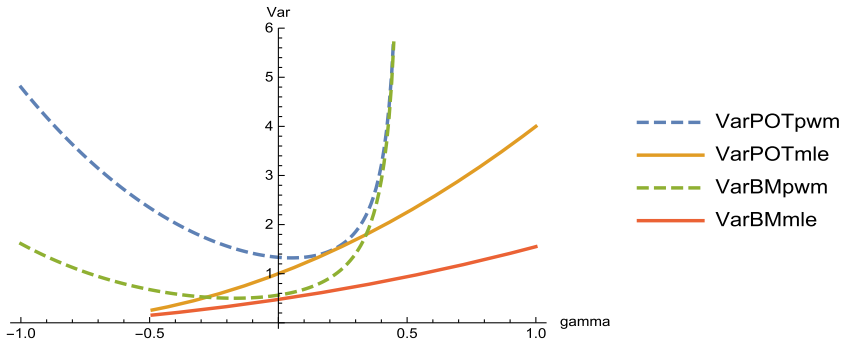
### 3. Theoretical comparisons: BM vs POT and MLE vs PWM

We compare in this section asymptotic properties of classical MLE and PWM estimators of the extreme value index, under the BM and POT methods. The methodological background is provided for instance, in Chapter 5 of the textbook by Beirlant et al. [2].

The POT method uses observations above some high threshold or top order statistic and the underlying approximate model is the Generalized Pareto distribution (Balkema and de Haan [1], Pickands [20]). Estimators of the shape parameter  $\gamma$ , as well as location and scale parameters have been proposed and widely studied, including MLE and PWM (Hosking and Wallis [16]). For their asymptotic properties – under basically the same conditions as under BM in Theorem 2.2 – we refer to de Haan and Ferreira [6]. Asymptotic normality of PWM estimators under BM has been established only recently by Ferreira and de Haan [14] and a comparison of PWM estimators under BM and POT has been carried out. The aim of the present section is to include our new asymptotic results for MLE estimators under BM, completing the picture in the comparison of the four different cases BM/POT and MLE/PWM. Only estimation of the extreme value index  $\gamma_0$  is considered here, but similar comparison could be performed for the location, scale or quantile estimators.

Recall that asymptotic normality of MLE (resp. PWM estimator) holds for  $\gamma > -1/2$  (resp.  $\gamma < 1/2$ ). The number  $k$  of selected observations corresponds to the number of blocks in BM and of selected top order statistics in POT. Similarly as in Ferreira and de Haan [14], our comparative study is restricted to the range  $\rho \in [-1, 0]$  where second-order conditions for BM and POT are comparable (cf. Drees et al. [10] or Ferreira and de Haan [14]). In the following we compare MLE/PWM under BM/POT methods through their: (i) asymptotic variances (VAR), (ii) asymptotic biases (BIAS), (iii) optimal asymptotic mean square errors (AMSE) and optimal number of observations minimizing AMSE ( $k_0$ ).





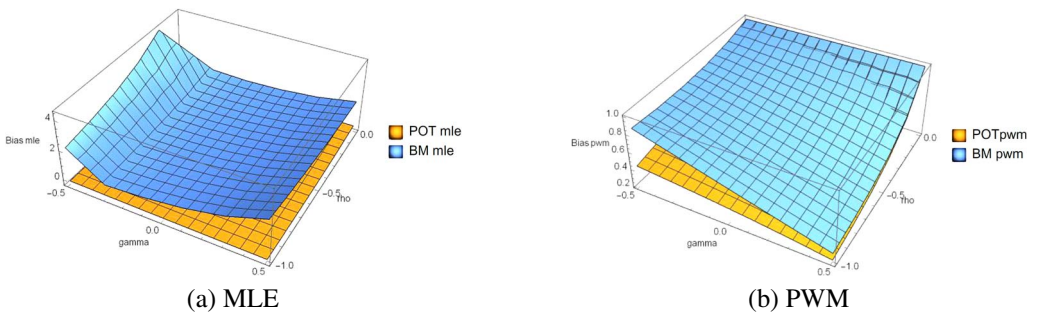
**Figure 1.** Asymptotic variances of estimators of the extreme value index  $\gamma$ . The straight lines corresponds to the MLE under BM and POT while the dashed lines correspond to PWM under BM and POT.

**(i) Asymptotic variances**

The asymptotic variance depends on  $\gamma$  only and is plotted in Figure 1 where straight lines stand for MLE and dashed lines for PWM estimators. Among all four different cases, BM-MLE is the one with the smallest variance within its range. Moreover, for both estimators, BM has the lowest variance indicating that BM is preferable to POT when variance is concerned.

**(ii) Asymptotic biases**

The asymptotic biases depend on  $\gamma$  and  $\rho$  and are shown in Figures 2–3: POT-MLE is the one with the smallest bias also in absolute value when compared to BM-MLE, contrary to what was observed for variance. This is in agreement with what has been observed when comparing BM-PWM and POT-PWM, also shown in Figures 2–3 already analysed in Ferreira and de Haan [14]. There is again the indication that the POT method is favourable to BM when concerning bias.



**Figure 2.** Asymptotic bias of estimators of the extreme value index  $\gamma$ : blue color for BM and orange for POT.

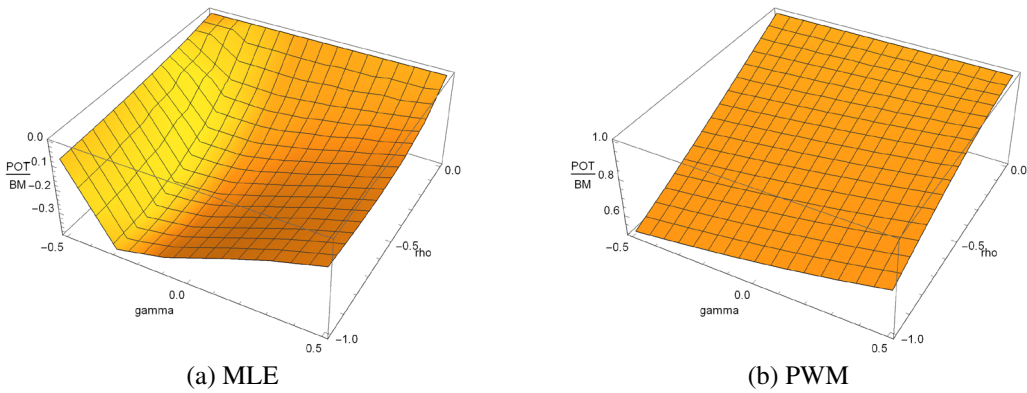


Figure 3. Ratios of asymptotic bias:  $BIAS_{POT} / BIAS_{BM}$ .

**(iii) Optimal asymptotic MSEs and optimal number of observations**

Another way to compare the estimators that combines both variance and bias information is through mean square error. One can compare these for the optimal number of observations  $k_0$  that is, that value for which the asymptotic mean square error (AMSE) is minimal. Similarly as in Ferreira and de Haan [14], we restrict our comparison to  $\rho < 0$ . Then, under the conditions of Theorem 2.2, we have

$$k_0 \sim \frac{n}{(\frac{1}{s})^{\leftarrow}(n)} \left( \frac{VAR(\gamma)}{BIAS^2(\gamma, \rho)} \right)^{1/(1-2\rho)}, \quad n \rightarrow \infty,$$

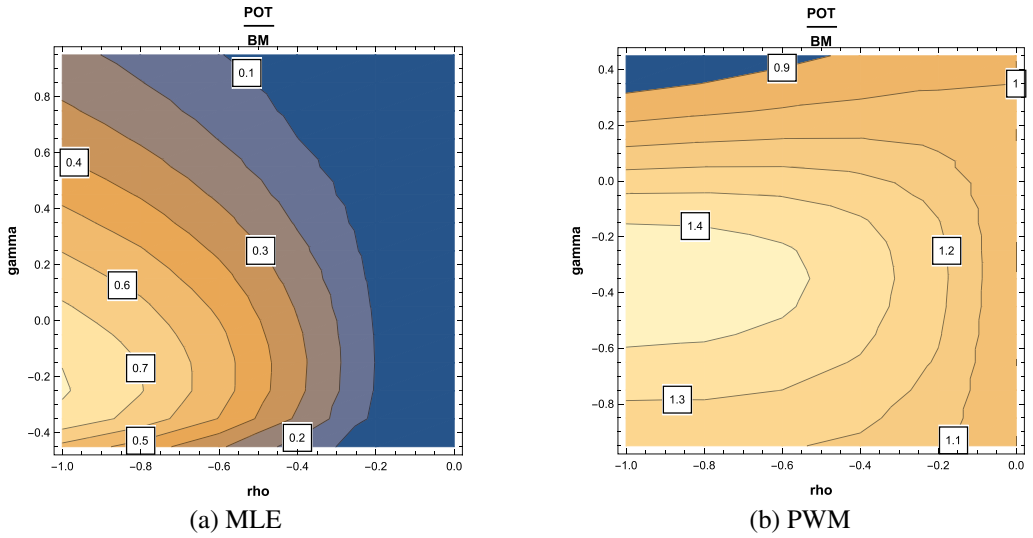
with  $s(\cdot)$  a decreasing and  $2\rho - 1$  regularly varying function such that  $A^2(t) = \int_t^\infty s(u) du$ . It follows in particular that the optimal  $k_0$  is different but of the same order for both estimators and methods. As for the optimal AMSE, we have

$$AMSE \sim \frac{1 - 2\rho}{-2\rho} \frac{(\frac{1}{s})^{\leftarrow}(n)}{n} (BIAS^2(\gamma, \rho))^{1/(1-2\rho)} (VAR(\gamma))^{-2\rho/(1-2\rho)}, \quad n \rightarrow \infty.$$

When considering ratios of optimal AMSE (or optimal number  $k_0$  of selected observations), the regularly varying function cancels out and the asymptotic ratio does not depend on  $n$  but only on  $\gamma$  and  $\rho$ .

Figure 4 shows the contour plots of the ratio  $AMSE_{POT} / AMSE_{BM}$  for MLE and PWM estimators. It is surprising to see a reverse behaviour in both cases: in the range of parameters considered, POT is preferable when MLE are considered, while BM is mostly preferable for PWM estimators.

In Figure 5 are shown  $(BIAS^2(\gamma, \rho))^{1/(1-2\rho)} (VAR(\gamma))^{-2\rho/(1-2\rho)}$  for comparing optimal AMSE among all combinations. The green surface corresponds to MLE-POT that has always the minimal optimal AMSE in the range of parameters considered. Finally, Figure 6 reports for MLE the asymptotic ratio of optimal numbers of selected observations, that is  $k_{0,POT} / k_{0,BM}$ . We

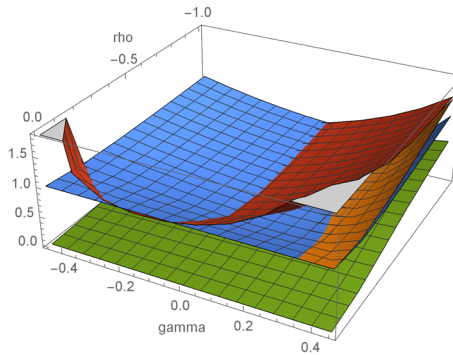


**Figure 4.** Contour plot of ratios of optimal AMSE:  $AMSE_{POT} / AMSE_{BM}$ .

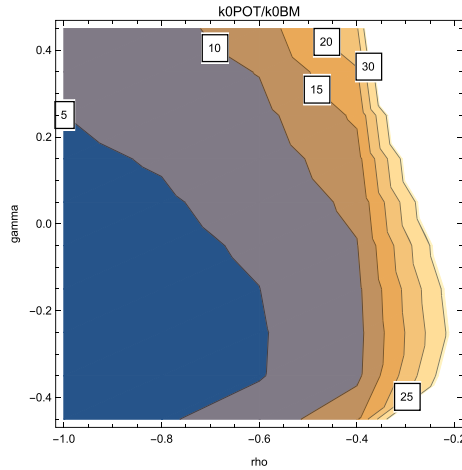
can see that the optimal number of observations is larger for POT, which is in agreement with the PWM case considered in previous studies.

### 4. Main proofs

We start by introducing some material that will be useful for the proofs. More technical material is still postponed to the Appendices.



**Figure 5.** Comparison of optimal AMSE: the lowest green surface corresponds to MLE-POT.



**Figure 6.** Ratio of asymptotic optimal sizes for the MLE:  $k_{0,POT}/k_{0,BM}$ .

### 4.1. Local asymptotic normality of the GEV model

If the observations  $(X_i)_{i \geq 1}$  are exactly  $GEV(\gamma_0, \mu_0, \sigma_0)$  distributed, then the choice of constants

$$a_m = \sigma_0 m^{\gamma_0} \quad \text{and} \quad b_m = \mu_0 + \sigma_0 \frac{m^{\gamma_0} - 1}{\gamma_0} \tag{4.1}$$

ensures that the normalised block maxima  $((M_{i,m} - b_m)/a_m)_{k \geq 1}$  are i.i.d. with distribution  $G_{\gamma_0}$ . The issue of model misspecification is irrelevant in that particular case.

In this simple i.i.d. setting, a key property in the theory of ML estimation is differentiability in quadratic mean (see, e.g., van der Vaart [24], Chapter 7). A statistical model defined by the family of densities  $\{p_\theta(x), \theta \in \Theta\}$  is called differentiable in quadratic mean at the point  $\theta_0$  if there exists a measurable function  $\dot{\ell}_{\theta_0}$  called the *score function* such that

$$\int_{\mathbb{R}} \left[ \sqrt{p_{\theta_0+h}(x)} - \sqrt{p_{\theta_0}(x)} - \frac{1}{2} h^T \dot{\ell}_{\theta_0}(x) \sqrt{p_{\theta_0}(x)} \right]^2 dx = o(\|h\|^2), \quad \text{as } h \rightarrow 0.$$

The following proposition corresponds to Proposition 3.2 in Bücher and Segers [3].

**Proposition 4.1.** *The three-parameter GEV model with log-likelihood  $\ell(\theta, x)$  defined in Equation (2.2) is differentiable in quadratic mean at  $\theta_0 = (\gamma_0, \sigma_0, \mu_0) \in \Theta$  if and only if  $\gamma_0 > -1/2$ . The score function is then given by  $\dot{\ell}_{\theta_0}(x) = \frac{\partial \ell}{\partial \theta}(\theta_0, x)$ .*

Differentiability in quadratic mean implies that the score function is centred with finite variance equal to the information matrix, i.e.

$$\int_{\mathbb{R}} \dot{\ell}_{\theta_0}(x) p_{\theta_0}(x) dx = 0 \quad \text{and} \quad \int_{\mathbb{R}} \dot{\ell}_{\theta_0}(x) \dot{\ell}_{\theta_0}(x)^T p_{\theta_0}(x) dx = I_{\theta_0}. \tag{4.2}$$

Another important consequence of differentiability in quadratic mean is the local asymptotic normality property of the local score process. The following corollary follows from Proposition 4.1 by a direct application of Theorem 7.2 in Van der Vaart [24].

**Corollary 4.1.** *Assume that  $F = \text{GEV}(\gamma_0, \mu_0, \sigma_0)$  with  $\gamma_0 > -1/2$  and that the constants  $a_m > 0, b_m \in \mathbb{R}$  are given by (4.1). Then the local log-likelihood process (2.10) satisfies*

$$\tilde{L}_{k,m}(h) = \tilde{L}_{k,m}(0) + h^T \tilde{G}_{k,m} - \frac{1}{2} h^T I_{\theta_0} h + o_P(1),$$

where the  $o_P(1)$  term is for a fixed  $h \in \mathbb{R}^3$ , and

$$\tilde{G}_{k,m} = \frac{1}{\sqrt{k}} \sum_{i=1}^k \frac{\partial \ell}{\partial \theta} \left( \theta_0, \frac{M_{i,m} - b_m}{a_m} \right) \xrightarrow{d} \mathcal{N}(0, I_{\theta_0}).$$

Note the similarity between Theorem 2.1 and Corollary 4.1. In Theorem 2.1 however, the  $o_P(1)$  is uniform on compact sets and the model misspecification  $F \in \mathcal{D}(G_{\gamma_0})$  results in a bias term  $\lambda b$  for the asymptotic distribution  $\tilde{G}_{k,m}$ .

### 4.2. The empirical quantile process associated to BM

The starting point of the proof of Proposition 2.1 and Theorem 2.1 is to rewrite the local log-likelihood process (2.10) in terms of the (normalized) empirical quantile process

$$Q_{k,m}(s) = \frac{M_{\lceil ks \rceil:k,m} - b_m}{a_m}, \quad 0 < s < 1, \tag{4.3}$$

where  $M_{1:k,m} \leq \dots \leq M_{k:k,m}$  are the order statistics of the block maxima sample  $(M_{k,m})_{1 \leq k \leq m}$  defined by (2.3) and  $\lceil x \rceil$  denotes the smallest integer larger than or equal to  $x$ . The local log-likelihood process (2.10) can be rewritten as

$$\tilde{L}_{k,m}(h) = k \int_0^1 \ell \left( \theta_0 + \frac{h}{\sqrt{k}}, Q_{k,m}(s) \right) ds - k \log(a_m). \tag{4.4}$$

Convergence (2.4) ensures the convergence of the empirical quantile process  $Q_{k,m}$  to the ‘‘true’’ quantile function  $Q_{\gamma_0}$  defined in (2.12). The following expansion of the empirical quantile process is taken from Ferreira and de Haan [14], Theorem 2.1.

**Proposition 4.2.** *Let  $X_1, X_2, \dots$  be independent and identically distributed random variables with cumulative distribution function  $F \in D(G_{\gamma_0})$  with  $\gamma_0 > -1/2$ , verifying (2.7) and, assume (2.8). For a specific choice of the second-order auxiliary functions  $a$  and  $A$ , as  $n \rightarrow \infty$ ,*

$$\sqrt{k}(Q_{k,m}(s) - Q_{\gamma_0}(s)) = \frac{B_k(s)}{s(-\log s)^{\gamma_0+1}} + \lambda H_{\gamma_0, \rho} \left( \frac{1}{-\log s} \right) + R_{k,m}(s), \tag{4.5}$$

where  $B_k, k \geq 1$ , denotes an appropriate sequence of standard Brownian bridges and the remainder term  $R_{k,m}$  satisfies, for  $0 < \varepsilon < 1/2$ ,

$$R_{k,m}(s) = s^{-1/2-\varepsilon} (1-s)^{-1/2-\gamma_0-\rho-\varepsilon} o_P(1) \tag{4.6}$$

uniformly for  $s \in [\frac{1}{k+1}, \frac{k}{k+1}]$ .

**Remark 4.1.** For Proposition 4.2, the auxiliary functions  $a$  and  $A$  have to be specially chosen for establishing uniform second-order regular variation bounds refining (2.7), see Lemma 4.2 in [14]. However, this choice is useful for the proofs only and is irrelevant for the statements of the main results in Section 2.2, similarly as in [14].

The following proposition provides useful technical bounds for the proof of the main results. Note that while  $O_P(1)$  refers to be bounded away from infinity,  $e^{O_P(1)}$  means bounded away from zero and infinity, in probability.

**Proposition 4.3.** *Under the conditions of Proposition 4.2, as  $n \rightarrow \infty$ ,*

$$(-\log s)^{\gamma_0} \left( 1 + (\gamma_0 + h_1/\sqrt{k}) \frac{Q_{k,m}(s) - h_2/\sqrt{k}}{1 + h_3/\sqrt{k}} \right) = e^{O_P(1)}$$

and

$$(-\log s)^{-1} \left( 1 + (\gamma_0 + h_1/\sqrt{k}) \frac{Q_{k,m}(s) - h_2/\sqrt{k}}{1 + h_3/\sqrt{k}} \right)^{-1/(\gamma_0+h_1/\sqrt{k})} = e^{O_P(1)}$$

uniformly for  $s \in [\frac{1}{k+1}, \frac{k}{k+1}]$  and  $h \in H_n$  as in Proposition 2.1.

For the proof of Proposition 4.3, we need the following lemma.

**Lemma 4.1.** *Let  $Z_{1:k} < \dots < Z_{k:k}$  be the order statistics of i.i.d. random variables  $Z_1, \dots, Z_k$  with standard Fréchet distribution. Then,*

$$\log\{(-\log s)Z_{[ks]:k}\} = O_P(1),$$

where the  $O_P(1)$ -term is uniform for  $s \in [\frac{1}{k+1}, \frac{k}{k+1}]$ .

**Proof of Lemma 4.1.** An equivalent statement is, with  $U$  standard uniform,

$$\log \left\{ \frac{-\log U_{[ks]:k}}{-\log s} \right\} = O_P(1).$$

We use Shorack and Wellner [22] (inequality 1 on page 419): for some  $M > 1$

$$\frac{1}{M} \leq \frac{U_{[ks]:k}}{s} \leq M \quad \text{for } \frac{1}{k+1} \leq s < 1, \tag{4.7}$$

$$\frac{1}{M} \leq \frac{1 - U_{[ks]:k}}{1 - s} \leq M \quad \text{for } 0 < s \leq \frac{k}{k+1}. \tag{4.8}$$

Relation (4.7) implies, for  $s \geq 1/(k+1)$ ,

$$1 - \frac{\log M}{-\log s} \leq \frac{-\log U_{[ks]:k}}{-\log s} \leq 1 + \frac{\log M}{-\log s}.$$

Both sides are bounded for  $0 < s \leq 1/2$ . Relation (4.8) implies, for  $s \leq k/(k+1)$ ,

$$\frac{1-s}{-\log s} \frac{1 - U_{[ks]:k}}{1-s} \leq \frac{-\log U_{[ks]:k}}{-\log s} \leq \frac{-\log\{1 - (1 - U_{[ks]:k})\}}{-\log s} \leq \frac{1 - U_{[ks]:k}}{1-s} \frac{1}{U_{[ks]:k}} \frac{1-s}{-\log s}.$$

Both sides are bounded for  $1/2 \leq s < 1$ . □

**Proof of Proposition 4.3.** Let  $Z$  be a unit Fréchet random variable, that is, with distribution function  $F(x) = e^{-1/x}$ ,  $x > 0$ , and  $\{Z_{i:k}\}_{i=1}^k$  be the order statistics from the associated i.i.d. sample of size  $k$ ,  $Z_1, \dots, Z_k$ . Note that  $M_{i:k,m} =^d V(mZ_{i:k})$ . From Lemma 4.2 in [14],

$$1 + \gamma \frac{V(mZ_{[ks]:k}) - b_m}{\sigma} = 1 + \gamma \frac{a_m^0}{\sigma} \frac{V(mZ_{[ks]:k}) - b_m}{a_m^0}$$

is bounded (above and below) by

$$Z_{[ks]:k}^{\gamma_0} + \left( \gamma \frac{a_m^0}{\sigma} - \gamma_0 \right) \frac{Z_{[ks]:k}^{\gamma_0} - 1}{\gamma_0} + \gamma \frac{a_m^0}{\sigma} A_0(m) H_{\gamma_0, \rho}(Z_{[ks]:k}) \pm \varepsilon \gamma \frac{a_m^0}{\sigma} Z_{[ks]:k}^{\gamma_0 + \rho \pm \delta} A_0(m)$$

for each  $\varepsilon, \delta > 0$  provided  $k$  and  $m$  are large enough. Hence,

$$(-\log s)^{\gamma_0} \left\{ 1 + \gamma \frac{a_m^0}{\sigma} \left( \frac{V(mZ_{[ks]:k}) - b_m}{a_m^0} - \frac{\mu - b_m}{a_m^0} \right) \right\}$$

is bounded (above and below) by,

$$\begin{aligned} & ((-\log s)Z_{\lceil ks \rceil:k})^{\gamma_0} + \left(\gamma \frac{a_m^0}{\sigma} - \gamma_0\right) \frac{((-\log s)Z_{\lceil ks \rceil:k})^{\gamma_0} - (-\log s)^{\gamma_0}}{\gamma_0} \\ & + \gamma \frac{a_m^0}{\sigma} A_0(m)(-\log s)^{\gamma_0} H_{\gamma_0,\rho}(Z_{\lceil ks \rceil:k}) - \gamma \frac{a_m^0}{\sigma} (-\log s)^{\gamma_0} \frac{\mu - b_m}{a_m^0} \\ & \pm \varepsilon \gamma \frac{a_m^0}{\sigma} ((-\log s)Z_{\lceil ks \rceil:k})^{\gamma_0} Z_{\lceil ks \rceil:k}^{\rho \pm \delta} A_0(m). \end{aligned} \tag{4.9}$$

Applying Lemma 4.1, the first term in (4.9) is bounded (above and below) uniformly in  $s$ , in probability.

It remains to verify that the other terms are  $o_p(1)$  uniformly in  $s$ . Note that

$$\sup_{(k+1)^{-1} \leq s \leq k(k+1)^{-1}} \frac{(-\log s)^{\gamma_0}}{k^{1/2-\delta}} = \begin{cases} O\left(\frac{(\log k)^{\gamma_0}}{k^{1/2-\delta}}\right), & \gamma_0 > 0, \\ O(k^{-1/2-\gamma_0+\delta}), & \gamma_0 \leq 0. \end{cases}$$

Using this with  $\delta < \min(1/2, \gamma_0 + 1/2)$ , Lemma 4.1 and (2.13), the second term in (4.9) is  $o_p(1)$  uniformly in  $s$ . For the third term, by

$$A_0(m)(-\log s)^{\gamma_0} H_{\gamma_0,\rho}(Z_{\lceil ks \rceil,k}) = \sqrt{k} A_0(m) \frac{(-\log s)^{\gamma_0}}{\sqrt{k}} H_{\gamma_0,\rho}(Z_{\lceil ks \rceil,k})$$

and,

$$\sup_{(k+1)^{-1} \leq s \leq k(k+1)^{-1}} \frac{(-\log s)^{\gamma_0}}{\sqrt{k}} H_{\gamma_0,\rho}\left(\frac{1}{-\log s}\right) = O((\log k)^\xi k^{-1/2-(\gamma_0 \vee 0)})$$

for some  $\xi \in \mathbb{R}$ , it follows that it is also  $o_p(1)$  uniformly in  $s$ . The last two terms follow similarly.

For the second statement just note that

$$(-\log s)^{-\gamma_0/(\gamma_0+h_1/\sqrt{k})} = (-\log s)^{-1}(-\log s)^{h_1/(\gamma_0\sqrt{k})(1+o(1))},$$

where the second factor converges to 1 uniformly in  $s$ . □

The following auxiliary result closely related to Proposition 4.3 will be useful in our proofs of Proposition 2.1 and Theorem 2.1.

**Lemma 4.2.** *Under the conditions of Proposition 4.2, as  $n \rightarrow \infty$ ,*

$$(-\log s)^{-1} \left(1 + (\gamma_0 + h_1/\sqrt{k}) \frac{Q_{\gamma_0}(s) - h_2/\sqrt{k}}{1 + h_3/\sqrt{k}}\right)^{-1/(\gamma_0+h_1/\sqrt{k})} = 1 + o(1),$$

uniformly for  $s \in [\frac{1}{k+1}, \frac{k}{k+1}]$  and  $h \in H_n$  as in Proposition 2.1.



**Proof.** Using (2.9) and expanding,

$$\begin{aligned}
 & 1 + (\gamma_0 + h_1/\sqrt{k}) \frac{Q_{\gamma_0}(s) - h_2/\sqrt{k}}{1 + h_3/\sqrt{k}} \\
 &= (-\log s)^{-\gamma_0} + O\left(\max(h_1, h_2, h_3) \frac{1 + (-\log s)^{\gamma_0}}{\sqrt{k}}\right),
 \end{aligned}$$

where the O-term is in fact a o-term uniformly in  $s$  as seen in the proof of Proposition 4.3. The result follows as in the last part of the proof of Proposition 4.3 for the second statement.  $\square$

### 4.3. Proofs of Proposition 2.1, Theorems 2.1 and 2.2

Before proving Proposition 2.1, we first check that, with probability tending to one the local log-likelihood process  $\tilde{L}_{k,m}$  is finite and twice differentiable on  $H_n$ .

**Lemma 4.3.** *Under the assumptions of Proposition 2.1, we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}[\tilde{L}_{k,m}(h) > -\infty \text{ for all } h \in H_n] = 1. \tag{4.10}$$

Furthermore,  $\tilde{L}_{k,m}$  is smooth on  $H_n$  as soon as it is finite on  $H_n$ .

**Proof.** In view of Equations (2.1), (2.2) and (2.10),  $\tilde{L}_{k,m}(h)$  is finite on  $H_n$  as soon as

$$1 + (\gamma_0 + h_1/\sqrt{k}) \frac{Q_{k,m}(1/(k+1)) - h_2/\sqrt{k}}{1 + h_3/\sqrt{k}} > 0 \quad \text{for all } h = (h_1, h_2, h_3) \in H_n.$$

Proposition 4.3 entails that the left hand side is asymptotically  $(\log k)^{-\gamma_0} e^{O_P(1)}$  uniformly on  $H_n$  so that it remains positive on  $H_n$  with probability tending to one. Equations (2.1)–(2.2) imply that the function  $\theta \mapsto \ell(\theta, x)$  is smooth when it is finite. We deduce that  $\tilde{L}_{k,m}$  is smooth on  $H_n$  as soon as it is finite on  $H_n$ .  $\square$

**Proof of Proposition 2.1.** According to Lemma 4.3, the local log-likelihood process  $\tilde{L}_{k,m}$  is smooth on  $H_n$  with probability tending to one. Differentiating Equation (4.4), we get

$$\frac{\partial^2 \tilde{L}_{k,m}}{\partial h \partial h^T}(h) = \int_0^1 \frac{\partial^2 \ell}{\partial \theta \partial \theta^T} \left( \theta_0 + \frac{h}{\sqrt{k}}, Q_{k,m}(s) \right) ds.$$

By the definition (2.15) of the information matrix,

$$\begin{aligned}
 \frac{\partial^2 \tilde{L}_{k,m}}{\partial h \partial h^T}(h) + I_{\theta_0} &= \int_0^1 \left( \frac{\partial^2 \ell}{\partial \theta \partial \theta^T} \left( \theta_0 + \frac{h}{\sqrt{k}}, Q_{k,m}(s) \right) - \frac{\partial^2 \ell}{\partial \theta \partial \theta^T}(\theta_0, Q_{\gamma_0}(s)) \right) ds \\
 &= \int_0^{\frac{1}{k}} (\dots) ds + \int_{\frac{1}{k}}^{\frac{k-1}{k}} (\dots) ds + \int_{\frac{k-1}{k}}^1 (\dots) ds \\
 &= \text{I} + \text{II} + \text{III}.
 \end{aligned}$$

We will show that these three terms are  $o_P(1)$  uniformly on  $H_n$ , which proves Equation (2.14).

First consider the boundary terms I and III. Since  $Q_{k,m}(s)$  is constant on  $[0, \frac{1}{k}]$ , we have

$$I = \frac{1}{k} \frac{\partial^2 \ell}{\partial \theta \partial \theta^T} \left( \theta_0 + \frac{h}{\sqrt{k}}, Q_{k,m} \left( \frac{1}{k} \right) \right) - \int_0^{\frac{1}{k}} \frac{\partial^2 \ell}{\partial \theta \partial \theta^T} (\theta_0, Q_{\gamma_0}(s)) ds. \quad (4.11)$$

The integral term vanishes as  $k \rightarrow \infty$  because the integral is well defined on  $[0, 1]$  (see Eq. (2.15) or alternatively use the upper bound for the second derivative provided by Proposition B.1). To deal with the first term, we use the upper bound for the second derivative provided by Proposition B.1:

$$\left\| \frac{\partial^2 \ell}{\partial \theta \partial \theta^T} \left( \theta_0 + \frac{h}{\sqrt{k}}, Q_{k,m} \left( \frac{1}{k} \right) \right) \right\| \leq C \max(z^{-\varepsilon}, z^{1+\varepsilon}, z^{2\gamma_0-\varepsilon}, z^{1+2\gamma_0+\varepsilon})$$

with

$$\begin{aligned} z &= z \left( \theta_0 + \frac{h}{\sqrt{k}}, Q_{k,m} \left( \frac{1}{k} \right) \right) \\ &= \left( 1 + (\gamma_0 + h_1/\sqrt{k}) \frac{Q_{k,m}(1/k) - h_2/\sqrt{k}}{1 + h_3/\sqrt{k}} \right)^{-1/(\gamma_0 + h_1/\sqrt{k})}. \end{aligned}$$

Proposition 4.3 with  $s = \frac{1}{k}$  provides the asymptotic  $z = e^{O_P(1)} \log k$  uniformly for  $h \in H_n$ . We deduce that the first term in (4.11) is asymptotically

$$e^{O_P(1)} \frac{1}{k} \max((\log k)^{-\varepsilon}, (\log k)^{1+\varepsilon}, (\log k)^{2\gamma_0-\varepsilon}, (\log k)^{1+2\gamma_0+\varepsilon}) = o_P(1)$$

uniformly for  $h \in H_n$ . Hence,  $I = o_P(1)$  uniformly for  $h \in H_n$ . The proof for the boundary term III is similar and details are omitted.

Next, consider the main term II. By Taylor formula, we have

$$\begin{aligned} II &= \int_{\frac{1}{k}}^{\frac{k-1}{k}} \left( \frac{\partial^2 \ell}{\partial \theta \partial \theta^T} \left( \theta_0 + \frac{h}{\sqrt{k}}, Q_{k,m}(s) \right) - \frac{\partial^2 \ell}{\partial \theta \partial \theta^T} (\theta_0, Q_{\gamma_0}(s)) \right) ds \\ &= II_a + II_b \end{aligned}$$

with

$$II_a = \frac{1}{\sqrt{k}} \int_{\frac{1}{k}}^{\frac{k-1}{k}} \int_0^1 h^T \frac{\partial^3 \ell}{\partial^2 \theta \partial \theta^T} \left( \theta_0 + \frac{uh}{\sqrt{k}}, (1-u)Q_{\gamma_0}(s) + uQ_{k,m}(s) \right) du ds, \quad (4.12)$$

$$\begin{aligned} II_b &= \int_{\frac{1}{k}}^{\frac{k-1}{k}} \int_0^1 (Q_{k,m}(s) - Q_{\gamma_0}(s)) \\ &\quad \times \frac{\partial^3 \ell}{\partial x \partial \theta \partial \theta^T} \left( \theta_0 + \frac{uh}{\sqrt{k}}, (1-u)Q_{\gamma_0}(s) + uQ_{k,m}(s) \right) du ds. \end{aligned} \quad (4.13)$$

Using the notation  $z = z(\theta'_n, x'_n)$  with

$$x'_n = (1 - u)Q_{\gamma_0}(s) + uQ_{k,m}(s) \quad \text{and} \quad \theta'_n = \theta_0 + \frac{uh}{\sqrt{k}}, \quad u \in [0, 1],$$

Proposition B.1 provides the upper bound

$$\left\| \frac{\partial^3 \ell}{\partial^2 \theta \partial \theta^T}(\theta'_n, x'_n) \right\| \leq C \max(z^{-\varepsilon}, z^{1+\varepsilon}, z^{3\gamma_0-\varepsilon}, z^{1+3\gamma_0+\varepsilon}).$$

Using the fact that  $uh \in H_n$  and that  $z(\theta'_n, x'_n)$  is between  $z(\theta'_n, Q_{\gamma_0}(s))$  and  $z(\theta'_n, Q_{k,m}(s))$ , Proposition 4.3 and Lemma 4.2 imply

$$z = z(\theta'_n, x'_n) = e^{O_P(1)}(-\log s)$$

uniformly for  $s \in [\frac{1}{k}, \frac{k-1}{k}]$ ,  $u \in [0, 1]$  and  $h \in H_n$ . Using these bounds, we obtain

$$\left\| \frac{\partial^3 \ell}{\partial^2 \theta \partial \theta^T}(\theta'_n, x'_n) \right\| = e^{O_P(1)} \max(s^{-\varepsilon}, (1-s)^{-\varepsilon}, (1-s)^{3\gamma_0-\varepsilon})$$

and, since condition (2.13) implies  $\|h\| = O(k^\delta)$ ,

$$\Pi_a = O_P(k^{\delta-1/2}) \int_{\frac{1}{k}}^{\frac{k-1}{k}} \max(s^{-\varepsilon}, (1-s)^{-\varepsilon}, (1-s)^{3\gamma_0-\varepsilon}) ds.$$

When  $3\gamma_0 - \varepsilon > -1$  the integral converges as  $k \rightarrow \infty$  and,  $\Pi_a = O_P(k^{\delta-1/2}) = o_P(1)$  since  $\delta < 1/2$ . When  $\gamma_0 \leq -1/3$ , the integral diverges at rate  $O(k^{-1-3\gamma_0+\varepsilon})$  so that  $\Pi_a = O_P(k^{\delta-3/2-3\gamma_0+\varepsilon}) = o_P(1)$  since  $\delta + \varepsilon < \gamma_0 + 1/2$ .

Similarly for the term  $\Pi_b$ , Propositions 4.3 and B.1 together with Lemma 4.2 imply

$$\begin{aligned} & \left\| \frac{\partial^3 \ell}{\partial x \partial \theta \partial \theta^T}(\theta'_n, x'_n) \right\| \\ & \leq C \max(z^{\gamma_0-\varepsilon}, z^{1+\gamma_0+\varepsilon}, z^{3\gamma_0-\varepsilon}, z^{1+3\gamma_0+\varepsilon}) \\ & \leq e^{O_P(1)} \max((-\log s)^{\gamma_0-\varepsilon}, (-\log s)^{1+\gamma_0+\varepsilon}, (-\log s)^{3\gamma_0-\varepsilon}, (-\log s)^{1+3\gamma_0+\varepsilon}), \end{aligned}$$

uniformly for  $s \in [\frac{1}{k}, \frac{k-1}{k}]$ ,  $u \in [0, 1]$  and  $h \in H_n$ .

From the law of the iterated logarithm,

$$B_k(s) = O_p(s^{1/2-\varepsilon}(1-s)^{1/2-\varepsilon}) \quad \text{uniformly on } (0, 1) \tag{4.14}$$

and,

$$H_{\gamma_0, \rho} \left( \frac{1}{-\log s} \right) = O(s^{-\varepsilon}(1-s)^{-\varepsilon+\min(-\gamma_0, 0)}) \quad \text{uniformly on } (0, 1). \tag{4.15}$$

Combining these two with Proposition 4.2 it follows,

$$\sqrt{k}(Q_{k,m}(s) - Q_{\gamma_0}(s)) = O_P(s^{-1/2-\varepsilon}(1-s)^{-1/2-\gamma_0-\varepsilon} + s^{-\varepsilon}(1-s)^{\min(-\gamma_0,0)-\varepsilon}) \quad (4.16)$$

uniformly on  $(0, 1)$ .

Combining the previous bound for the derivative and (4.16), we deduce similarly as before

$$\Pi_b = O_P\left(\frac{1}{\sqrt{k}}\right) \int_{\frac{1}{k}}^{\frac{k-1}{k}} \max(s^{-1/2-2\varepsilon}, (1-s)^{-1/2+2\gamma_0-2\varepsilon}, (1-s)^{3\gamma_0-2\varepsilon}) ds = o_P(1)$$

uniformly in  $h \in H_n$  for  $\varepsilon$  small enough. □

**Proof of Theorem 2.1.** Integrating Equation (2.14), we obtain directly

$$\begin{aligned} \frac{\partial \tilde{L}_{k,m}}{\partial h}(h) &= \frac{\partial \tilde{L}_{k,m}}{\partial h}(0) - I_{\theta_0} h + o_P(1), \\ \tilde{L}_{k,m}(h) &= \tilde{L}_{k,m}(0) + h^T \frac{\partial \tilde{L}_{k,m}}{\partial h}(0) - \frac{1}{2} h^T I_{\theta_0} h + o_P(1) \end{aligned}$$

uniformly on compact sets. This is exactly Equations (2.16) and (2.17) since

$$\frac{\partial \tilde{L}_{k,m}}{\partial h}(0) = \frac{1}{\sqrt{k}} \sum_{i=1}^k \frac{\partial \ell}{\partial \theta} \left( \theta_0, \frac{M_{i,m} - b_m}{a_m} \right) = \tilde{G}_{k,m}.$$

It remains to prove the asymptotic normality (2.18), that is,

$$\frac{\partial \tilde{L}_{k,m}}{\partial h}(0) \rightarrow^d \mathcal{N}(\lambda b, I_{\theta_0}). \quad (4.17)$$

Differentiating Equation (4.4), we obtain

$$\frac{\partial \tilde{L}_{k,m}}{\partial h}(0) = \sqrt{k} \int_0^1 \frac{\partial \ell}{\partial \theta}(\theta_0, Q_{k,m}(s)) ds = \text{I}' + \text{II}' + \text{III}'$$

where the three terms correspond to the integrals on  $[0, \frac{1}{k}]$ ,  $[\frac{1}{k}, \frac{k-1}{k}]$  and  $[\frac{k-1}{k}, 1]$ , respectively. Since  $Q_{k,m}(s)$  is constant on the first and last intervals, we have

$$\begin{aligned} \text{I}' &= \frac{1}{\sqrt{k}} \frac{\partial \ell}{\partial \theta} \left( \theta_0, Q_{k,m} \left( \frac{1}{k+1} \right) \right), \\ \text{III}' &= \frac{1}{\sqrt{k}} \frac{\partial \ell}{\partial \theta} \left( \theta_0, Q_{k,m} \left( \frac{k}{k+1} \right) \right). \end{aligned}$$

The first term is evaluated thanks to Propositions B.1 and 4.3:

$$\left\| \frac{\partial \ell}{\partial \theta} \left( \theta_0, Q_{k,m} \left( \frac{1}{k+1} \right) \right) \right\| \leq C \max(z^{-\varepsilon}, z^{1+\varepsilon}, z^{\gamma_0-\varepsilon}, z^{1+\gamma_0+\varepsilon})$$

with

$$z = \left(1 + \gamma_0 Q_{k,m} \left(\frac{1}{k+1}\right)\right)^{-1/\gamma_0} = e^{O_P(1)} \log k,$$

whence we deduce

$$I' = e^{O_P(1)} \frac{1}{\sqrt{k}} \max((\log k)^{1+\varepsilon}, (\log k)^{1+\gamma_0+\varepsilon}) = o_P(1).$$

With similar arguments, one can prove  $III' = o_P(1)$ .

For the second term, we use Taylor integral formula

$$\begin{aligned} & \frac{\partial \ell}{\partial \theta}(\theta_0, Q_{k,m}(s)) \\ &= \frac{\partial \ell}{\partial \theta}(\theta_0, Q_{\gamma_0}(s)) + \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, Q_{\gamma_0}(s))(Q_{k,m}(s) - Q_{\gamma_0}(s)) \\ & \quad + (Q_{k,m}(s) - Q_{\gamma_0}(s))^2 \int_0^1 (1-u) \frac{\partial^3 \ell}{\partial x^2 \partial \theta}(\theta_0, (1-u)Q_{\gamma_0}(s) + uQ_{k,m}(s)) du. \end{aligned}$$

From decomposition (4.5) for  $\sqrt{k}(Q_{k,m}(s) - Q_{\gamma_0}(s))$ , we get

$$\Pi' = \sqrt{k} \int_{\frac{1}{k}}^{\frac{k-1}{k}} \frac{\partial \ell}{\partial \theta}(\theta_0, Q_{k,m}(s)) ds = \Pi'_a + \Pi'_b + \Pi'_c + \Pi'_d + \Pi'_e,$$

with

$$\Pi'_a = \sqrt{k} \int_{\frac{1}{k}}^{\frac{k-1}{k}} \frac{\partial \ell}{\partial \theta}(\theta_0, Q_{\gamma_0}(s)) ds,$$

$$\Pi'_b = \int_{\frac{1}{k}}^{\frac{k-1}{k}} \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, Q_{\gamma_0}(s)) \frac{B_k(s)}{s(-\log s)^{\gamma_0+1}} ds,$$

$$\Pi'_c = \lambda \int_{\frac{1}{k}}^{\frac{k-1}{k}} \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, Q_{\gamma_0}(s)) H_{\gamma_0, \rho} \left(\frac{1}{-\log s}\right) ds,$$

$$\Pi'_d = \int_{\frac{1}{k}}^{\frac{k-1}{k}} \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, Q_{\gamma_0}(s)) R_{k,m}(s) ds,$$

$$\Pi'_e = \int_{\frac{1}{k}}^{\frac{k-1}{k}} \int_0^1 \sqrt{k}(Q_{k,m}(s) - Q_{\gamma_0}(s))^2 (1-u) \frac{\partial^3 \ell}{\partial x^2 \partial \theta}(\theta_0, (1-u)Q_{\gamma_0}(s) + uQ_{k,m}(s)) du ds.$$

We consider the different terms successively. Equation (4.2) implies

$$\int_0^1 \frac{\partial \ell}{\partial \theta}(\theta_0, Q_{\gamma_0}(s)) ds = 0$$

so that

$$\Pi'_a = -\sqrt{k} \int_0^{\frac{1}{k}} \frac{\partial \ell}{\partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) ds - \sqrt{k} \int_{1-\frac{1}{k}}^1 \frac{\partial \ell}{\partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) ds.$$

Proposition B.1 provides the upper bound

$$\begin{aligned} \left\| \frac{\partial \ell}{\partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) \right\| &\leq C \max((-\log s)^{-\varepsilon}, (-\log s)^{1+\varepsilon}, (-\log s)^{\gamma_0-\varepsilon}, (-\log s)^{1+\gamma_0+\varepsilon}) \\ &\leq C \max(s^{-\varepsilon}, (1-s)^{-\varepsilon}, (1-s)^{\gamma_0-\varepsilon}), \end{aligned}$$

whence we deduce  $\Pi'_a = O(\max(k^{-1/2+\varepsilon}, k^{-1/2-\gamma_0+\varepsilon})) = o(1)$  because  $\gamma_0 > -1/2$ . For term  $\Pi'_b$ , Proposition B.1 entails

$$\left\| \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) \right\| \tag{4.18}$$

$$\begin{aligned} &\leq C \max((-\log s)^{\gamma_0-\varepsilon}, (-\log s)^{1+\gamma_0+\varepsilon}, (-\log s)^{2\gamma_0-\varepsilon}, (-\log s)^{1+2\gamma_0+\varepsilon}) \\ &\leq C \max(s^{-\varepsilon}, (1-s)^{\gamma_0-\varepsilon}, (1-s)^{2\gamma_0-\varepsilon}). \end{aligned} \tag{4.19}$$

Combined with (4.14), we get

$$\left\| \frac{B_k(s)}{s(-\log s)^{\gamma_0+1}} \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) \right\| = \max(s^{-1/2-2\varepsilon}, (1-s)^{-1/2-2\varepsilon}, (1-s)^{\gamma_0-1/2-2\varepsilon}) O_P(1),$$

which implies, since  $\gamma_0 > -1/2$ ,

$$\Pi'_b = \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) \frac{B_k(s)}{s(-\log s)^{\gamma_0+1}} ds + o_P(1),$$

where the integral on  $[0, 1]$  is well defined.

Similarly for  $\Pi'_c$ , (4.15) together with (4.19) yields

$$\left\| H_{\gamma_0, \rho} \left( \frac{1}{-\log s} \right) \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) \right\| \leq C \max(s^{-2\varepsilon}, (1-s)^{-2\varepsilon}, (1-s)^{2\gamma_0-2\varepsilon}).$$

Because  $2\gamma_0 > -1$ , we get

$$\Pi'_c = \lambda \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) H_\rho \left( \frac{1}{-\log s} \right) ds + o(1),$$

where the integral on  $[0, 1]$  is well defined.

For  $\Pi'_d$  we use the uniform bound (4.6) and the upper bound (4.19) to get

$$\Pi'_d = \int_{\frac{1}{k}}^{\frac{k-1}{k}} \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) R_{k,m}(s) ds$$

$$\begin{aligned}
 &= o_P(1) \int_{\frac{1}{k}}^{\frac{k-1}{k}} C \max(s^{-1/2-2\varepsilon}, (1-s)^{-1/2-\rho-2\varepsilon}, (1-s)^{\gamma_0-1/2-\rho-2\varepsilon}) ds \\
 &= o_P(1).
 \end{aligned}$$

We consider finally the last term  $\Pi'_e$ . With the notations  $x'_n = (1-u)Q_{\gamma_0}(s) + uQ_{k,m}(s)$  and  $z = z(\theta_0, x'_n)$ , Proposition B.1 yields

$$\left\| \frac{\partial^3 \ell}{\partial^2 x \partial \theta}(\theta_0, x'_n) \right\| \leq C \max(z^{2\gamma_0-\varepsilon}, z^{1+2\gamma_0+\varepsilon}, z^{3\gamma_0-\varepsilon}, z^{1+3\gamma_0+\varepsilon}).$$

Using the fact that  $z(\theta_0, x'_n)$  is between  $z(\theta_0, Q_{\gamma_0}(s))$  and  $z(\theta_0, Q_{k,m}(s))$ , Corollary B.1 implies  $z(\theta_0, x'_n) = e^{O_P(1)}(-\log s)$  so that

$$\left\| \frac{\partial^3 \ell}{\partial^2 x \partial \theta}(\theta_0, x'_n) \right\| = O_P(1) \max(s^{-\varepsilon}, (1-s)^{2\gamma_0-\varepsilon}, (1-s)^{3\gamma_0-\varepsilon}).$$

Combining this bound with (4.16), we obtain similarly as before

$$\Pi'_e = O\left(\frac{1}{\sqrt{k}}\right) \int_{\frac{1}{k}}^{\frac{k-1}{k}} \max(s^{-1-3\varepsilon}, (1-s)^{-1+\gamma_0-3\varepsilon}) ds = o_P(1).$$

Collecting all the different terms, we get

$$\begin{aligned}
 \frac{\partial \tilde{L}_{k,m}}{\partial h}(0) &= \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, Q_{\gamma_0}(s)) \frac{B_k(s)}{s(-\log s)^{\gamma_0+1}} ds \\
 &\quad + \lambda \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \theta}(\theta_0, Q_{\gamma_0}(s)) H_\rho\left(\frac{1}{-\log s}\right) ds + o_P(1).
 \end{aligned}$$

The second term in the right-hand side is deterministic and corresponds to  $\lambda b$  with  $b$  defined by (2.19). The first term is an integral of the Brownian bridge that defines a centred Gaussian vector whose covariance only depends on the first-order parameter  $\gamma_0$ . Comparing with the special case of i.i.d. GEV random variables considered in Corollary 4.1, we identify the covariance which is equal to  $I_{\gamma_0}$ . This proves Equation (4.17) and concludes the proof of Theorem 2.1.  $\square$

**Proof of Theorem 2.2.** The proof of Theorem 2.2 relies on Theorem 2.1 and on the Argmax theorem. Consider the random processes

$$M_n(h) = \tilde{L}_{k,m}(h) - \tilde{L}_{k,m}(0), \quad h \in \mathbb{R}^3$$

and

$$M(h) = h^T(\lambda b + G) - \frac{1}{2} h^T I_{\theta_0} h, \quad h \in \mathbb{R}^3$$

with  $G$  a centred Gaussian random vector with covariance  $I_{\theta_0}$ . Let  $H_n$  be the closed ball of  $\mathbb{R}^3$  centred at 0 and with radius  $r_n \rightarrow \infty$  such that  $r_n = O(k^\delta)$  as in (2.13). Define the maximizer

$$\widehat{h}_n = \operatorname{argmax}_{h \in H_n} M_n(h).$$

In the case it is not unique, define  $\widehat{h}_n$  as the smallest maximizer in the lexicographic order. Theorem 2.1 implies that, for any compact  $K \subset \mathbb{R}^3$ ,  $M_n$  converge in distribution to  $M$  in  $\ell^\infty(K)$  as  $k \rightarrow \infty$ . The limit process  $M$  is continuous and has a unique maximizer given by

$$\widehat{h} = \operatorname{argmax}_{h \in \mathbb{R}^3} M(h) = I_{\theta_0}^{-1}(\lambda b + G).$$

The Argmax theorem (see van der Vaart [24], Corollary 5.58) implies that, provided the sequence  $\widehat{h}_n$  is tight, then  $\widehat{h}_n$  converge weakly to  $\widehat{h}$  as  $n \rightarrow \infty$ .

We now prove the tightness of the sequence  $\widehat{h}_n$ . Let  $\varepsilon > 0$ . There is  $R > 0$  such that

$$\mathbb{P}(\|\widehat{h}\| < R) > 1 - \varepsilon.$$

The relation

$$M(h) = M(\widehat{h}) - \frac{1}{2}(h - \widehat{h})^T I_{\theta_0}(h - \widehat{h}),$$

implies that

$$\max_{\|\widehat{h}-h\| \geq 1} M(h) = M(\widehat{h}) - \frac{1}{2}\lambda_{\min}$$

with  $\lambda_{\min} > 0$  the smallest eigenvalue of  $I_{\theta_0}$ . As a consequence,  $\|\widehat{h}\| < R$  implies

$$M(\widehat{h}) - \max_{\|h\|=R+1} M(h) = \max_{\|h\| \leq R} M(h) - \max_{\|h\|=R+1} M(h) \geq \frac{1}{2}\lambda_{\min}$$

and this occurs with probability at least  $1 - \varepsilon$ . Using the convergence in distribution of  $M_n$  to  $M$  in  $\ell^\infty(K)$  with  $K = \{h : \|h\| \leq R + 1\}$ , we deduce that for large  $n$

$$\max_{\|h\| \leq R} M_n(h) - \max_{\|h\|=R+1} M_n(h) \geq \frac{1}{4}\lambda_{\min} \tag{4.20}$$

with probability at least  $1 - 2\varepsilon$ . For large  $n$ ,  $H_n$  contains the ball  $\{h : \|h\| \leq R + 1\}$  (because  $r_n \rightarrow \infty$ ) and, according to Proposition 2.1,  $M_n$  is strictly concave on  $H_n$  with probability at least  $1 - \varepsilon$ . Then, Equation (4.20) together with the strict concavity of  $M_n$  implies that the maximizer  $\widehat{h}_n$  of  $M_n$  over  $H_n$  satisfies  $\|\widehat{h}_n\| \leq R$ . Hence, for large  $n$ ,  $\mathbb{P}(\|\widehat{h}_n\| \leq R) \geq 1 - 3\varepsilon$  and this proves the tightness of  $\widehat{h}_n$ . Note also that on this event,  $\widehat{h}_n$  belongs to the interior of  $H_n$  and is hence a critical point of  $M_n$ , i.e.

$$\frac{\partial M_k}{\partial h}(\widehat{h}_n) = 0 \quad \text{or equivalently} \quad \frac{\partial \widetilde{L}_{k,m}}{\partial h}(\widehat{h}_n) = 0.$$



Define

$$\widehat{\theta}_n = (\gamma_0 + k^{-1/2}\widehat{h}_{n,1}, b_m + a_mk^{-1/2}\widehat{h}_{n,2}, a_m(1 + k^{-1/2}\widehat{h}_{n,3})).$$

Equations (2.10) and (2.11) imply that

$$\widetilde{L}_{k,m}(\widehat{h}_n) = L_{k,m}(\widehat{\theta}_n) \quad \text{and} \quad \frac{\partial \widetilde{L}_{k,m}}{\partial h}(\widehat{h}_n) = k^{-1/2} \frac{\partial L_{k,m}}{\partial \theta}(\widehat{\theta}_n).$$

Hence, with probability tending to one,  $L_{k,m}$  has a local maximum at  $\widehat{\theta}_n$  with  $\frac{\partial L_{k,m}}{\partial \theta}(\widehat{\theta}_n) = 0$ , that is,  $\widehat{\theta}_n$  is a MLE and Eq. (2.20) is satisfied. Eq. (2.21) stating the asymptotic normality of  $\widehat{\theta}_k$  is a direct consequence of

$$\widehat{h}_n \xrightarrow{d} \widehat{h} = I_{\theta_0}^{-1}(\lambda b + G) \sim \mathcal{N}(\lambda I_{\theta_0}^{-1}b, I_{\theta_0}^{-1}).$$

The second part of Theorem 2.2, that is, the asymptotic uniqueness of the MLE, is a consequence of the strict concavity stated in Proposition 2.1: with large probability the log-likelihood function is strictly concave on  $H_n$  and hence the score equation  $\frac{\partial}{\partial h} \widetilde{L}_{k,m}(h) = 0$  has a unique solution on  $H_n$ . For  $n$  large, the normalised MLE

$$\widehat{h}_n^i = (\widehat{\gamma}_n^i, (\widehat{\mu}_n^i - b_m)/a_m, \widehat{\sigma}_n^i/a_m - 1), \quad i = 1, 2$$

belong to  $H_n$  with large probability and solve  $\frac{\partial}{\partial h} \widetilde{L}_{k,m}(h) = 0$ . This implies that  $\widehat{h}_n^1 = \widehat{h}_n^2$ , and hence  $\widehat{\theta}_n^1 = \widehat{\theta}_n^2$ , with probability tending to one. □

## Appendix A: Formulas for the information matrix and bias

According to Prescott and Walden [21] (see also Beirlant et al. [2], page 169), the information matrix of the GEV model at point  $\theta_0 = (\gamma_0 \neq 0, 0, 1)$  is given by

$$I_{\theta_0} = \begin{pmatrix} \frac{1}{\gamma_0^2} \left( \frac{\pi^2}{6} + \left(1 - \gamma_* + \frac{1}{\gamma_0}\right)^2 - \frac{2q}{\gamma_0} + \frac{p}{\gamma_0^2} \right) & -\frac{1}{\gamma_0} \left( q - \frac{p}{\gamma_0} \right) & -\frac{1}{\gamma_0^2} \left( 1 - \gamma_* - q + \frac{1-r+p}{\gamma_0} \right) \\ -\frac{1}{\gamma_0} \left( q - \frac{p}{\gamma_0} \right) & p & -\frac{p-r}{\gamma_0} \\ -\frac{1}{\gamma_0^2} \left( 1 - \gamma_* - q + \frac{1-r+p}{\gamma_0} \right) & -\frac{p-r}{\gamma_0} & \frac{1}{\gamma_0^2} (1 - 2r + p) \end{pmatrix},$$

where  $\Gamma$  is Euler’s Gamma function,  $\gamma_* = 0.5772157$  is Euler’s constant and

$$p = (1 + \gamma_0)^2 \Gamma(1 + 2\gamma_0), \quad q = (1 + \gamma_0) \Gamma'(1 + \gamma_0) + \left(1 + \frac{1}{\gamma_0}\right) \Gamma(2 + \gamma_0), \quad r = \Gamma(2 + \gamma_0);$$

for  $\theta_0 = (0, 0, 1)$  obtained by continuity, it equals

$$\begin{pmatrix} 2.42361 & 0.41184 & 0.332485 \\ 0.41184 & 1 & -0.422784 \\ 0.332485 & -0.422784 & 1.82368 \end{pmatrix}. \tag{A.1}$$

**Lemma A.1.** *The information matrix  $I_{\theta_0}$  is positive definite.*

**Proof.** The information matrix  $I_{\theta_0}$  is equal to the covariance of the score vector

$$\left( \frac{\partial}{\partial \gamma} \ell(\theta_0, X), \frac{\partial}{\partial \mu} \ell(\theta_0, X), \frac{\partial}{\partial \sigma} \ell(\theta_0, X) \right), \quad \theta_0 = (\gamma_0, 0, 1), \quad (\text{A.2})$$

where  $X \sim \text{GEV}(\gamma_0, 0, 1)$  and  $\ell(\theta, x)$  is the log-likelihood

$$\ell(\theta, x) = -\left(1 + \frac{1}{\gamma}\right) \log\left(1 + \gamma \frac{x - \mu}{\sigma}\right) - \left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-1/\gamma} - \log \sigma.$$

We have

$$\ell(\theta, x) = (1 + \gamma) \log z - z - \log \sigma \quad \text{with } z = z(\theta, x) = \left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-1/\gamma}.$$

For  $\gamma \neq 0$ , simple computations yield the partial derivatives, with  $z_0 = z(\theta_0, x)$ ,

$$\frac{\partial z}{\partial \gamma}(\theta_0, x) = z_0 \frac{z_0^{\gamma_0} - 1 - \gamma_0 \log z_0}{\gamma_0^2}, \quad \frac{\partial z}{\partial \mu}(\theta_0, x) = z_0^{1+\gamma_0}, \quad \frac{\partial z}{\partial \sigma}(\theta_0, x) = z_0 \frac{1 - z_0^{\gamma_0}}{\gamma_0}$$

and

$$\begin{aligned} \frac{\partial \ell}{\partial \gamma}(\theta_0, x) &= \log z_0 + (1 + \gamma_0 - z_0) \frac{z_0^{\gamma_0} - 1 - \gamma_0 \log z_0}{\gamma_0^2}, \\ \frac{\partial \ell}{\partial \mu}(\theta_0, x) &= (1 + \gamma_0 - z_0) z_0^{\gamma_0}, \\ \frac{\partial \ell}{\partial \sigma}(\theta_0, x) &= \frac{1 - z_0 - (1 + \gamma_0 - z_0) z_0^{\gamma_0}}{\gamma_0}. \end{aligned}$$

Replacing  $z_0$  in those equations by  $Z = z(\theta_0, X)$  which has a standard exponential distribution when  $X \sim \text{GEV}(\gamma_0, 0, 1)$ , we obtain a simpler representation of the score vector (A.2).

We reason by contradiction and assume that the information matrix is degenerate with rank equal or less than 2. Then the score vector is supported by an hyperplane with equation  $a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 = 0$ ,  $(a_1, a_2, a_3, a_4) \neq 0$ . This implies

$$a_1 \frac{\partial \ell}{\partial \gamma} + a_2 \frac{\partial \ell}{\partial \mu} + a_3 \frac{\partial \ell}{\partial \sigma} + a_4 = 0 \quad (\text{A.3})$$

for almost every  $z > 0$ . By continuity, the equality holds for all  $z > 0$ . But in the space of continuous functions on  $(0, \infty)$ , the functions  $f_1 = \frac{\partial \ell}{\partial \gamma}$ ,  $f_2 = \frac{\partial \ell}{\partial \mu}$ ,  $f_3 = \frac{\partial \ell}{\partial \sigma}$  and  $f_4 = 1$  are linearly independent (they involve different power functions and the logarithm function that are linearly independent). This contradicts the linear relation (A.3) and proves that the information matrix must be positive definite.

For  $\gamma = 0$ , note that the determinant of (A.1) equals 3.45103 hence, being positive ensures positive definiteness.  $\square$

The bias in Theorem 2.2 is given by  $I_{\theta_0}^{-1}b$  where the vector  $b$  can be computed exactly. Calculations are tedious and have been performed with Mathematica<sup>®</sup>. We show the results for  $b = (b_\gamma, b_\mu, b_\sigma)$  when  $\rho \leq 0$  and  $\gamma_0 \neq 0, \gamma_0 + \rho \neq 0$  (the other cases yield also finite values that can be obtained by continuity and integration):

$$\begin{aligned}
 b_\gamma &= \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \gamma}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) H_{\gamma_0, \rho} \left( \frac{1}{-\log s} \right) ds \\
 &= \begin{cases} \frac{1}{\gamma_0^3 \rho (\gamma_0 + \rho)} ((\gamma_0 + \rho)(1 + \gamma_0 - \gamma_* \gamma_0) - (\gamma_0 + \gamma_0^2(1 + \rho) + 2\rho(1 + \gamma_0))\Gamma(1 + \gamma_0) \\ \quad + (1 + \gamma_0)^2 \rho \Gamma(1 + 2\gamma_0) + \gamma_0^2 \Gamma(1 - \rho) - \gamma_0(1 + \gamma_0)\Gamma(2 - \rho) \\ \quad + \gamma_0(1 + \gamma_0)(1 - \rho)\Gamma(1 + \gamma_0 - \rho) - \gamma_0 \rho \Gamma'(2 + \gamma_0) - \gamma_0^2 \Gamma'(2 - \rho)), & \rho < 0, \\ \frac{1}{\gamma_0^4} ((1 + \gamma_0 - \gamma_0 \gamma_*)^2 + \gamma_0^2 \pi^2 / 6 + (1 + \gamma_0)^2 \Gamma(1 + 2\gamma_0) \\ \quad - 2(1 + \gamma_0)[(1 + \gamma_0)\Gamma(1 + \gamma_0) + \gamma_0 \Gamma'(1 + \gamma_0)]), & \rho = 0, \end{cases} \\
 b_\mu &= \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \mu}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) H_{\gamma_0, \rho} \left( \frac{1}{-\log s} \right) ds \\
 &= \begin{cases} \frac{1 + \gamma_0}{\gamma_0 \rho (\gamma_0 + \rho)} (-(\gamma_0 + \rho)\Gamma(1 + \gamma_0) + (1 + \gamma_0)\rho \Gamma(1 + 2\gamma_0) + \gamma_0(1 - \rho)\Gamma(1 + \gamma_0 - \rho)), & \rho < 0, \\ \frac{1 + \gamma_0}{\gamma_0^2} ((1 + \gamma_0)\Gamma(1 + 2\gamma_0) - \Gamma(2 + \gamma_0) - \gamma_0 \Gamma'(1 + \gamma_0)), & \rho = 0, \end{cases} \\
 b_\sigma &= \int_0^1 \frac{\partial^2 \ell}{\partial x \partial \sigma}(\theta_0, \mathcal{Q}_{\gamma_0}(s)) H_{\gamma_0, \rho} \left( \frac{1}{-\log s} \right) ds \\
 &= \begin{cases} \frac{1}{\gamma_0^2 \rho (\gamma_0 + \rho)} (-\gamma_0 - \rho + (1 + \gamma_0)(\gamma_0 + 2\rho)\Gamma(1 + \gamma_0) - (1 + \gamma_0)^2 \rho \Gamma(1 + 2\gamma_0) \\ \quad + \gamma_0 \Gamma(2 - \rho) - \gamma_0(1 + \gamma_0)(1 - \rho)\Gamma(1 + \gamma_0 - \rho)), & \rho < 0, \\ \frac{1}{\gamma_0^3} (-1 + \gamma_0(\gamma_* - 1) - (1 + \gamma_0)^2 \Gamma(1 + 2\gamma_0) \\ \quad + \Gamma(3 + \gamma_0) + \gamma_0(1 + \gamma_0)\Gamma'(1 + \gamma_0)), & \rho = 0. \end{cases}
 \end{aligned}$$

### Appendix B: Bounds for the derivatives of the likelihood

We provide in this section upper bounds for the partial derivatives of the GEV log-likelihood,

$$\ell(\theta, x) = -\left(1 + \frac{1}{\gamma}\right) \log\left(1 + \gamma \frac{x - \mu}{\sigma}\right) - \left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-1/\gamma} - \log \sigma,$$

for  $\theta = (\gamma, \mu, \sigma)$  and  $x$  such that  $1 + \gamma \frac{x - \mu}{\sigma} > 0$ .

**Proposition B.1.** *Let  $\theta_0 = (\gamma_0, 0, 1)$  with  $\gamma_0 \in \mathbb{R}$ . For all  $\varepsilon > 0$ , there exists a neighbourhood  $N_0$  of  $\theta_0$  and a constant  $C > 0$  such that, for all  $\theta \in N_0$  and  $x$  such that  $1 + \gamma(x - \mu)/\sigma > 0$ , we have*

$$\begin{aligned} \left| \frac{\partial \ell}{\partial x^i \partial \theta^{1-i}} \right| &\leq C \max(z^{i\gamma_0-\varepsilon}, z^{1+i\gamma_0+\varepsilon}, z^{\gamma_0-\varepsilon}, z^{1+\gamma_0+\varepsilon}), & i = 0, 1, \\ \left| \frac{\partial^2 \ell}{\partial x^i \partial \theta^{2-i}} \right| &\leq C \max(z^{i\gamma_0-\varepsilon}, z^{1+i\gamma_0+\varepsilon}, z^{2\gamma_0-\varepsilon}, z^{1+2\gamma_0+\varepsilon}), & i = 0, 1, 2, \\ \left| \frac{\partial^3 \ell}{\partial x^i \partial \theta^{3-i}} \right| &\leq C \max(z^{i\gamma_0-\varepsilon}, z^{1+i\gamma_0+\varepsilon}, z^{3\gamma_0-\varepsilon}, z^{1+3\gamma_0+\varepsilon}), & i = 0, 1, 2, 3, \end{aligned}$$

where

$$z = z(\theta, x) = \left( 1 + \gamma \frac{x - \mu}{\sigma} \right)^{-1/\gamma} > 0, \quad \text{for } 1 + \gamma \frac{x - \mu}{\sigma} > 0.$$

The notation  $\partial \theta$  denotes either  $\partial \gamma$ ,  $\partial \sigma$  or  $\partial \mu$  and, similarly for higher order derivatives,  $\partial \theta^2$  denotes  $\partial \gamma^2$ ,  $\partial \gamma \partial \mu$ ,  $\partial \gamma \partial \sigma$ ,  $\partial \mu^2$ ,  $\dots$

Note that the constant  $C > 0$  appearing in Proposition B.1 and in the proofs below may change from line to line. Proposition B.1 gathers with short notations several different inequalities. For instance, the first inequality with  $i = 1$  yields

$$\left| \frac{\partial \ell}{\partial x} \right| \leq C \max(z^{\gamma_0-\varepsilon}, z^{1+\gamma_0+\varepsilon}),$$

while the third inequality with  $i = 1$  yields

$$\begin{aligned} \left| \frac{\partial^3 \ell}{\partial x \partial \gamma \partial \sigma} \right| &\leq C \max(z^{\gamma_0-\varepsilon}, z^{1+\gamma_0+\varepsilon}, z^{3\gamma_0-\varepsilon}, z^{1+3\gamma_0+\varepsilon}) \\ &\leq \begin{cases} C \max(z^{\gamma_0-\varepsilon}, z^{1+3\gamma_0+\varepsilon}) & \text{if } \gamma_0 \geq 0 \\ C \max(z^{3\gamma_0-\varepsilon}, z^{1+\gamma_0+\varepsilon}) & \text{if } \gamma_0 \leq 0. \end{cases} \end{aligned}$$

For  $\theta = (\gamma, \mu, \sigma)$ , we have

$$\ell(\theta, x) = g\left(\gamma, \frac{x - \mu}{\sigma}\right) - \log \sigma, \quad g(\gamma, x) = -\left(1 + \frac{1}{\gamma}\right) \log(1 + \gamma x) - (1 + \gamma x)^{-1/\gamma}, \quad (\text{B.1})$$

where the function  $g$  is the log-likelihood of the one parameter GEV distribution with  $\gamma \in \mathbb{R}$  (i.e.,  $\mu = 0$  and  $\sigma = 1$ ). With the notation

$$z(\gamma, x) = (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0 \quad (\text{B.2})$$

we have,

$$g(\gamma, x) = (1 + \gamma) \log z(\gamma, x) - z(\gamma, x). \quad (\text{B.3})$$

The proof of Proposition B.1 relies on three lemmas providing upper bounds for the derivatives of  $z(\gamma, x)$  and  $g(\gamma, x)$ . Define  $h : \mathbb{R} \rightarrow \mathbb{R}$  by

$$h(x) = \begin{cases} (e^x - 1 - x)/x^2 & \text{for } x \neq 0, \\ 1/2 & \text{for } x = 0 \end{cases} \tag{B.4}$$

and denote by  $h^{(n)}(x)$  its derivative of order  $n = 0, 1, 2$ .

**Lemma B.1.** *The function  $h : \mathbb{R} \rightarrow \mathbb{R}$  in (B.4) is twice continuously differentiable and  $h^{(n)}(x) = O(\max(1, e^x))$  for all  $x \in \mathbb{R}$ , and  $n = 0, 1, 2$ .*

**Proof of Lemma B.1.** The function  $h$  can be represented as the power series  $h(x) = \sum_{n \geq 0} x^n / (n + 2)!$  and is hence indefinitely continuously differentiable. From the asymptotic behaviour

$$h(x) \sim -\frac{1}{x} \quad \text{as } x \rightarrow -\infty, \quad h(x) \sim \frac{e^x}{x^2} \quad \text{as } x \rightarrow +\infty,$$

we deduce that the function  $x \mapsto |h(x)| / \max(1, e^x)$  is bounded on  $\mathbb{R}$  since it is continuous with vanishing limits at  $\pm\infty$ . This proves the existence of  $C > 0$  such that  $|h(x)| \leq C \max(1, e^x)$  for all  $x \in \mathbb{R}$ . The upper bound for the second and third derivatives is proved similarly since simple computations show that

$$h'(x) \sim \frac{1}{x^2} \quad \text{as } x \rightarrow -\infty, \quad h'(x) \sim \frac{e^x}{x^2} \quad \text{as } x \rightarrow +\infty$$

and

$$h''(x) \sim -\frac{2}{x^3} \quad \text{as } x \rightarrow -\infty, \quad h''(x) \sim \frac{e^x}{x^2} \quad \text{as } x \rightarrow +\infty.$$

□

From Lemma B.1, it follows

$$|h^{(n)}(\gamma \log z)| = O(\max(1, z^\gamma)), \quad n = 0, 1, 2. \tag{B.5}$$

We also use throughout the elementary bound  $|\log z|^k = O(z^{\pm\delta})$ , for all  $k \in \mathbb{N}$  and  $\delta > 0$ .

**Lemma B.2.** *Consider the function  $z(\gamma, x)$  defined by Equation (B.2). For all  $\gamma_0 \in \mathbb{R}$  and  $\varepsilon > 0$ , there exists  $C > 0$  and  $\delta > 0$  such that, for all  $\gamma \in (\gamma_0 - \delta, \gamma_0 + \delta)$  and  $1 + \gamma x > 0$ ,*

$$\begin{aligned} \left| \frac{\partial z}{\partial x^i \partial \gamma^{1-i}} \right| &\leq C \max(z^{1+i\gamma_0 \pm \varepsilon}, z^{1+\gamma_0 \pm \varepsilon}), & i = 0, 1, \\ \left| \frac{\partial^2 z}{\partial x^i \partial \gamma^{2-i}} \right| &\leq C \max(z^{1+i\gamma_0 \pm \varepsilon}, z^{1+2\gamma_0 \pm \varepsilon}), & i = 0, 1, 2, \\ \left| \frac{\partial^3 z}{\partial x^i \partial \gamma^{3-i}} \right| &\leq C \max(z^{1+i\gamma_0 \pm \varepsilon}, z^{1+3\gamma_0 \pm \varepsilon}), & i = 0, 1, 2, 3. \end{aligned}$$

**Proof of Lemma B.2.** Recall the definition (B.4) of the function  $h$ . The first-order partial derivatives of  $z$  equal

$$\frac{\partial z}{\partial x} = -z^{1+\gamma} \quad \text{and} \quad \frac{\partial z}{\partial \gamma} = z(\log z)^2 h(\gamma \log z).$$

Assuming  $\gamma \in (\gamma_0 - \delta, \gamma_0 + \delta)$ , we deduce

$$\left| \frac{\partial z}{\partial x} \right| \leq C \max(z^{1+\gamma_0 \pm \delta}) \quad \text{and} \quad \left| \frac{\partial z}{\partial \gamma} \right| \leq C \max(z^{1 \pm \delta}, z^{1+\gamma_0 \pm 2\delta}) \tag{B.6}$$

from which the bounds of the first-order partial derivatives of  $z$  follow. The second-order partial derivatives of  $z$  are given by

$$\begin{aligned} \frac{\partial^2 z}{\partial x^2} &= -(1 + \gamma)z^\gamma \frac{\partial z}{\partial x}, \\ \frac{\partial^2 z}{\partial x \partial \gamma} &= -(1 + \gamma)z^\gamma \frac{\partial z}{\partial \gamma} - (\log z)z^{1+\gamma}, \\ \frac{\partial^2 z}{\partial \gamma^2} &= \{(\log z + 2)(\log z)h(\gamma \log z) + \gamma(\log z)^2 h'(\gamma \log z)\} \frac{\partial z}{\partial \gamma} + z(\log z)^3 h'(\gamma \log z). \end{aligned}$$

Combined with the previous bounds and (B.5), we obtain, for  $\gamma \in (\gamma_0 - \delta, \gamma_0 + \delta)$ ,

$$\begin{aligned} \left| \frac{\partial^2 z}{\partial x^2} \right| &\leq C \max(z^{1+2\gamma_0 \pm 2\delta}), \\ \left| \frac{\partial^2 z}{\partial \gamma \partial x} \right| &\leq C \max(z^{1+\gamma_0 \pm 2\delta}, z^{1+2\gamma_0 \pm 2\delta}), \\ \left| \frac{\partial^2 z}{\partial \gamma^2} \right| &\leq C \max(z^{1 \pm 2\delta}, z^{1+2\gamma_0 \pm 4\delta}) \end{aligned}$$

from which the bounds for the second-order partial derivatives of  $z$  follow. The case of third-order derivatives can be dealt with similarly and we omit the details.  $\square$

**Lemma B.3.** Consider the function  $g(\gamma, x)$  defined by Equation (B.3). For all  $\gamma_0 \in \mathbb{R}$  and  $\varepsilon > 0$ , there exists  $C > 0$  and  $\delta > 0$  such that, for all  $\gamma \in (\gamma_0 - \delta, \gamma_0 + \delta)$  and  $1 + \gamma x > 0$ ,

$$\begin{aligned} \left| \frac{\partial g}{\partial x^i \partial \gamma^{1-i}} \right| &\leq C \max(z^{i\gamma_0 - \varepsilon}, z^{1+i\gamma_0 + \varepsilon}, z^{\gamma_0 - \varepsilon}, z^{1+\gamma_0 + \varepsilon}), \quad i = 0, 1, \\ \left| \frac{\partial^2 g}{\partial x^i \partial \gamma^{2-i}} \right| &\leq C \max(z^{i\gamma_0 - \varepsilon}, z^{1+i\gamma_0 + \varepsilon}, z^{2\gamma_0 - \varepsilon}, z^{1+2\gamma_0 + \varepsilon}), \quad i = 0, 1, 2, \\ \left| \frac{\partial^3 g}{\partial x^i \partial \gamma^{3-i}} \right| &\leq C \max(z^{i\gamma_0 - \varepsilon}, z^{1+i\gamma_0 + \varepsilon}, z^{3\gamma_0 - \varepsilon}, z^{1+3\gamma_0 + \varepsilon}), \quad i = 0, 1, 2, 3. \end{aligned}$$

**Proof of Lemma B.3.** Equation (B.3) expressing  $g(\gamma, x)$  in terms of  $z(\gamma, x)$  entails

$$\frac{\partial g}{\partial x} = \left( \frac{1 + \gamma}{z} - 1 \right) \frac{\partial z}{\partial x}, \quad \frac{\partial g}{\partial \gamma} = \left( \frac{1 + \gamma}{z} - 1 \right) \frac{\partial z}{\partial \gamma} + \log z.$$

Using the upper bound for the first derivatives of  $z(\gamma, x)$  in Lemma B.2, we get, for  $\gamma \in (\gamma_0 - \delta, \gamma_0 + \delta)$ ,

$$\left| \frac{\partial g}{\partial x} \right| \leq C \max(z^{\gamma_0 \pm \varepsilon}, z^{1 + \gamma_0 \pm \varepsilon}), \quad \left| \frac{\partial g}{\partial \gamma} \right| \leq C \max(z^{\pm \varepsilon}, z^{1 \pm \varepsilon}, z^{\gamma_0 \pm \varepsilon}, z^{1 + \gamma_0 \pm \varepsilon}).$$

This proves the upper bounds for the first derivatives of  $g(\gamma, x)$  given in Lemma B.3. Note that we can handle the  $\pm$  sign since we have  $-\varepsilon < \varepsilon < 1 - \varepsilon < 1 + \varepsilon$ , for small  $\varepsilon > 0$ . When considering the maximum of the power functions, only the extreme exponents  $-\varepsilon < 1 + \varepsilon$  matter. A similar argument holds for  $\gamma_0 - \varepsilon < \gamma_0 + \varepsilon < 1 + \gamma_0 - \varepsilon < 1 + \gamma_0 + \varepsilon$ .

Similarly for the second-order derivatives of  $g(\gamma, x)$ , the upper bounds are derived from the similar upper bounds for the partial derivatives of  $z(\gamma, x)$  in Lemma B.2 (with  $\varepsilon$  replaced by  $\varepsilon/2$ ) together with the formulas

$$\begin{aligned} \frac{\partial^2 g}{\partial x^2} &= \left( \frac{1 + \gamma}{z} - 1 \right) \frac{\partial^2 z}{\partial x^2} - \frac{1 + \gamma}{z^2} \left( \frac{\partial z}{\partial x} \right)^2, \\ \frac{\partial^2 g}{\partial x \partial \gamma} &= \left( \frac{1 + \gamma}{z} - 1 \right) \frac{\partial^2 z}{\partial x \partial \gamma} + \frac{1}{z} \frac{\partial z}{\partial x} - \frac{1 + \gamma}{z^2} \frac{\partial z}{\partial x} \frac{\partial z}{\partial \gamma}, \\ \frac{\partial^2 g}{\partial \gamma^2} &= \left( \frac{1 + \gamma}{z} - 1 \right) \frac{\partial^2 z}{\partial \gamma^2} + \frac{2}{z} \frac{\partial z}{\partial \gamma} - \frac{1 + \gamma}{z^2} \left( \frac{\partial z}{\partial \gamma} \right)^2. \end{aligned}$$

Partial derivatives of order 3 are dealt with similarly. □

**Proof of Proposition B.1.** Equation (B.1) implies that the partial derivatives of  $\ell$  are closely related to those of  $g$ . For the first derivatives, we have

$$\begin{aligned} \frac{\partial \ell}{\partial x}(\theta, x) &= \frac{1}{\sigma} \frac{\partial g}{\partial x} \left( \gamma, \frac{x - \mu}{\sigma} \right), & \frac{\partial \ell}{\partial \mu}(\theta, x) &= -\frac{1}{\sigma} \frac{\partial g}{\partial x} \left( \gamma, \frac{x - \mu}{\sigma} \right), \\ \frac{\partial \ell}{\partial \gamma}(\theta, x) &= \frac{\partial g}{\partial \gamma} \left( \gamma, \frac{x - \mu}{\sigma} \right), & \frac{\partial \ell}{\partial \sigma}(\theta, x) &= -\frac{x - \mu}{\sigma^2} \frac{\partial g}{\partial x} \left( \gamma, \frac{x - \mu}{\sigma} \right) - \frac{1}{\sigma}. \end{aligned}$$

These equalities together with Lemma B.3 yield

$$\left| \frac{\partial \ell}{\partial x} \right| = \frac{1}{\sigma} \left| \frac{\partial g}{\partial x} \right| \leq C \max(z^{\gamma_0 - \varepsilon}, z^{1 + \gamma_0 + \varepsilon}),$$

which corresponds to the first inequality in Proposition B.1 with  $i = 0$ . We also obtain

$$\left| \frac{\partial \ell}{\partial \gamma} \right| = \left| \frac{\partial g}{\partial \gamma} \right| \leq C \max(z^{-\varepsilon}, z^{1 + \varepsilon}, z^{\gamma_0 - \varepsilon}, z^{1 + \gamma_0 + \varepsilon}),$$

$$\left| \frac{\partial \ell}{\partial \mu} \right| = \frac{1}{\sigma} \left| \frac{\partial g}{\partial x} \right| \leq C \max(z^{\gamma_0 - \varepsilon}, z^{1 + \gamma_0 + \varepsilon})$$

which implies the inequality in Proposition B.1 with  $i = 1$  and the derivatives taken with respect to  $\gamma$  and  $\mu$  respectively. The case of the derivative with respect to  $\sigma$  is slightly more difficult: we use the inequality

$$\left| \frac{x - \mu}{\sigma} \right| = \left| \frac{z^{-\gamma} - 1}{\gamma} \right| \leq C \max(z^{-\gamma_0 - \delta}, z^{-\gamma_0 + \delta}, 1), \quad \gamma \in (\gamma_0 - \delta, \gamma_0 + \delta),$$

which implies

$$\begin{aligned} \left| \frac{\partial \ell}{\partial \sigma} \right| &= \frac{1}{\sigma} \left| \frac{z^{-\gamma} - 1}{\gamma} \frac{\partial g}{\partial x} + 1 \right| \leq C \max(z^{-\gamma_0 - \delta}, z^{-\gamma_0 + \delta}, 1) \max(z^{\gamma_0 - \varepsilon}, z^{1 + \gamma_0 + \varepsilon}) \\ &\leq C \max(z^{-\varepsilon'}, z^{1 + \varepsilon'}, z^{\gamma_0 - \varepsilon'}, z^{1 + \gamma_0 + \varepsilon'}) \end{aligned}$$

for sufficiently small  $\varepsilon'$ . For the first inequality, we use Lemma B.3. This proves the first inequality in Proposition B.1 with  $i = 1$  and the derivatives taken with respect to  $\sigma$ .

The case of second-order derivatives is dealt similarly with the relations

$$\begin{aligned} \frac{\partial^2 \ell}{\partial x^2} &= \frac{1}{\sigma^2} \frac{\partial^2 g}{\partial x^2}, & \frac{\partial^2 \ell}{\partial \mu^2} &= \frac{1}{\sigma^2} \frac{\partial^2 g}{\partial x^2}, \\ \frac{\partial^2 \ell}{\partial \gamma \partial x} &= \frac{1}{\sigma} \frac{\partial^2 g}{\partial \gamma \partial x}, & \frac{\partial^2 \ell}{\partial \sigma^2} &= 2 \frac{x - \mu}{\sigma^3} \frac{\partial g}{\partial x} + \frac{(x - \mu)^2}{\sigma^4} \frac{\partial^2 g}{\partial x^2} + \frac{1}{\sigma^2}, \\ \frac{\partial^2 \ell}{\partial \mu \partial x} &= -\frac{1}{\sigma^2} \frac{\partial^2 g}{\partial x^2}, & \frac{\partial^2 \ell}{\partial \gamma \partial \mu} &= -\frac{1}{\sigma} \frac{\partial^2 g}{\partial \gamma \partial x}, \\ \frac{\partial^2 \ell}{\partial \sigma \partial x} &= -\frac{x - \mu}{\sigma^3} \frac{\partial^2 g}{\partial x^2}, & \frac{\partial^2 \ell}{\partial \gamma \partial \sigma} &= -\frac{x - \mu}{\sigma^2} \frac{\partial^2 g}{\partial \gamma \partial x}, \\ \frac{\partial^2 \ell}{\partial \gamma^2} &= \frac{\partial^2 g}{\partial \gamma^2}, & \frac{\partial^2 \ell}{\partial \mu \partial \sigma} &= \frac{1}{\sigma^2} \frac{\partial g}{\partial x} + \frac{x - \mu}{\sigma^3} \frac{\partial^2 g}{\partial x^2}. \end{aligned}$$

Using these relations, the second inequality in Proposition B.1 follows from Lemma B.3. Checking all the different cases is relatively tedious but elementary. Details are omitted. Similar formulas hold for derivatives of order 3 and the resulting bounds have been checked with Mathematica<sup>®</sup>; cf. supplemental article Dombry and Ferreira [9]. □

## Acknowledgements

The authors would like to thank Laurens de Haan for his useful suggestions and support. We would like to thank two anonymous referees, the Associate Editor and the Editor for their comments that improved the paper. In particular the proof of Lemma A.1 is based on the Associate Editor's suggestion.



Research partially funded by VolkswagenStiftung Support for Europe – WEX-MOP and, FCT – Fundação para a Ciência e a Tecnologia, Portugal, through UID/MAT/00006/2013 and UID/Multi/04621/2013.

## Supplementary Material

**Mathematica<sup>®</sup> code and output on partial derivatives of the GEV log-likelihood** (DOI: [10.3150/18-BEJ1032SUPP](https://doi.org/10.3150/18-BEJ1032SUPP); .pdf). This supplement provides the Mathematica<sup>®</sup> code and output to obtain explicit derivatives up to third order of the log-likelihood of the generalized extreme value distribution.

## References

- [1] Balkema, A.A. and de Haan, L. (1974). Residual life time at great age. *Ann. Probab.* **2** 792–804. [MR0359049](#)
- [2] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. With contributions from Daniel De Waal and Chris Ferro. [MR2108013](#)
- [3] Bücher, A. and Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes* **20** 839–872. [MR3737387](#)
- [4] Bücher, A. and Segers, J. (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462. [MR3706798](#)
- [5] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. London: Springer. [MR1932132](#)
- [6] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. [MR2234156](#)
- [7] Dietrich, D., de Haan, L. and Hüslér, J. (2002). Testing extreme value conditions. *Extremes* **5** 71–85. [MR1947789](#)
- [8] Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21** 420–436. [MR3322325](#)
- [9] Dombry, C. and Ferreira, A. (2019). Supplement to “Maximum likelihood estimators based on the block maxima method.” DOI:[10.3150/18-BEJ1032SUPP](https://doi.org/10.3150/18-BEJ1032SUPP).
- [10] Drees, H., de Haan, L. and Li, D. (2003). On large deviation for extremes. *Statist. Probab. Lett.* **64** 51–62. [MR1995809](#)
- [11] Drees, H., de Haan, L. and Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *J. Statist. Plann. Inference* **136** 3498–3538. [MR2284668](#)
- [12] Drees, H., Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14** 1179–1201. [MR2071420](#)
- [13] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance*. *Applications of Mathematics (New York)* **33**. Berlin: Springer. [MR1458613](#)
- [14] Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298. [MR3285607](#)
- [15] Gumbel, E.J. (1958). *Statistics of Extremes*. New York: Columbia Univ. Press. [MR0096342](#)

- [16] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29** 339–349. [MR0906643](#)
- [17] Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27** 251–261. [MR0797563](#)
- [18] Markovich, N. (2007). *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*. Wiley Series in Probability and Statistics. Chichester: Wiley. [MR2364666](#)
- [19] Novak, S.Y. (2011). *Extreme Value Methods with Applications to Finance*. Monographs on Statistics and Applied Probability **122**. Boca Raton, FL: CRC Press. [MR2933280](#)
- [20] Pickands, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. [MR0423667](#)
- [21] Prescott, P. and Walden, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67** 723–724. [MR0601119](#)
- [22] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: Wiley. [MR0838963](#)
- [23] Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72** 67–90. [MR0790201](#)
- [24] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics **3**. Cambridge: Cambridge Univ. Press. [MR1652247](#)

Received February 2017 and revised March 2018