



Empirical Bayes inference for the block maxima method

SIMONE A. PADOAN ^a and STEFANO RIZZELLI ^b

¹*Department of Decision Sciences, Bocconi University, Milano, Italy, simone.padoan@unibocconi.it*

²*Department of Statistical Sciences, Catholic University, Milano, Italy, stefano.rizzelli@unicatt.it*

The block maxima method is one of the most popular approaches for extreme value analysis with independent and identically distributed observations in the domain of attraction of an extreme value distribution. The lack of a rigorous study on the Bayesian inference in this context has limited its use for statistical analysis of extremes. In this paper we propose an empirical Bayes procedure for inference on the block maxima law and its related quantities. We show that the posterior distributions of the tail index of the data distribution and of the return levels (representative of future extreme episodes) are consistent and asymptotically normal. These properties guarantee the reliability of posterior-based inference. We also establish contraction rates of the posterior predictive distribution, the key tool in Bayesian probabilistic forecasting. Posterior computations are readily obtained via an efficient adaptive Metropolis-Hasting type of algorithm. Simulations show its excellent inferential performances already with modest sample sizes. The utility of our proposal is showcased analysing extreme winds generated by hurricanes in Southeastern US.

Keywords: Contraction Rate; Extreme Quantiles; Posterior Consistency; Return Levels; Tail Index; Wind Speed.

1. Introduction

Extreme Value Theory (EVT) provides a mathematical foundation to analyse extreme events regardless of the unknown underlying distribution. In this paper we focus on the univariate *Block-Maxima* (BM) approach, see, e.g., Chapter 5 in [1] and Chapter 3 in [5]. Namely, we consider a dataset of n values from an unknown distribution, which are then divided in k blocks say of m observations and from which k maxima are computed. These, if suitably normalised, are asymptotically distributed according to the Generalised Extreme Value (GEV) distribution for an increasing block-size, provided that some weak conditions are satisfied, see Chapter 1 in [10]. EVT's supreme aim is to provide the basis for the statistical prediction of future extreme events, a vital task for risk management. In statistics, a robust approach to prediction is through the Bayesian paradigm, which naturally takes into account model uncertainty. Several Bayesian analyses of univariate BM and other extremes have been proposed over time, see, e.g., Chapter 11 in [1], Chapter 9 in [5] and [7,8,28,29,36], to name a few. Nevertheless, the potential of Bayesian inference does not seem to be fully exploited in earlier BM literature.

The first motivation is a reluctance to use asymptotic models, as the degree of accuracy of Bayesian inference based on them is not fully understood. In real applications end users analyse BM pretending that they are exactly distributed according to a GEV distribution, the so-called *vanilla* approach, ignoring that it is in fact a misspecified model for the data, since BM with fixed block-size are only approximately GEV distributed. The vanilla approach leads to the naive conclusion that estimators of the GEV parameters are asymptotically unbiased, and end users are inclined to believe that both a large block-size and number of blocks are necessary to have negligible bias in practice. This is a problem as a large data sample may not be available in many applications. In the frequentist context, recent theory states that there is a potential asymptotic bias whenever the block-size does not increase sufficiently fast along with the number of blocks, thus involving the typical applications where one worries to have

a very large block-size or, in alternative, a large number of blocks. In fact, although basic estimators such as the Maximum Likelihood (ML) and Probability Weighted Moments (PWM) (e.g., [22,23]) have been introduced decades ago, for a long time the only available asymptotic theory has been that derived under the unrealistic vanilla framework (see, e.g., [3,14,35]; see also [1] and [10] and the references therein for other estimators), while a proper theory that takes the model misspecification issue into account is only recently developing, see [15–17]. We stress that standard likelihood-based inference results under model misspecification (e.g., Theorem 5.5 in [25]) do not apply for the block maxima method, as can be understood from the latter references. A theory for Bayesian analysis of BM based on the misspecified GEV model is still missing.

The second motivation regards the complexity of the GEV class. Specifically, the GEV is a three parameter (location, scale and shape) superset of distributions, comprehending three different subfamilies, known as short-, light- and heavy-tailed, depending on the values of the shape parameter. The latter, known as the tail index, describes the tail heaviness of a distribution (see Chapter 1 in [10]). Notably, its sign affects the GEV support, thus the GEV is an *irregular* class. Location and scale are fixed parameters under the unrealistic vanilla framework, while they are block-size dependent in the more realistic BM approach. Accordingly, the elicitation of a prior distribution for such parameters, capable of taking the block-size into account, is not an easy task under a genuinely Bayesian approach. Moreover, the expression of the posterior distribution is not available in closed-form, therefore posterior-based inference relies on the generation of samples from the posterior via Markov Chain Monte Carlo (MCMC) methods. In this regard, the dependence among parameters entails that standard MCMC procedures can be computationally and statistically inefficient when the prior distribution is not appropriately specified. Note that in some literature other limiting distributions, like for example the Generalised Pareto, are embedded in a wider class of distributions in order to model simultaneously the bulk and the tail of a distribution. The resulting model is flexible enough for modelling the entire dataset and although is still misspecified, classical Bayesian methods are not necessarily subject to the issues discussed here (e.g., [9]).

In this article, we propose an Empirical Bayes (EB) procedure for statistical inference on the BM law and its related quantities that circumvents the above mentioned complications. Our setting leverages on the GEV approximation, but regarding the GEV distribution as a misspecified model for the data. The most natural choice, in our opinion, is to define an overall prior distribution consisting of a data-independent prior for the tail index and data-dependent prior distributions for the location and scale parameters, overcoming the difficulties of an orthodox prior selection. Our EB method does not allow to obtain a genuine posterior distribution since it does not follow an authentic Bayesian formulation, but as typically happens for the latter, the former produces a proper distribution, which is shown to comply with desirable asymptotic properties. In the sequel, we use the term “posterior” instead of “quasi-posterior”, for simplicity. EB has a long tradition as an informal Bayesian approach that helps in the selection of the hyperparameters, which are estimated from the data, and gives rise to more robust inferential procedures than those based on more discretionary prior distribution specification (see, e.g., Chapter 4.6 in [27]). In our context, since location and scale parameters depend on the way that the data are blocked the benefit of EB in the prior distribution elicitation of such parameters is even more prominent.

The first main contribution of this paper is to derive a solid theory for the proposed method to guarantee trustworthy inferences. We give simple conditions on the prior distribution to obtain a consistent posterior distribution of the GEV parameters. In particular, we quantify the rate at which the posterior distribution concentrates around the true parameter values, since this is useful to understand whether accurate inference is achievable in practice with the available sample size. We then establish that the posterior distribution is asymptotically close to a normal distribution. This result is particularly useful to: show that posterior credible intervals have good frequentist coverage probability (close to the nom-

inal confidence level); to derive simple approximate credible regions for the entire parameter vector, preventing practical complications in the derivation of such regions for multidimensional parameters. We complete our analysis of posterior distributions extending the aforementioned theoretical properties, i.e. consistency with suitable contraction rates and asymptotic normality, to that of the so-called *return level* corresponding to the *return period* T , namely the value that is expected to be exceeded by the block maximum on average every T time periods (e.g., Chapter 3 in [5]). The scope of the a posteriori distribution of the return level is however limited to estimation activities (e.g. to produce point or interval estimates), while prediction of extreme events, which is a far more ambitious task, goes beyond it.

The ultimate goal of an extreme values analysis is the prediction of future extreme values, potentially larger than those observed in the available dataset. Since the use of posterior predictive distributions is undoubtedly one of the strengths of the Bayesian paradigm, from a forecasting view point, we focus then on the posterior predictive distribution of the block maximum. So far, the use of posterior predictive distribution has been limited to point forecasting of the annual maximum, as far as we know (see, e.g., [6]). In this regard, our second main contribution is to provide for the first time a general definition of the posterior predictive distribution of a future unobserved block maximum, embedding as a special case that of the annual maxima, with which the canonical return level point forecast is derived. Our general formulation allows for a proper *probabilistic forecasting* approach to extreme events, which, to the best of our knowledge, has not been discussed in earlier contributions. Moreover, we establish important asymptotic results that guarantee the accuracy of posterior-predictive-based inference.

The third main contribution of this paper is practical. Firstly, we provide simple concrete examples of prior distributions that meet the conditions required by our theory. Secondly, we propose to adopt a simple adaptive random-walk Metropolis-Hasting algorithm (see [21]) for the calculation of the posterior distribution, since by embedding our proposed prior distributions it turns out to be computationally and statistically efficient. By an extensive simulation study we show that the empirical posterior distributions obtained via this sampling scheme comply with the theory. Most importantly, we demonstrate that accurate inference is achievable with moderate dimensions of block-size and blocks number, realistically available in the vast majority of applications.

Recently, [30] have established consistency of an empirical Bayes approach to inference on multivariate max-stable distributions, whose margins are all of one the following three types: reverse-Weibull, Gumbel or Fréchet. This article, dealing with the (simpler) univariate case, aims at a greater degree of generality, insofar as working with the GEV family allows to cover all the aforementioned distributions simultaneously. Compared to the earlier work, we provide here a more complete picture: a finer theory for the method, along with an algorithm and codes for its implementation. This work is accessible to a wider scientific community, due to simplicity and popularity of the univariate case, and can be used as a preliminary reading before approaching the multivariate theory.

Our methods and data have been incorporated into the R package `ExtremeRisks`, freely available on CRAN. The paper layout is the following. Section 2 explains in detail our statistical context and introduces our empirical Bayes method. Section 3 provides the asymptotic theory for the posterior distribution of the parameters of the GEV family, the return levels, and for the posterior predictive distribution. Section 4 describes practical aspects of the posterior distribution computation. The finite sample performance of the proposed methods is examined via simulation in Section 5 and on periodic maxima wind speed data generated by hurricanes in Southeastern US, in Section 6. The article ends with a conclusive discussion in Section 7 and with the main proofs in 8.

2. Background

2.1. Setting and notation

Let X_1, \dots, X_m be independent and identically distributed (iid) random variables with common (unknown) distribution F which is assumed to be in the domain of attraction of a distribution G_γ , where $\gamma \in \mathbb{R}$ is the tail index, in symbols $F \in \mathcal{D}(G_\gamma)$. This means that there are sequences $a_m > 0$ and $b_m \in \mathbb{R}$ such that

$$\lim_{m \rightarrow \infty} F^m(a_mx + b_m) = G_\gamma(x), \quad (1)$$

for all continuity points of G_γ , where the latter is a GEV distribution. From a purely distributional view point GEV distribution function is of the form $G_\theta(x) := G_\gamma((x - \mu)/\sigma)$ for $x \in \mathcal{S}_\theta$, where $\theta := (\gamma, \mu, \sigma)^\top$ ranges over $\mathbb{R}^2 \times (0, \infty)$, while

$$G_\gamma(z) = \begin{cases} \exp\left(- (1 + \gamma z)_+^{-1/\gamma}\right), & \gamma \neq 0, \\ \exp(-\exp(-z)), & \gamma = 0, \end{cases}$$

and \mathcal{S}_θ is the set $[\mu - \sigma/\gamma, \infty)$ or \mathbb{R} or $(-\infty, \mu - \sigma/\gamma]$, if $\gamma > 0$ or $\gamma = 0$ or $\gamma < 0$, respectively, see Chapter 1 in [10] for details. In particular, for $x \in \mathbb{R}$, $(x)_+ = \max(0, x)$. Parameters μ and σ are the location and scale, while γ is informative on how heavy the distribution tail is. We recall that G_θ is a *max-stable* distribution, i.e. there are a suitable positive function s and a real-valued function c such that $G_\theta^t(s(t)x + c(t)) = G_\theta(x)$ for each $t \geq 1$. The corresponding density is $g_\theta(x) = g_\gamma((x - \mu)/\sigma)/\sigma$, with

$$g_\gamma(z) = \begin{cases} (1 + \gamma z)_+^{-1/\gamma - 1} \exp\left(- (1 + \gamma z)_+^{-1/\gamma}\right), & \gamma \neq 0, \\ \exp(-\exp(-z)), & \gamma = 0. \end{cases} \quad (2)$$

The vanilla approach simply assumes that data are exactly distributed according to G_θ , while the BM method considered in this paper adopts a different data generating mechanism. Specifically, we consider X_1, \dots, X_n iid random variables, $n = 1, 2, \dots$, whose distribution satisfies $F \in \mathcal{D}(G_\gamma)$. We then assume that the n variables can be split into $k \geq 1$ blocks of size $m \geq 1$, so that $n = mk$. The BM method is now concerned with the analysis of the k block-maxima, where the i -th block maximum is given by

$$M_{m,i} = \max(X_{(i-1)m+1}, \dots, X_{im}), \quad i \in \{1, \dots, k\},$$

and whose distribution is F^m . By the domain of attraction condition in (1), also known as *first-order* condition, the BM distribution can be *approximated* by the GEV distribution G_{θ_m} with $\theta_m := (\gamma, b_m, a_m)$, for a large enough given block-size m . The first-order condition is equivalent to the condition

$$\lim_{m \rightarrow \infty} \frac{V(mx) - V(m)}{a(m)} = \frac{x^\gamma - 1}{\gamma} =: Q_{G_\gamma}(1 - e^{-1/x}), \quad \forall x > 0, \quad (3)$$

where a is a positive function, $V(y) := F^{\leftarrow}(e^{-1/y})$, $y > 0$, $Q_{G_\gamma}(p) := G_\gamma^{\leftarrow}(1 - p)$, $p \in (0, 1)$ and f^{\leftarrow} denotes the generalised left-continuous inverse of a non-decreasing right-continuous function f (see Chapter 1 in [10]), on which basis a possible selection for the norming constants is $b_m = V(m)$ and $a_m = a(m)$. Result (3) is also useful because it allows to approximate an extreme quantile of the distribution F via

$$Q_F(p) := F^{\leftarrow}(1 - p) \approx Q_{G_{\theta_m}}(p_m) := b_m + a_m Q_{G_\gamma}(p_m), \quad (4)$$

for a fixed small exceeding probability p and for a large enough m , where $p_m = 1 - (1 - p)^m$. In this context, to study the asymptotic behaviour of some estimators (see, e.g., the estimators in Chapter 3 of [10] and [15,16], to name a few), the following *second-order* condition has been introduced

$$\lim_{t \rightarrow \infty} \frac{\frac{V(tx) - V(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma, \rho}(x), \quad \forall x > 0, \quad (5)$$

where $A(t)$ is a positive or negative function satisfying $A(t) \rightarrow 0$ as $t \rightarrow \infty$ such that $|A|$ is a regularly varying function with index $\rho \leq 0$, named the second-order parameter, while $H_{\gamma, \rho}$ is a non-null function whose expression depends on γ and ρ , see [12] and Appendix B of [10] for details. These are useful in practice for the derivation of the non-negligible bias factor of an estimator, due to model misspecification. In particular, Dombry and Ferreira [16] use them for developing the asymptotic theory of ML estimation in the frequentist context.

In this paper, we extend their results to the Bayesian paradigm. We conclude this section with some notation used throughout the paper. For any pair of probability measures F, H over a Borel subspace of \mathbb{R} , with Lebesgue densities f, h , we denote by

$$\mathcal{H}(f, h)^2 = \int (\sqrt{f(x)} - \sqrt{h(x)})^2 dx$$

the squared Hellinger distance. For a real valued function f on \mathbb{R} , f' and f'' denote its first and second derivative. The operations cX and X/c , where X is either a vector or a matrix and c is a scalar, are meant componentwise. If X_k has asymptotic distribution F , we use the symbol $X_k \sim F$. For $d = 1, 2, \dots$, $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ -field of \mathbb{R}^d .

2.2. Inference

We specify here the inferential setting used to establish the asymptotic theory for our EB method, presented in the next section. Assume that the sequence X_1, \dots, X_n follows a specific distribution $F_0 \in \mathcal{D}(G_{\gamma_0})$, with $\gamma_0 > -1$. Let $M_{m,i}$, $i = 1, \dots, k$, be the sequence of BM with joint distribution $\prod_{i=1}^k F_0^m(\cdot)$ and whose probability density function is denoted by $f_0^{(m)}$. We assume that both m and k go to infinity and, to avoid asymptotic results based on double limits, in the sequel we assume for simplicity that m depends on k , say $m \equiv m_k$, and that $m \rightarrow \infty$ as $k \rightarrow \infty$.

For a large fixed m , we assume that the family $\{G_{\theta}^k, \theta = (\gamma, b_m, a_m)^\top \in \Theta = (-1, \infty) \times \mathbb{R} \times (0, \infty)\}$ is used as the misspecified statistical model for the sequence $M_{m,i}$, $i = 1, \dots, k$. Note that without loss of generality here and in the sequel we use the simplified notation θ in place of θ_m since the parameter space Θ does not depend on m and, accordingly, the reference statistical model is the same no matter what the block-size is. Analogously, we also use the symbol θ_0 in place of $\theta_{m,0} = (\gamma_0, b_{m,0}, a_{m,0})^\top$. We denote the likelihood function relative to the GEV misspecified class by $L_k(\theta) = \prod_{i=1}^k g_{\theta}(M_{m,i})$, for all $\theta \in \Theta$. Given the GEV log-density,

$$l_{\theta}(x) = \begin{cases} \log g_{\theta}(x), & x \in \mathcal{S}_{\theta}, \\ -\infty, & \text{otherwise,} \end{cases}$$

for all $\theta \in \Theta$, where g_θ is as in (2), then the log-likelihood is simply defined as $l_k(\theta) = \log L_k(\theta) = \sum_{i=1}^k l_\theta(M_{m,i})$. Accordingly, we denote by

$$S_{k,\theta_0}^\top = \left(\frac{1}{\sqrt{k}} \sum_{i=1}^k \frac{\partial l_\theta}{\partial \theta_j}(M_{m,i}) \right)_{j=1,2,3} \Big|_{\theta=\theta_0}, \quad (6)$$

the score process vector of the log-likelihood evaluated at θ_0 .

Similarly to that in Section 4 of [30], the empirical Bayes procedure we propose is based on an overall prior distribution on θ that has a Lebesgue density taking general form

$$\pi_k(\theta) = \pi_{\text{sh}}(\gamma) \pi_{\text{loc}} \left(\frac{b_m - \widehat{b}_{m,k}}{\widehat{a}_{m,k}} \right) \frac{1}{\widehat{a}_{m,k}} \pi_{\text{sc}} \left(\frac{a_m}{\widehat{a}_m} \right) \frac{1}{\widehat{a}_{m,k}}, \quad \forall \theta \in \Theta, \quad (7)$$

where the prior distributions on the location and scale parameters are data-dependent. More precisely, π_{sh} , π_{loc} and π_{sc} are generic probability kernels for the shape γ , rescaled location b_m and rescaled scale a_m parameters, respectively, which are not depending on the data (see Section 3 for more details), while the rescaling terms $\widehat{b}_{m,k}$ and $\widehat{a}_{m,k}$ are estimators of a_m and b_m , respectively. The benefit of the prior specification as in (7) is to address the problem of placing enough mass around true parameters that change with the sample size, with a really simple form. The corresponding EB posterior distribution of the parameters of GEV family is therefore defined by

$$\Psi_k(B) := \frac{\int_B L_k(\theta) \pi_k(\theta) d\theta}{\int_\Theta L_k(\theta) \pi_k(\theta) d\theta}, \quad \forall B \in \mathcal{B}(\Theta),$$

where $\mathcal{B}(\Theta)$ is the class of Borel-subsets of Θ . In the next section we establish the asymptotic properties of such EB posterior distribution.

We point out that our theory benefits from the asymptotic results on the local reparametrised likelihood and score processes developed by Dombry and Ferreira [16]. To save space we refer to Section 3.2 of Supplementary Material [31] for a detailed account, while below we only outline the basic idea. We remind that $G_{\bar{\theta}_0}$ with $\bar{\theta}_0 = (\gamma_0, 0, 1)^\top$ arises as the limit distribution for suitably normalised maxima, see (1). Accordingly, the likelihood theory in [16] is established focusing on a GEV likelihood function defined using normalised maxima as data and the reparametrization

$$\bar{\theta} = r(\theta) := (\gamma, (b_m - b_{m,0})/a_{m,0}, a_m/a_{m,0})^\top, \quad (8)$$

whose corresponding MLE sets then out to estimate $\bar{\theta}_0$. Nevertheless, the likelihood defined on normalised BM is linked to that of unnormalised BM and the asymptotic results can be rephrased for estimation of θ_0 and related quantities. This is especially relevant when interest goes on the far side of inference on the tail behaviour, extending e.g. to extrapolation beyond observed levels, see Section 3.2. A similar reasoning applies to the posterior distribution, namely $\bar{\Psi}_k(B) = \Psi_k(r^{-1}(B))$, where $\bar{\Psi}_k$ is a posterior using normalised BM as data. Since in applications the goal is often both to assess the tail heaviness and to predict future extreme values, in the next section we present the asymptotic results regarding G_{θ_0} and its related quantities, while postponing the technical discussion on $G_{\bar{\theta}_0}$ to Section 8 and Supplementary Material [31].

Extreme value analyses are often based on the annual maximum as it allows for easily interpretable results. According to EVT, the distribution of the block maximum can be approximated by a GEV one as long as m is large and, as the theory in Section 3 suggests, the proposed inferential procedures are accurate provided that the number of maxima k is neither too large (compared to the block-size) nor

too small (to have a sufficient amount of information). What matters in practice is to select m and k that allow to achieve good inferential results. In applications where the interest is in inferring features of the distribution of a maximum over m^* observations, this is still possible by starting from the analysis of maxima with block-size m different from m^* and exploiting max-stability. Hence, the choice of m need not concur with m^* . A method for appropriately selecting the block-size m is proposed in Section 6 of Supplementary Material [31].

3. Theory for Empirical Bayes Inference

3.1. GEV distribution: posterior asymptotic properties

In order to establish our theory we exploit the important notion of *contiguity*, we refer to Chapter 6 of [37] for a general discussion and to [4] for an application to approximate statistical modelling of time series. Let \mathcal{F}_0^k and \mathcal{G}_0^k be the joint probability measures of the normalised maxima $(M_{m,i} - \beta(m))/a_{m,0}$, $i = 1, \dots, k$, and of k iid random variables with G_{γ_0} distribution, respectively. In particular, $\beta: \mathbb{N} \rightarrow \mathbb{R}$ is a suitable function satisfying condition (c) of Theorem 1 that generalises the centering through $b_{m,0}$. See also Section 3.3 (first paragraph) and Section 3.5 of Supplementary Material [31] for further comments and technical details. Let E_k be any measurable set sequence. Then, \mathcal{F}_0^k is said to be contiguous with respect to \mathcal{G}_0^k , in symbols $\mathcal{F}_0^k \triangleleft \mathcal{G}_0^k$, if $\mathcal{G}_0^k(E_k) = o(1)$ implies that $\mathcal{F}_0^k(E_k) = o(1)$, as $k \rightarrow \infty$.

Establishing the asymptotic behaviour of the numerator in the posterior distribution formula is essential to obtain posterior contraction rates and asymptotic normality. This is done in the standard setting where the statistical model is well specified, by showing the existence of tests for the true parameter, with type I and II error probabilities having a suitable decay rate (see, e.g., [20,37]). Thanks to the contiguity result in Theorem 1, this method can be adapted to the nonstandard setting where the misspecified model G_{γ_0} is considered in place of the unknown distribution F_0^m .

Theorem 1. *If the following conditions are valid together, they are sufficient conditions for $\mathcal{F}_0^k \triangleleft \mathcal{G}_0^k$:*

(a) V_0 is twice differentiable, the following function is regularly varying of order $\rho \leq 0$

$$tV_0''(t)/V_0'(t) - \gamma_0 + 1 =: A_0(t) \rightarrow 0, \quad t \rightarrow \infty; \quad (9)$$

(b) $\sqrt{k}A(m) \rightarrow \lambda \in \mathbb{R}$ as $k \rightarrow \infty$;

(c) $(\beta(m) - V_0(m))/(a(m)A_0(m)) \rightarrow \lambda' \in \mathbb{R}$ as $k \rightarrow \infty$ and, if $\lambda \neq 0$, there is an integer $m' \geq 1$ such that

$$\max_{m \geq m'} \|a(m)f_0^{(m)}(a(m) \cdot + \beta(m))/g_{\gamma_0}\|_{\infty} < \infty.$$

Condition (a) requires that $F_0^{\leftarrow}(\exp(-1/t))$ is a smooth quantile function and the rate function A_0 , defined as in (9) (see [11]), is essentially regularly varying. This condition is reasonably mild and is satisfied by standard models considered in the simulation study of Section 5, see Section 1.1 of Supplementary Material [31]. Note that, under assumption (a), the second order condition (5) is satisfied with rate functions $A(m) = A_0(m)$ as in (9) and $a(m) = mV_0'(m)$. For this reason, assumption (a) is referred to as second order von Mises-type condition, see also Theorem 2.3.12 in [10] for a similar, alternative formulation. In addition, as $k \rightarrow \infty$

$$\mathcal{H}(f_0^{(m)}, g_{\theta_0}) = O(A_0(m)),$$

which is a very important property to derive the asymptotic results for density estimation in Theorem 2 and for prediction in Section 3.3, see Section 3.5 of Supplementary Material [31]. Condition (b) is the standard assumption adopted in EVT to quantify the bias amount of estimators arising from misspecified extreme value models (e.g., [10]). Finally, condition (c) essentially requires that for a large block-size m the approximating GEV density does not vanish faster than the true unknown one near the endpoints. This condition, also used by [30] to study multivariate maxima, does not appear over restrictive, we have indeed verified that standard models considered in the simulation study of Section 5 satisfy it, see Section 1.2 of Supplementary Material [31].

We next specify simple conditions on our proposed data-dependent prior π_k with generic form in (7), on the basis of which we establish our asymptotic theory.

Condition 1. The densities π_{sh} , π_{loc} and π_{sc} satisfy the following conditions:

- (a) π_{sh} is a positive and continuous on $(\gamma_0 \pm \eta)$, for an $\eta > 0$.
- (b) there is $\eta \in (0, 1)$ and an integrable continuous function $u_{\text{sc}} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that
 - (b.1) $\inf_{x \in [1 \pm \eta]} \pi_{\text{sc}}(x) > 0$;
 - (b.2) $\sup_{t \in (1 \pm \eta)} \pi_{\text{sc}}(x/t) \leq u_{\text{sc}}(x)$, for all $x > 0$.
- (c) there is $\eta \in (0, 1)$ and an integrable continuous function $u_{\text{loc}} : \mathbb{R} \rightarrow \mathbb{R}_+$ such that
 - (c.1) $\inf_{x \in [-\eta, +\eta]} \pi_{\text{loc}}(x) > 0$;
 - (c.2) $\sup_{t_1 \in (1 \pm \eta), t_2 \in (-\eta, +\eta)} \pi_{\text{loc}}((x - t_2)/t_1) \leq u_{\text{loc}}(x)$, for all $x > 0$.

Furthermore, the estimators $\widehat{a}_{m,k}$ of $a_{m,0}$ and $\widehat{b}_{m,k}$ of $b_{m,0}$ are such that:

- (d) $\widehat{a}_{m,k}/a_{m,0} = 1 + o_p(1)$ and $(\widehat{b}_{m,k} - b_{m,0})/a_{m,0} = o_p(1)$, as $k \rightarrow \infty$.

Conditions (a)–(c) are satisfied by most of the usual probability density kernels. Prior densities of the shape parameter with bounded support (e.g. uniform) are also allowed, as long as the latter contains γ_0 as an interior point. Condition (d) is satisfied by classical estimators, such as the ML in Theorem 2 of [15] and the PWM in Theorem 2.3 of [17]. We are now ready to establish posterior contraction rates.

Theorem 2. Let X_1, \dots, X_n be iid random variables with distribution $F_0 \in \mathcal{D}(G_{\gamma_0})$, where $\gamma_0 > -1/2$. Let $M_{m,i}$, $i = 1, \dots, k$ be the corresponding BM. Work with a prior density that is specified as in (7) and complies with Condition 1. Work under conditions in Theorem 1(a)–(c). Let C_k be a sequence of positive real numbers satisfying $C_k \rightarrow \infty$ and $C_k = o(\sqrt{k})$ as $k \rightarrow \infty$, and set $\epsilon_k = C_k/\sqrt{k}$, $k = 1, 2, \dots$. Then, there exist constants $c_1 > 0$ and $c_2 > 0$, such that, with probability tending to 1 as $k \rightarrow \infty$:

- (a) $\Psi_k \left(\left\{ \theta \in \Theta : \left\| \left(\gamma - \gamma_0, \frac{b_m - b_{m,0}}{a_{m,0}}, \frac{a_m}{a_{m,0}} - 1 \right) \right\|_1 > \epsilon_k \right\} \right) \leq e^{-c_1 C_k^2}$;
- (b) $\Psi_k \left(\left\{ \theta \in \Theta : \mathcal{H}(g_\theta, f_0^{(m)}) > \epsilon_k \right\} \right) \leq e^{-c_2 C_k^2}$.

A main implication of results (a) and (b) in Theorem 2 is that the posterior distribution of θ , based on unnormalised BM, provides consistent estimation of the unknown parameter θ_0 and unknown true BM density $f_0^{(m)}$, cumulating its mass in a neighbourhood of those. See Proposition 2 in Section 4.3 of Supplementary Material [31] for the explicit result on the posterior consistency. The result at point (a) gives a refined Bayesian analog of the MLE consistency result in [15]. It is due to the fact that with high probability the posterior of θ , based on normalised BM, concentrates most of his mass on a ball centered at θ_0 , whose radius ϵ_k decreases with k , while out of the latter the residual mass decreases (exponentially fast) as k increases see Section 8.2. From a practical view point, this allows to check

whether accurate posterior-based inference on the original parameter θ_0 is achievable with finite samples. In Section 5 we indeed assess the degree of posterior concentration via simulation showing that with several standard statistical models accurate inference can be obtained using moderate dimensions for m and k . We recall that the true density $f_0^{(m)}$ of the unnormalised BM becomes (topologically) undistinguishable from g_{θ_0} as m increases. Since the posterior distribution concentrates on a set of parameters θ such that g_θ is close to g_{θ_0} (in Hellinger metric), then this allows for the concentration result in point (b), see Section 4.1.1 in Supplementary Material [31] and Section 8.2. Its relevant theoretical implications for statistical prediction are highlighted in Section 3.3. Theorem 2 is also important from a technical view point because is preparatory for establishing posterior asymptotic normality. In the sequel, the d -variate normal distribution is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix, respectively. When $d = 1$, we write $\mathcal{N}(\mu, \sigma^2)$. We denote by $\mathcal{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ the probability measure of such a distribution. Moreover, the Fisher information matrix of the distribution $G_{\bar{\theta}_0}$ is denoted by \mathbf{I}_0 , see Section 3.2 of Supplementary Material [31] and [32] for details. We recall that S_{k, θ_0} is the score vector process given in (6).

Theorem 3. *Assume that the conditions of Theorem 2 are satisfied. Then, as $k \rightarrow \infty$*

$$\sup_{B \in \mathcal{B}(\mathbb{R}^3)} |\Psi_k(\{\theta \in \Theta : \sqrt{k}(r(\theta) - \bar{\theta}_0) \in B\}) - \mathcal{N}(B; \mathbf{I}_0^{-1} S_{k, \theta_0}, \mathbf{I}_0^{-1})| = o_p(1).$$

This result establishes that the posterior distribution of $\bar{\theta}$, based on normalised BM, is asymptotically close to a normal distribution centred at $\bar{\theta}_0 + k^{-1/2} \mathbf{I}_0^{-1} S_{k, \theta_0}$ and with covariance matrix $k^{-1} \mathbf{I}_0^{-1}$, as k increases. In particular, $\mathbf{I}_0^{-1} S_{k, \theta_0}$ is also asymptotically normally distributed and behaves like the normalised MLE, i.e. $\sqrt{k}(r(\hat{\theta}_k) - \bar{\theta}_0) \sim \mathcal{N}(\lambda \mathbf{I}_0^{-1} \mathbf{b}, \mathbf{I}_0^{-1})$ as $k \rightarrow \infty$, where \mathbf{b} is a bias vector term, see Section 3.2 of Supplementary Material [31] and Theorems 2.1-2.2 in [16] for details. Accordingly, $k^{-1/2} \mathbf{I}_0^{-1} S_{k, \theta_0}$ is asymptotically negligible as $k \rightarrow \infty$ and the posterior is therefore asymptotically centered at $\bar{\theta}_0$, consistently with the previous findings. Since the reparametrization $r(\theta)$ is obtained via the simple linear transformation in (8), the main practical implication of Theorem 3 is that also the posterior distribution Ψ_k based on nonnormalised BM asymptotically resembles a normal distribution, say $\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, with $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ that are suitable linear transformations of the score vector and of the inverse Fisher information matrix, respectively. In turn, also the univariate marginal distribution $\Psi_{k, j}$ of the individual parameter $\theta^{(j)}$, i.e. the j -th component of θ for $j = 1, 2, 3$, obtained from Ψ_k , is asymptotically normal. We can show that for any $\alpha \in (0, 1)$, the asymmetric $(1 - \alpha)100\%$ -credible interval given by the $(1 - \alpha)$ -quantile of $\Psi_{k, j}$, $Q_{\Psi_{k, j}}(1 - \alpha/2)$ and $Q_{\Psi_{k, j}}(\alpha/2)$, respectively, with $j = 1, 2, 3$, has coverage probability that asymptotically agrees with the nominal level $1 - \alpha$, whenever $\sqrt{k} A_0(m) = o(1)$, i.e. $\lambda = 0$ and so $\mathbf{I}_0^{-1} S_{k, \theta_0} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_0^{-1})$.

Corollary 1. *For any $\alpha \in (0, 1)$, let $I_{k, \alpha}^A = [Q_{\Psi_{k, j}}(1 - \alpha/2); Q_{\Psi_{k, j}}(\alpha/2)]$, for $j \in \{1, 2, 3\}$. Under the assumptions of Theorem 3, if $\lambda = 0$, as $k \rightarrow \infty$*

$$\mathbb{P}\left(\{\theta_0^{(j)} \in I_{k, \alpha}^A\}\right) = 1 - \alpha + o(1), \quad j = 1, 2, 3.$$

Theorem 3 can also be exploited to draw practical guidelines for constructing (approximate) credible Highest Posterior Density (HPD) regions. The derivation of HPD regions is a complex task when the expression of the posterior distribution is not known in closed-form. This is especially true for multidimensional parameters, as pointed out, e.g., on page 262 of [34] and in the references therein. Since $\Psi_{k, j}$, $j = 1, 2, 3$, is asymptotically similar to a normal distribution, say $\mathcal{N}(\mu_{k, j}, \sigma_{k, j}^2)$, then the

latter can be used to define HPD intervals with $(1-\alpha)$ -credible level. Note that, $\mu_{k,j}$ and $\sigma_{k,j}^2$ depend on the true unknown parameter $\theta_0^{(j)}$ and therefore they cannot be used for interval estimation in practice. We can replace them by the posterior mean $\widehat{\mu}_{k,j}$ and variance $\widehat{\sigma}_{k,j}^2$ and then use the symmetric interval $I_{k,\alpha}^S = [\widehat{\mu}_{k,j} - z_{\alpha/2}\widehat{\sigma}_{k,j}, \widehat{\mu}_{k,j} + z_{\alpha/2}\widehat{\sigma}_{k,j}]$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$ -quantile of an univariate standard normal distribution. The posterior asymptotic normality result is particularly useful when the interest is in computing a HPD region for the true unknown parameter vector θ_0 . Again, since Ψ_k can be approximated by $\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, then a HPD region with $(1-\alpha)100\%$ -credibility for θ_0 is given by the random symmetric ellipsoid

$$E_{k,\alpha} = \widehat{\boldsymbol{\mu}}_k + \widehat{\boldsymbol{\Sigma}}_k^{1/2} \mathbf{B}_2\left(\mathbf{0}, \sqrt{\chi_{3,1-\alpha}^2}\right), \quad (10)$$

where $\mathbf{B}_p(\mathbf{0}, r)$ is the closed L_p -norm-ball in \mathbb{R}^3 whose center is the origin $\mathbf{0}$ and the radius is r , $\chi_{3,1-\alpha}^2$ is the $(1-\alpha)$ -quantile of the chi-squared distribution with 3 degrees of freedom and $\widehat{\boldsymbol{\mu}}_k$ and $\widehat{\boldsymbol{\Sigma}}_k$ are the posterior mean and covariance, since $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are not available.

3.2. Return levels: posterior asymptotic properties

Estimation of extreme events through the BM approach can be achieved estimating the so-called return level corresponding to a prefixed return period T (T -return level for brevity). For instance, suppose X_1, \dots, X_m describe the random behaviour of an observational phenomenon, and the sequence $M_{m,i}$, $i = 1, \dots, k$ with $m = 366$ is hence representative of yearly maxima. Then, the return level x_T is the value that is expected to be exceeded by the annual maximum on average once every T years. From a probabilistic view point, it is the $(1-p)$ -quantile of the distribution F_0^m , where $p = 1/T$ is a small exceedance probability such that $p = 1 - F_0^m(x_T)$. The distribution F_0 is unknown in real applications, therefore, for any $p \in (0, 1)$, statistical inference on the unknown quantile $Q_{F_0^m}(p)$ can be based on the approximation given by the GEV quantile $Q_{G_{\theta_0}}(p)$, where in particular

$$Q_{G_{\theta_0}}(p) = b_m + a_m \frac{(-\log(1-p))^{-\gamma} - 1}{\gamma}. \quad (11)$$

For any $p \in (0, 1)$, the map $q : \Theta \rightarrow \mathbb{R} : \boldsymbol{\theta} \mapsto Q_{G_{\boldsymbol{\theta}}}(p)$ is continuous. Thus, the posterior distribution Ψ_k on $\boldsymbol{\theta}$ induces a posterior distribution on the GEV quantile q , which is given by $\Omega_k := \Psi_k \circ q^{-1}$. Next, we provide a series of asymptotic results on the posterior distribution Ω_k , which guarantee the reliability of quantile-based inference.

Theorem 4. *Assume that the conditions of Theorem 2 are satisfied. Let C_k be a sequence of positive real numbers satisfying $C_k \rightarrow \infty$ and $C_k = o(\sqrt{k})$ as $k \rightarrow \infty$, and set $\epsilon'_k = a_{m,0}C_k/(\sqrt{k}|b_{m,0}|)$, $k = 1, 2, \dots$. Then, for every $p < 1 - e^{-1}$, there is a constant $c > 0$ such that, with probability tending to 1 as $k \rightarrow \infty$:*

$$\Omega_k \left(q \in \mathbb{R} : |q/Q_{F_0^m}(p) - 1| > \epsilon'_k \right) \leq e^{-cC_k^2}.$$

The result in Theorem 4 implies that for any $p \in (0, 1)$, the posterior distribution Ω_k , based on unnormalised maxima, is consistent and allows therefore for increasingly accurate inference on the true unknown quantile $Q_{F_0^m}(p)$, as k increases. Unlike the result in Theorem 2, in this one the contraction rate of Ω_k depends on the tail heaviness of F_0 . The lighter is the tail of F_0 , the narrower is the neighbourhood of $Q_{F_0^m}(p)$ on which Ω_k concentrates, since $a_{m,0}/b_{m,0} = \max(\gamma_0, 0) + o(1)$ as

$k \rightarrow \infty$. Hence, in practice, with short- or light-tailed distributions accurate estimate of BM quantiles are achieved with smaller sample sizes than in the heavy-tailed case. The next result shows that the quantile posterior distribution is also asymptotically normal.

Theorem 5. *Assume that the conditions of Theorem 2 are satisfied. Assume that $p < 1 - e^{-1}$ and set $Q_k = |b_{m,0}| \sqrt{k} / a_{m,0}$. Then, there is a nonnull vector of constants $\mathbf{D}_0 \in \mathbb{R}^3$ and a constant $v \in \mathbb{R}$ such that, as $k \rightarrow \infty$*

$$\sup_{B \in \mathcal{B}(\mathbb{R})} \left| \Omega_k \left(q \in \mathbb{R} : Q_k \left(\frac{q}{Q_{F_0^m}(p)} - 1 \right) \in B \right) - \mathcal{N}(B; \mathbf{D}_0^\top \mathbf{I}_0^{-1} S_{k,\theta_0} + v, \mathbf{D}_0^\top \mathbf{I}_0^{-1} \mathbf{D}_0) \right| = o_p(1).$$

The exact expression of \mathbf{D}_0 and v and all the details are provided in Section 8.6. The main practical implication of Theorem 5 is that the posterior distribution Ω_k is asymptotically close to a normal distribution, say $\mathcal{N}(\mu_k, \sigma_k^2)$. Once more, a benefit of posterior asymptotic normality is that it allows to understand whether $(1 - \alpha)$ -credibility intervals, for $\alpha \in (0, 1)$, defined using Ω_k have coverage probability with asymptotic nominal level $(1 - \alpha)$, in the frequentist sense. In this regard, when $\lambda = 0$ then we have $v = 0$ and $\mathbf{D}_0^\top \mathbf{I}_0^{-1} S_{k,\theta_0} \xrightarrow{d} \mathcal{N}(0, \mathbf{D}_0^\top \mathbf{I}_0^{-1} \mathbf{D}_0)$ as $k \rightarrow \infty$. We show that the coverage probability of asymmetric $(1 - \alpha)$ 100%-credible intervals defined by the $1 - \alpha/2$ and $\alpha/2$ quantiles of Ω_k , $Q_{\Omega_k}(1 - \alpha/2)$ and $Q_{\Omega_k}(\alpha/2)$, respectively, asymptotically agrees with the nominal level $1 - \alpha$.

Corollary 2. *For any $\alpha \in (0, 1)$, let $I_{k,\alpha}^A = [Q_{\Omega_k}(1 - \alpha/2); Q_{\Omega_k}(\alpha/2)]$. Under the assumptions of Theorem 5, if $\lambda = 0$, as $k \rightarrow \infty$*

$$\mathbb{P} \left(\{Q_{F_0^m}(p) \in I_{k,\alpha}^A\} \right) = 1 - \alpha + o(1).$$

Another benefit of the approximation $\mathcal{N}(\mu_k, \sigma_k^2)$, for large k , is that we can derive for a small p an approximate HPD set with $(1 - \alpha)$ -credibility for $Q_{F_0^m}(p)$. This is given, likewise Section 3.1, by the symmetric interval $I_{k,\alpha}^S = [\hat{\mu}_k - z_{\alpha/2} \hat{\sigma}_k; \hat{\mu}_k + z_{\alpha/2} \hat{\sigma}_k]$, where $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the posterior mean and standard deviation of Ω_k .

Finally we recall that, for a small $p \in (0, 1)$, the extreme quantile $Q_{F_0}(p)$ can be approximated by the right-hand formula in (4), which is equal to that in (11) with m in front of the logarithm. The posterior distribution Ψ_k induces a posterior distribution on the approximate extreme quantile $Q_{G_{\theta_0}}(p_m)$, which is given by $\tilde{\Omega}_k := \Psi_k \circ \tilde{q}^{-1}$, where $\tilde{q} : \Theta \rightarrow \mathbb{R} : \theta \mapsto Q_{G_\theta}(p_m)$, with p_m as in (4). Thus, as done above for return levels, we define an asymmetric $(1 - \alpha)$ 100%-credible interval $I_{k,\alpha}^A = [Q_{\tilde{\Omega}_k}(1 - \alpha/2); Q_{\tilde{\Omega}_k}(\alpha/2)]$, where $Q_{\tilde{\Omega}_k}(p)$ is the $(1 - p)$ -quantile of $\tilde{\Omega}_k$, and an approximate HPD set with $(1 - \alpha)$ -credibility by the symmetric interval $I_{k,\alpha}^S = [\hat{\mu}_{\tilde{q};k} - z_{\alpha/2} \hat{\sigma}_{\tilde{q};k}; \hat{\mu}_{\tilde{q};k} + z_{\alpha/2} \hat{\sigma}_{\tilde{q};k}]$, where $\hat{\mu}_{\tilde{q};k}$ and $\hat{\sigma}_{\tilde{q};k}$ are the mean and standard deviation of $\tilde{\Omega}_k$.

3.3. Posterior predictive distribution

One unquestionable strength of the Bayesian paradigm is to be able to carry out prediction through the use of the posterior predictive distribution, which incorporates uncertainty on the model. In a BM approach, given a sequence of iid maxima $M_{m,i}$, $i = 1, \dots, k$, the distribution of the maximum M_{m^*} of a block of $m^* \geq m$ future iid unobservable random variables can be described by the posterior predictive distribution

$$\widehat{G}_k^{(m^*)}(x) = \int_{\Theta} G_\theta^{m^*/m}(x) d\Psi_k(\theta), \quad \forall x \in \mathbb{R}, \quad (12)$$

where G_θ approximates the unknown distribution F_0^m . Its posterior predictive density is

$$\widehat{g}_k^{(m^*)}(x) = \int_{\Theta} (\partial/\partial x) G_\theta^{m^*/m}(x) d\Psi_k(\theta), \quad \forall x \in \mathbb{R}.$$

For $p \in (0, 1)$, let $Q_{\widehat{G}_k^{(m^*)}}(p)$ be the $(1-p)$ -quantile of $\widehat{G}_k^{(m^*)}$. Consider now the special case $m^* = m$ in (12). The computation of $Q_{\widehat{G}_k^{(m)}}(1/T)$ provides an approach to point forecasting of the block maximum through the T -return level (see, e.g., [6]). The predictive distribution (12) allows for a more comprehensive probabilistic forecasting approach. See, e.g., [26] for Bayesian approaches to probabilistic forecasting that rely on the posterior predictive distribution. To illustrate this point, consider a large time window W (e.g. 50 years) and set $m^* = W \cdot m'$, where m' is the block-size corresponding to a time unit of W (e.g. $m' = 366$) and may differ from m . Beyond point forecasts (through predictive mean, quantiles, etc.), the predictive distribution in formula (12) allows to formulate richer prediction techniques than the point forecasting of return levels, and accounting then for far more extreme events. An example is given by predictive regions to which the maximum level belongs to in any particular time window W , with a fixed probability. Given the vital importance of predicting extreme events, the next result establishes key inferential properties of the posterior predictive distribution and density, not already discussed in the earlier literature, as far as we know.

Corollary 3. *Work under the assumptions and with the notation in Theorem 2. Let $K \subset (0, 1)$ be any compact set and, for any $\alpha \in (0, 1)$, let $R_k \in \mathcal{B}(\mathbb{R})$ satisfy $G_k^{(m^*)}(R_k) = 1 - \alpha$. Then, if $C_k/\sqrt{\log k} \rightarrow \infty$ and $m^* \equiv m_k^*$ satisfies $m^*/m \rightarrow t$, with $t > 0$, as $k \rightarrow \infty$:*

- (a) $\mathcal{H}(\widehat{g}_k^{(m^*)}, f_0^{(m^*)}) = O_p(\epsilon_k)$;
- (b) $\sup_{B \in \mathcal{B}(\mathbb{R})} |\widehat{G}_k^{(m^*)}(B) - F_0^{m^*}(B)| = O_p(\epsilon_k)$;
- (c) $F_0^{m^*}(R_k) = 1 - \alpha + O_p(\epsilon_k)$
- (d) $\sup_{p \in K} \left| \frac{Q_{\widehat{G}_k^{(m^*)}}(p)}{Q_{F_0^{m^*}}(p)} - 1 \right| = O_p\left(\epsilon_k \frac{a_{m^*,0}}{|b_{m^*,0}|}\right)$.

Points (a)–(b) guarantee that the predictive density $\widehat{g}_k^{(m^*)}$ and distribution $\widehat{G}_k^{(m^*)}$ are good approximations of the density $f_0^{(m^*)}$ and distribution $F_0^{m^*}$ of a future block-maximum M_{m^*} , for large enough m^* , m and k , providing therefore a reliable basis for the probabilistic forecasting. A predictive region R_k can be obtained by means of the quantiles of the predictive distribution in (12) or the highest predictive density regions. Point (c) establishes that predictive regions based on the predictive distribution are reliable for forecasting future BM. Finally, point (d) also guarantees that the return level functionals $Q_{F_0^{m^*}}(1/T)$, with T ranging in a certain set and m^* possibly different from m , can be accurately estimated. In particular, the predictive-based trajectory $\{Q_{\widehat{G}_k^{(m^*)}}(1/T), T_a, \leq T \leq T_b\}$ is a reliable proxy for the unknown return level path $\{Q_{F_0^{m^*}}(1/T), T_a, \leq T \leq T_b\}$ relative to future return period interval $[T_a, T_b]$.

4. Computational aspects

We resort to a MCMC computational method for the empirical calculation of the posterior distribution Ψ_k , due to the lack of its explicit formula. In particular, to draw samples from the posterior distribution

Algorithm 1: Adaptive Random-Walk Metropolis-Hastings MCMC scheme

```

1 Initialize: Set  $\theta^{(0)}$ ,  $\kappa^{(0)}$  and  $\Sigma^{(0)}$ ;
2 for  $j = 0$  to  $M$  do
3   Draw proposal  $\theta' \sim \phi_3(\theta^{(j)}, \kappa^{(j)}\Sigma^{(j)})$ ;
4   Compute acceptance probability  $\eta = \min\left(\frac{L_k(\theta'|\mathbf{y})\pi_k(\theta')}{L_k(\theta^{(j)}|\mathbf{y})\pi_k(\theta^{(j)})}, 1\right)$ ;
5   Draw  $U_1 \sim \mathcal{U}(0, 1)$ . If  $\eta > U_1$  then set  $\theta^{(j+1)} = \theta'$  else set  $\theta_1^{(j+1)} = \theta^{(j)}$ ;
6   Update  $\Sigma^{(j)}$  according to (13);
7   Update  $\kappa^{(j)}$  according to (14);

```

Ψ_k we use the Adaptive (Gaussian) random-walk Metropolis-Hastings (AMH) scheme discussed in [18], which is a special case of the AMH class of algorithms introduced in [21].

The main elements are the likelihood function defined in Section 2.2 that is here denoted by $L_k(\theta|\mathbf{y})$, where the data sample of maxima is denoted by $\mathbf{y} = (y_1, \dots, y_k)$ to simplify the notation, and the data-dependent prior $\pi_k(\theta)$ in (7). A prior density π_k , satisfying Condition 1, is not difficult to specify. A simple example is indeed easily given by taking in (7) the product of the following densities: $\pi_{\text{sh}}(\gamma) = (1 - T_1(-1))^{-1} t_1(\gamma) \mathbb{I}(-1 < \gamma < \infty)$, where t_ν and T_ν are the Student- t density and distribution function, respectively, with ν degrees of freedom; $\pi_{\text{loc}}((b_m - \widehat{b}_{m,k})/\widehat{a}_{m,k})/\widehat{a}_{m,k} = \phi((b_m - \widehat{b}_{m,k})/\widehat{a}_{m,k})/\widehat{a}_{m,k}$, where ϕ is standard normal density and $\widehat{a}_{m,k}$ and $\widehat{b}_{m,k}$ are the ML estimators of $a_{m,0}$ and $b_{m,0}$; $\pi_{\text{sc}}(a_m/\widehat{a}_{m,k})/\widehat{a}_{m,k} = \xi(a_m; 1, \widehat{a}_{m,k})/\widehat{a}_{m,k}$, where $\xi(x; \varphi, \zeta)$, $x > 0$ is a Gamma density with shape $\varphi > 0$ and scale $\zeta > 0$. Other simple choices can be readily obtained, but we hereafter focus on this basic option.

A short summary of the algorithm is as follows. The current state of the chain $\theta^{(j)}$ at time j is potentially updated by the proposal $\theta' \sim h(\theta|\theta^{(j)}) = \phi_3(\theta^{(j)}, \kappa^{(j)}\Sigma^{(j)})$, where $\phi_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a d -dimensional Gaussian density function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Following Haario, Saksman and Tamminen [21], the proposal covariance matrix $\Sigma^{(j)}$ is specified as

$$\Sigma^{(j+1)} = \begin{cases} (1 + [\kappa^{(j)}]^2/j)\mathbf{J}_3, & j \leq 100, \\ \frac{1}{j-1} \sum_{s=1}^j (\theta^{(s)} - \tilde{\theta}^{(j)})(\theta^{(s)} - \tilde{\theta}^{(j)})^\top + ([\kappa^{(j)}]^2/j)\mathbf{J}_3, & j > 100, \end{cases} \quad (13)$$

where \mathbf{J}_d is the d -dimensional identity matrix, $\tilde{\theta}^{(j)} = j^{-1}(\theta^{(1)} + \dots + \theta^{(j)})$, and $\kappa^{(j)} > 0$ is a scaling parameter that affects the acceptance rate of proposed parameter values. According to Garthwaite, Fan and Sisson [18] we adaptively update κ using a Robbins-Monro process so that

$$\log \kappa^{(j+1)} = \log \kappa^{(j)} + a(\eta^{(j)} - \eta^*), \quad (14)$$

where $a = (2\pi)^{1/2} \exp(\zeta_0^2/2)/(2\zeta_0)$ is a steplength constant, $\zeta_0 = -1/\Phi(\pi^*/2)$, and where Φ is the univariate standard Gaussian distribution function. The parameter η^* is the desired overall sampler acceptance probability, which we specified as $\eta^* = 0.234$, according to Gelman, Gilks and Roberts [19]. Given the symmetry of the proposal, i.e. $h(\theta'|\theta) = h(\theta|\theta')$, the acceptance probability of the update $\theta^{(j+1)} = \theta'$ reduces to

$$\eta^{(j)} = \min\left(\frac{L_k(\theta'|\mathbf{y})\pi_k(\theta')}{L_k(\theta^{(j)}|\mathbf{y})\pi_k(\theta^{(j)})}, 1\right),$$

otherwise we set $\theta^{(j+1)} = \theta^{(j)}$ with rejection probability $1 - \eta^{(j)}$. This algorithm is summarised in Algorithm 1. Finally, let $\theta_1^*, \dots, \theta_N^*$ be a sample generated from the posterior Ψ_k , then, for any $p \in (0, 1)$, a sample q_1^*, \dots, q_N^* from quantile posterior Ω_k is obtained by exploiting the relation between θ and q given in (11). Moreover an approximation of the posterior predictive distribution $\widehat{G}_k^{(m^*)}$ can be obtained via simulation by considering the Monte Carlo counterpart of formula (12), i.e. $\widehat{G}_k^{(m^*)} \approx N^{-1} \sum_{i=1}^N G_{\theta_i^*}^{m^*/m}(x)$.

5. Simulation study

In this section, we assess for finite sample sizes the behaviour of the posterior distributions Ψ_k , Ω_k and $\tilde{\Omega}_k$, computed via the MCMC method described in Section 4, and the performance of the resulting inference. We consider nine distributions: three in the domain of attraction of the Fréchet, Gumbel and reverse-Weibull families, respectively. To save space, we focus the discussion on the Half-Cauchy, Gamma with both shape and rate parameters equal to 2 and Power-law distribution $F(x) = 1 - K(x^* - x)^\alpha$, where $x^* = 1$, $\alpha = 3$ and $K = 1/9$ are the end-point of the distribution, the shape parameter and a positive constant, respectively. With these models the tail index is $\gamma = 1, 0, -1/3$. The simulation results for the other six models are discussed in Section 2.2. of Supplementary Material [31].

Our theory in Section 3 holds as long as the conditions in Theorem 1 are satisfied, in particular, as long as its condition (b) is fulfilled. The function $|A_0|$ is regularly varying of order $\rho \leq 0$ and it asymptotically behaves as Cm^ρ for most of the distributions in the Fréchet domain of attraction, where $C \neq 0$ and $\rho < 0$. With the unit-Fréchet, standard-Pareto and Half-Cauchy distributions $\rho = -\infty, -1, -2$ and condition (b) is satisfied setting $k = \lceil m^{-2\rho} (\log(m))^{-\alpha} \rceil$ for some $\alpha \in \mathbb{R}$, where $\lceil x \rceil$ denotes the closest integer number to x , for any $x \in \mathbb{R}$. However, with the other domains of attraction different formulas for k may be needed. To avoid using different ways to select k depending on the distribution taken into account, we simply put $k = \lceil m/\sqrt{\log(m)} \rceil$ for all the models and doing so condition (b) is always satisfied except for the Gamma model. In this way, we are able to verify whether the posterior distributions behaviour and the corresponding inference agree with theory of Section 3 but also if the results are still satisfactory when the theoretical conditions are violated. We consider the block-sizes $m = 40, 60, 109, 234$ which entail the numbers of blocks $k = 20, 30, 50, 100$ and sample sizes $n = km = 800, 1800, 5450, 23400$. For each model we perform the following steps. In the first one, we simulate n observations and derive on their basis k maxima. We run the AMH algorithm described in Section 4 generating 50,000 values of which we retain $N = 20,000$ realizations as a sample from the posterior Ψ_k , after a burn-in period of 30,000. With this setting, the generation of a posterior sample is fast, it takes 42 seconds using a MacBook Pro with M1 CPU and 8 GB of RAM. A diagnostic study to verify the convergence of the produced Markov chains is reported in Section 2.1 of Supplementary Material [31]. In the second step, we compute the following summaries. The uniform-norms

$$\tilde{\gamma}_k = \|\boldsymbol{\gamma}_k^* - \gamma_0\|_\infty, \quad \tilde{b}_k = \|(\mathbf{b}_k^* - b_{m,0})/a_{m,0}\|_\infty, \quad \tilde{a}_k = \|(\mathbf{a}_k^*/a_{m,0} - 1)\|_\infty, \quad (15)$$

are calculated first, where $\boldsymbol{\gamma}_k^* = (\gamma_{k,1}^*, \dots, \gamma_{k,N}^*)^\top$, $\mathbf{b}_k^* = (b_{k,1}^*, \dots, b_{k,N}^*)^\top$ and $\mathbf{a}_k^* = (a_{k,1}^*, \dots, a_{k,N}^*)^\top$ are the values sampled from Ψ_k , $(\gamma_0, b_{m,0}, a_{m,0})$ are the true parameters with $a_{m,0} = mV_0'(m)$, $b_{m,0} = V_0(m)$. We compute the proportion of times that the Manhattan-norm of posterior draw exceeds the radius ϵ_k , i.e.

$$\tilde{p}_N = \frac{1}{N} \sum_{i=1}^N \mathbb{I} \left(\left\| \left(\gamma_{k,i}^* - \gamma_0, \frac{b_{k,i}^* - b_{m,0}}{a_{m,0}}, \frac{a_{k,i}^*}{a_{m,0}} - 1 \right) \right\|_1 > \epsilon_k \right), \quad (16)$$

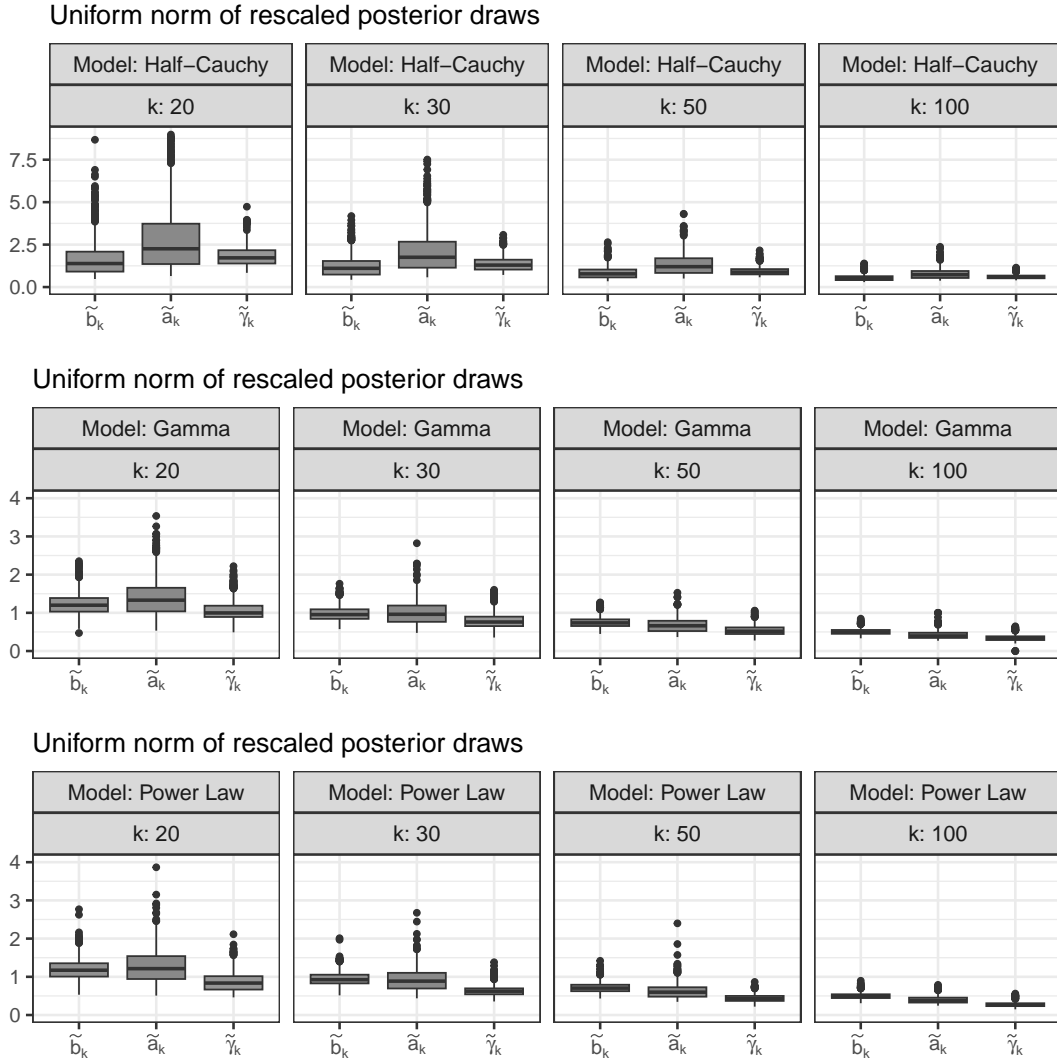


Figure 1: The box-plots report the Monte Carlo distribution of the statistics (uniform-norms of the rescaled posterior draws) defined in formula (15) for the Half-Cauchy, Gamma and Power-law distributions and for increasing value of k .

where $\epsilon_k = C_k/\sqrt{k}$ and $C_k = o(\sqrt{k})$ according to Theorem 2. We set $C_k = \log^2(k)$ obtaining $\epsilon_k \approx 2$ for all the k values (see Table 1), which is a fairly small radius for all the considered models. In the third step, we compute: asymmetric- and symmetric-95%-credibility intervals for γ_0 , $a_{m,0}$ and $b_{m,0}$ (see Section 3.1); symmetric-95%-credibility region for θ_0 (see (10)); asymmetric- and symmetric-95%-credibility intervals for the return level $Q_{F_0^m}(1/100)$ (see Section 3.2) and finally asymmetric- and symmetric-95%-credibility intervals for the extreme quantile $Q_{F_0}(0.001)$ (see Section 3.2).

We repeat these three steps $M = 1000$ times and with the obtained results we assess the following concentration properties of posterior distributions and coverage properties of posterior credible sets. A

k	C_k	ϵ_k	$R(k)$	Half-Cauchy			Gamma			Power-law		
				$b_{m,0}$	$a_{m,0}$	P_k	$b_{m,0}$	$a_{m,0}$	P_k	$b_{m,0}$	$a_{m,0}$	P_k
20	8.97	2.01	0.447	25.8	25.5	94.0	2.8	0.58	99.9	4.4	0.20	100
30	11.57	2.11	0.262	38.5	38.2	95.4	3.0	0.58	100	4.5	0.18	100
50	15.30	2.16	0.096	69.7	69.4	96.6	3.4	0.57	100	4.6	0.14	100
100	21.21	2.12	0.011	149.3	149.0	98.3	4.4	0.56	100	4.8	0.08	100

Table 1. Behaviour of the empirical posterior distribution of θ for increasing k . The seventh, tenth and thirteenth columns report the percentage in (17) computed with the Student- t , Gamma and Power-law models and different values of k and where $R(k) = e^{-0.01C_k^2}$.

sufficient condition for the posterior distribution to be consistent is that the theoretical counterpart of the summaries in (15) converges to zero. Figure 1 reports via box-plots the Monte Carlo distribution of such summaries obtained with different models, along the rows, and different k values, along the columns. In all the cases the right end point of the distribution is close to zero and the spread of the distribution is small already with $k = 20$ (although some outliers are present). As k increases, the range of the distributions shrinks considerably, becoming a very small interval in proximity to zero in the case $k = 100$. This suggests that the posterior concentration property theoretically envisaged by consistency seems to hold in practice. Similar results are obtained also with other models, see Section 2.2 of Supplementary Material [31]. According to Theorem 2, the posterior distribution places out of $B_1(\mathbf{0}, \epsilon_k)$ (the L_1 -norm-ball with centre $\mathbf{0}$ and radius ϵ_k) an amount of mass that goes to zero at the rate $R(k) := e^{-cC_k^2}$, for a certain $c > 0$. The posterior mass outside of $B_1(\mathbf{0}, \epsilon_k)$ is smaller than $R(k)$, with probability tending one. For practical purposes, in order to have an insight on the concentration speed of Ψ_k around θ_0 , we compute the percentage of times that the proportion defined in (16) is smaller than $R(k)$, where for the latter term we have set $c = 0.01$, namely

$$P_k = \frac{1}{M} \sum_{i=1}^M \mathbb{I}(\tilde{p}_{N,i} < R(k)) \cdot 100\%, \quad (17)$$

where $\tilde{p}_{N,i}$ is the i th Monte Carlo realisation of the proportion in (16). Table 1 collects the results. Considering the Half-Cauchy result, with $k = 20$, the posterior distribution places more than approximately 45% of its mass out of $B_1(\mathbf{0}, 2)$ only 6% of the time, and so on decreasing up to the point that, in the case $k = 100$, only 1.7% of the times more than 1% of the posterior mass lies out of $B_1(\mathbf{0}, 2)$. Even better results are obtained with the Gamma and Power-law models. Similar results are obtained with the other six models, see Section 2.2 of Supplementary Material [31]. These results also suggest that the posterior distribution concentrates around θ_0 already with $k = 20$ and it rapidly increases its concentration degree as k increases. From the practical point of view, accurate posterior-based inference is achievable with inexpensive values of m and k .

Finally we study the Monte Carlo coverage probability of the credible regions and intervals discussed in Sections 3.1 and 3.2. The results for the different models and dimensions of k and m are reported along the first two rows of each section between the dashed horizontal lines of Table 2. In the third row, the results obtained with frequentist MLE-based confidence intervals and regions are also reported in italics. Note that the confidence intervals for the return level and extreme quantile are obtained by applying a type of delta method. When the column Type reports the letter S and A the coverage probabilities of symmetric intervals and regions and asymmetric intervals are reported.

Overall, all the coverage probabilities are close to the 95% nominal level already with $k = 20$ and then they get even closer as k increases. This finding is consistent with the previous outcome, as expected.

Model	k	m	Type	Coverage probability					
				γ_0	$b_{m,0}$	$a_{m,0}$	θ_0	$Q_{F_0^m}(0.01)$	$Q_{F_0}(0.001)$
Half-Cauchy	20	40	S	94.6	93.7	93.7	92.4	97.3	96.4
			A	94.9	92.1	91.8	–	93.7	93.4
			S	89.6	88.1	87.7	45.4	82.3	83.4
	30	60	S	93.7	95.9	95.4	93.0	95.5	93.5
			A	94.4	94.1	93.3	–	94.0	93.8
			S	90.8	91.2	88.3	49.1	81.8	83.3
	50	109	S	94.5	95.4	95.3	94.2	95.4	94.3
			A	94.8	95.0	93.8	–	94.2	94.0
			S	91.9	92.9	91.6	52.7	84.0	85.1
	100	234	S	94.9	94.3	95.7	94.7	95.1	96.0
			A	95.1	94.2	94.4	–	94.9	94.2
			S	94.4	95.6	94.8	54.6	87.0	90.9
Gamma	20	40	S	97.0	93.8	96.6	93.1	96.5	96.2
			A	96.4	94.5	95.9	–	95.3	93.8
			S	80.7	91.2	88.6	67.3	84.0	90.4
	30	60	S	96.3	94.8	96.2	93.7	96.3	96.8
			A	96.5	94.9	94.9	–	94.4	93.7
			S	83.5	91.9	90.7	73.7	89.6	94.8
	50	109	S	95.3	95.9	95.5	93.9	96.0	96.3
			A	95.0	96.2	95.4	–	94.2	94.4
			S	86.8	93.4	93.0	78.3	92.6	97.8
	100	234	S	94.9	95.7	95.1	94.3	94.9	94.9
			A	94.4	94.6	95.0	–	94.7	94.8
			S	91.3	93.3	92.6	81.9	95.0	98.1
Power-law	20	40	S	95.2	94.0	95.2	93.0	97.3	97.5
			A	94.3	94.6	95.3	–	94.4	94.1
			S	67.8	91.0	88.5	42.4	90.7	94.7
	30	60	S	95.5	94.2	96.1	93.8	96.6	96.3
			A	95.3	94.3	95.2	–	94.2	94.8
			S	74.4	92.5	91.2	45.8	94.6	97.1
	50	109	S	95.6	94.3	95.1	94.1	95.6	95.2
			A	95.8	94.5	95.4	–	95.6	95.7
			S	81.1	93.8	92.6	55.7	98.4	99.5
	100	234	S	94.4	94.6	94.7	94.5	95.4	94.3
			A	94.6	95.1	94.7	–	94.3	94.4
			S	84.9	95.0	92.6	55.5	99.3	98.7

Table 2. Coverage probability of Symmetric (S) and Asymmetric (A) 95%-credible intervals for γ_0 , $b_{m,0}$, $a_{m,0}$, $Q_{F_0^m}(0.01)$ and $Q_{F_0}(0.001)$ and S-95%-credible region for θ_0 , obtained with the Half-Cauchy, Gamma and Power-law models and different values of m and k . Analogous results obtained with the frequentist MLE-based confidence intervals and regions are reported in italics.

The coverage of symmetric- and asymmetric-95% credible intervals is almost the same when estimat-

ing the parameters γ_0 , $a_{m,0}$, $b_{m,0}$. While, asymmetric intervals outperform the symmetric ones when estimating $Q_{F_0^m}(0.01)$ and $Q_{F_0}(0.001)$. The smallest coverages (although still good) are obtained with the credible region $E_{k,0.05}^S$, but this is expected as estimating a three-dimensional parameter is harder than estimating a scalar one. Similar results are obtained with the other six models, see Section 2.2 of Supplementary Material [31]. The take home message is that in practice the good properties suggested by our asymptotic theory can be observed even when m and k are modest in size. The good news is that our Empirical Bayes method considerably outperforms the MLE approach. Finally, in Section 2.3 of Supplementary Material [31] we have compared the performance of our Empirical Bayes approach with that of a posterior distribution obtained with standard non-data-dependent prior.

6. Atlantic basin hurricanes wind speed analysis

The effects of wind speed and heavy rainfall generated by hurricanes are catastrophic on the affected areas. Such catastrophes have been experienced for example with Hurricane Katrina in August 2005. It caused 1,836 deaths and approximately \$125 billion in damage in the US, where the Maximum Wind Speed (MWS) reached a peak of about 280 km/h. See National Hurricane Center (NHC)'s Tropical Cyclone Reports at <https://www.nhc.noaa.gov/data/tcr/> for similar episodes.

Understanding how frequently extreme wind speeds occur is important for planning actions that can mitigate a hurricane's devastating consequences. To this aim, we consider the MWS generated by hurricanes in the Atlantic basin from 1851 to 2021 that have been collected in the Revised Atlantic Hurricane Database developed by the NHC. To better assess the risk associated with hurricanes that hit inland, we focus on the restricted area of the southeast of US, see the top-left panel of Figure 2. The red circles show the locations where hurricanes were recorded and the circles size highlights the wind speed intensity, which ranges between approximately 102 km/h and 296 km/h. We then focus on the data of the specific location marked by the blue cross, which was hit by a category 2 hurricane in 1985, named Kate, whose highest wind speed recorded there was approximately 171 km/h. We recall that Saffir-Simpson Hurricane Wind Scale classifies hurricane's wind speed as category: 1 if between 119-153 km/h; 2 if between 154-177 km/h; 3 if between 178-208 km/h; 4 if between 209-251 km/h and 5 if higher than 252 km/h, see <https://www.nhc.noaa.gov/aboutsshws.php> for details. Alternatively, one may want to consider data simultaneously on multiple sites, but unfortunately the empirical Bayes method for multivariate extremes proposed by Padoan and Rizzelli [30] is not yet mature enough to be used in practice.

Applying the selection criterion described in Section 5 of Supplementary Material [31] to the sequence of daily wind speed maxima from 1976 to 2021, we obtained that a block-size $m = 122$, resulting into a number of blocks $k = 120$, is a suitable choice. We apply our empirical Bayes procedure to the block maxima computing the posterior distribution of the GEV parameters via the algorithm described in Section 4. All the details are provided in Section 6 of Supplementary Material [31]. The posterior density of the location, scale parameters and tail index are reported from the bottom-left to bottom-right plots of Figure 2. Their shape is similar to that of a normal distribution, with that of the tail index that is slightly asymmetric on the right. The Posterior Mean (PM) and Asymmetric- and Symmetric-95%-Credibility Intervals (A- and S-95%-CI) are reported in the second, third and fourth columns of the upper part of Table 3 respectively. The former interval supports the hypothesis that the distribution of block maximum is heavy-tailed but with finite first four moments, while according to latter one a light-tailed distribution is also plausible. The forecast of extreme wind speeds is performed in two different ways: using our posterior predictive distribution of a block maximum associated to a certain time window W (see Section 3.3) and a point forecast of the annual return level, that is obtained as the quantile of the annual posterior predictive distribution, which is in turn a special case

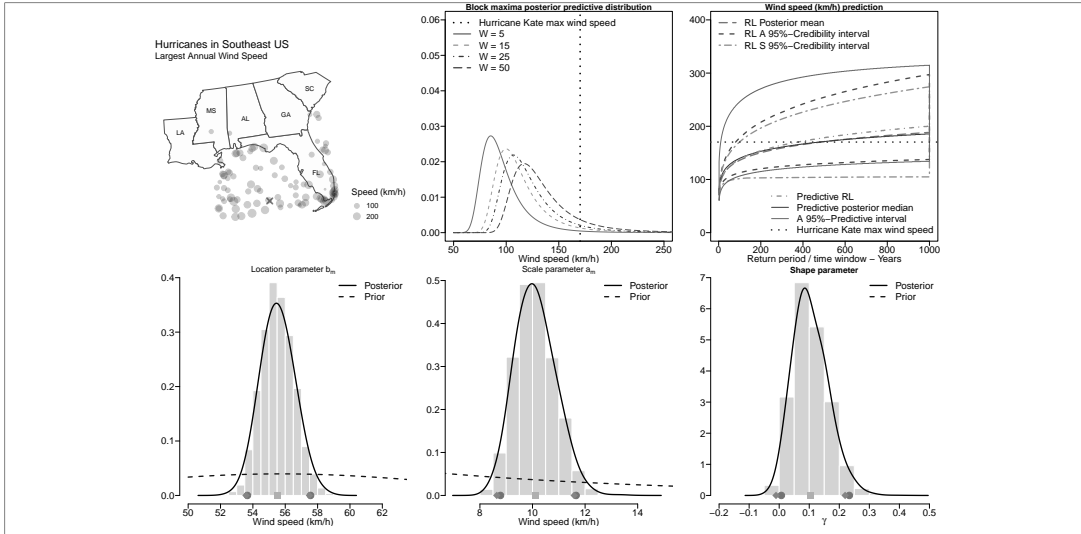


Figure 2: Southeast US map with location and intensities of the recorded hurricane wind speeds from 1851 to 2021 (top-left). Posterior predictive distributions of the largest wind speed over 5-, 15-, 25- and 50-years time window (top-middle). T -return level posterior mean and asymmetric- and symmetric-95% credibility intervals and T -return level point forecast for $T = 3W$ and posterior predictive median and asymmetric-95% predictive intervals of the largest wind speed over W -years time window with $W \in [2, \dots, 1000]$ (top-right). Posterior and prior distributions of the location, and scale parameters and tail index (form bottom-left to bottom-right). Squares are the posterior means. Circles and diamonds are the asymmetric- and symmetric-95% credibility intervals

of our more general posterior predictive distribution. The middle-top panel of Figure 2 displays the posterior predictive distribution of the 5-, 15-, 25-, and 50-years maximum. The posterior distributions of the 15-, 45-, 75- and 150-return levels associated to the four-month maximum are also computed for comparison purposes, but are reported in Figure 10 of Supplementary Material [31] to save space. According to the latter distributions a return level of intensity equal to that recorded with hurricane Kate is implausible, while in accordance with the former distributions we have that, for instance, the 50-years maximum is expected to exceed such a level with a chance of approximately 11%. The bottom part of Table 3 reports the PM and the A- and S-95%-CI relative to the posterior distributions of the quarterly return level (with return periods converted in years) and the Posterior Predictive Median (PPM) and Asymmetric-95%-Predictive Intervals (A-95%-PI) relative to the posterior predictive distributions. The latter intervals are remarkably greater than the former ones entailing that, for example, on the one hand it's hardly credible that within each quarter the largest wind speed exceeds an intensity as that generated by a category 1 hurricane with high chance (comparable to $1 - 1/15$), on the other hand it's reasonably likely (approximately 12%) that such an intensity is exceeded by the 5-years largest wind speed. Finally, the top-right panel of Figure 2 displays the PM and relative A- and S-95%-CI and the T -return level point forecast for $T = 3 \cdot W$ and the PPM and relative A-95%-PI for a time window of $W \in [2, \dots, 1000]$ years. Once again the A-95%-PI are much larger than A- and S-95%-CI uniformly over the entire time range. Concluding, when assessing extreme wind speeds, probabilistic forecasting turns out to be a more precautionary approach than the canonical method, issuing point forecasts of

Table 3. Posterior Mean (PM) and Asymmetric and Symmetric posterior 95%-Credibility Intervals (A- and S-95%-CI) for parameters and return levels that are also estimated by Posterior Predictive Median (PPM) and Asymmetric 95%-Predictive Intervals (A-95%-PI)

Parameter	PM	A-95%-CI	S-95%-CI	PPM	A-95%-PI
Location	55.5	[53.7, 57.6]	[53.6, 57.5]		
Scale	10.1	[8.8, 11.7]	[8.6, 11.6]		
Tail index	0.10	[0.01, 0.23]	[-0.01, 0.22]		
5-years RL	85.9	[80.2, 94.12]	[78.8, 92.9]	92.0	[70.5, 153.5]
10-years RL	102.7	[92.5, 119.3]	[89.0, 116.5]	102.0	[78.5, 171.0]
25-years RL	111.1	[97.9, 132.8]	[93.0, 129.1]	116.0	[89.5, 198.0]
50-years RL	123.1	[105.2, 153.9]	[97.8, 148.4]	127.5	[98.0, 221.5]

the form of return level estimates based on the posterior or the predictive distribution. Its use is then advisable to achieve richer extreme values analyses.

7. Discussion

We discuss empirical Bayes inference for the block maxima approach from the theoretical and practical perspectives. The proposed inferential method relies on the realistic assumption that block maxima are not exactly distributed according to the GEV distribution, which best describes the practical situations faced in applications. The accuracy of the inference based on our posterior distributions is guaranteed by large-sample theory. An empirical study reveals that it is satisfactory even with moderate sample sizes.

We defer to future work considerations on an even more general framework that allows to handle nonstationary block maxima and additional information, e.g. using covariates. These extensions are by far the most interesting for applications, but a still open challenge is to derive an equally solid theory also in this case. Another important way of studying extremes is through the Peaks over a Threshold approach, which is very popular in real analyses of univariate extreme events. A reformulation of our empirical Bayes method in this context would therefore be of considerable practical utility. Whether the posterior distributions of this alternative method allows for equally good performance needs to be investigated.

8. Main proofs

Here we make extensive use of the notation introduced in Sections 3.2–3.4 of Supplementary Material [31], to which we refer the reader for a comprehensive list of symbols. In particular, we denote by $\ell_k(\delta) := \log \bar{L}_k(\bar{\theta}_0 + \delta/\sqrt{k})$ the local log likelihood process, where $\delta \in \sqrt{k}(\Theta - \bar{\theta}_0)$ and $\bar{L}_k(\bar{\theta}) = \prod_{i=1}^k g_{\bar{\theta}}((M_{m,i} - b_{m,0})/a_{m,0})$ is the likelihood function obtained with reparametrisation r and rescaled block maxima. We also make use of the alternative reparametrisation

$$\tilde{r}: \Theta \mapsto \Theta : \theta \mapsto (\gamma, (b_m - \beta(m)/a_{m,0}), a_m/a_{m,0})^\top =: \tilde{\theta}$$

and corresponding posterior probability measure $\tilde{\Psi}_k(\cdot) = \Psi_k \circ \tilde{r}^{-1}(\cdot)$ and likelihood function $\tilde{L}_k(\tilde{\theta}) = \prod_{i=1}^k g_{\tilde{\theta}}((M_{m,i} - \beta(m))/a_{m,0})$. Finally, we denote by \mathcal{Q}_k the k -fold product measure pertaining to E_0^m , i.e. the probability measure of the joint distribution of BM $M_{m,1}, \dots, M_{m,k}$.

8.1. Proof of Theorem 1

If condition (b) is satisfied with $\lambda = 0$, the result in the statement of Theorem 1 follows from Lemma 3 of Supplementary Material [31]. If instead $\lambda \neq 0$, condition (c) allows to apply Lemmas 1 and 2 of Supplementary Material [31], leading to the same final result.

8.2. Proof of Theorem 2

First observe that by conditions (b)-(c) of Theorem 1 we have $(\beta(m) - b_{m,0})/a_{m,0} = O(1/\sqrt{k})$. Thus, there exists a constant $a > 0$ such that, by triangular inequality,

$$\Psi_k \left(\left\{ \theta \in \Theta : \left\| \gamma - \gamma_0, \frac{b_m - b_{m,0}}{a_m}, \frac{a_m}{a_{m,0}} - 1 \right\|_1 > \epsilon_k \right\} \right) \leq \tilde{\Psi}_k(U_{a\epsilon_k}^c).$$

For $\epsilon > 0$ and k large enough, $\tilde{\Psi}_k(U_{a\epsilon_k}^c) = \tilde{\Psi}_k(U_\epsilon \setminus U_{a\epsilon_k}) + \tilde{\Psi}_k(U_\epsilon^c)$. By Lemma 22 of Supplementary Material [31], as $k \rightarrow \infty$, we have $\tilde{\Psi}_k(U_\epsilon^c) = o_p(e^{-ck})$, with $c > 0$. Moreover, by similar arguments to those in formula (27) of Supplementary Material [31], Condition 1 and Lemma 21 of Supplementary Material [31] lead to conclude that, if ϵ is small enough, as $k \rightarrow \infty$

$$\tilde{\Psi}_k(U_\epsilon \setminus U_{a\epsilon_k}) \leq \tilde{\phi}_k + O_p(k^{3/2})(1 - \tilde{\phi}_k) \int_{U_\epsilon \setminus U_{a\epsilon_k}} \frac{\tilde{L}_k(\tilde{\theta})}{\tilde{L}_k(\tilde{\theta}_0)} \pi_{\text{sh}}(\gamma) u_{\text{loc}}(\mu) u_{\text{sc}}(\sigma) d\tilde{\theta},$$

with probability tending to 1, where

$$\tilde{\phi}_k \equiv \phi_k((M_{m,i} - \beta(m))/a_{m,0}, 1 \leq i \leq k)$$

and $\phi_k(\mathbf{y})$, $\mathbf{y} \in \mathbb{R}^k$, is a test functional defined as in Lemma 20 of Supplementary Material [31], with aC_k in place of C_k . Accordingly, the contiguity relation established in Theorem 1 allows to deduce that $\tilde{\phi}_k = o_p(e^{-c'C_k^2})$ and

$$k^{3/2}(1 - \tilde{\phi}_k) \int_{U_\epsilon^c} \frac{\tilde{L}_k(\tilde{\theta})}{\tilde{L}_k(\tilde{\theta}_0)} \pi_{\text{sh}}(\gamma) u_{\text{loc}}(\mu) u_{\text{sc}}(\sigma) d\tilde{\theta} = o_p(e^{-c''C_k^2}),$$

for some positive constants c', c'' . The result in point (a) of Theorem 2 now follows. The result in point (b) of Theorem 2 follows from an application of Lemmas 2 and 4 of Supplementary Material [31].

8.3. Proof of Theorem 3

The result is established following the main arguments of the proofs of Theorem 10.1 in [37] and Theorem 2.1 in [24], with a few adaptations. Thus, we only highlight the main changes. The proof consists of three parts: preliminaries, intermediate result and conclusion.

8.3.1. Preliminaries

Let $\delta_{k,0} = \mathbf{I}_0^{-1} S_{k,\theta_0}$ and $\tilde{\Phi}_k := \mathcal{N}(\cdot; \theta_0 + \delta_{k,0}/\sqrt{k}; k^{-1} \mathbf{I}_0^{-1})$. For all the sets of the form $\Delta_k = \sqrt{k}(\Theta_k - \tilde{\theta}_0)$, with $\Theta_k \in \mathcal{B}(\Theta)$ of positive Lebesgue measure, define $\tilde{\Psi}_k(\Delta_k) := \tilde{\Psi}_k(\tilde{\theta}_0 + \Delta_k/\sqrt{k})$ and $\tilde{\Phi}_k(\Delta_k) = \tilde{\Phi}_k(\tilde{\theta}_0 + \Delta_k/\sqrt{k})$. Note that $\tilde{\Psi}_k$ and $\tilde{\Phi}_k$ are the probability measures corresponding to the empirical

Bayes posterior distribution of the local parameter δ and of the randomly located Gaussian distribution with mean vector $\delta_{k,0}$ and covariance matrix \mathbf{I}_0^{-1} , respectively. Define the event $Z_k := \{\bar{\Psi}_k(\Theta_k) > 0\}$ and

$$\bar{\Phi}_k^{\Delta_k}(\cdot) = \frac{\bar{\Phi}_k(\cdot \cap \Delta_k)}{\bar{\Phi}_k(\Delta_k)}, \quad \bar{\Psi}_k^{\Delta_k}(\cdot) = \begin{cases} \frac{\bar{\Psi}_k(\cdot \cap \Delta_k)}{\bar{\Psi}_k(\Delta_k)}, & \mathbb{1}_{Z_k} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Observe that $\bar{\Phi}_k(\Delta_k) = \bar{\Phi}_k(\Theta_k) > 0$ with probability one, thus the random probability measure $\bar{\Phi}_k^{\Delta_k}$ is well defined. Let $0 < C_k \uparrow \infty$ and $C_k = o(\sqrt{k})$, as $k \rightarrow \infty$, and choose $\Theta_k = \{\bar{\theta} \in \Theta : \|\bar{\theta} - \bar{\theta}_0\|_1 < \epsilon_k\}$, where $\epsilon_k = C_k/\sqrt{k}$. Then,

$$\begin{aligned} \mathcal{I}(\bar{\Psi}_k, \bar{\Phi}_k) &\leq \mathcal{I}(\bar{\Psi}_k, \bar{\Phi}_k) \mathbb{1}_{Z_k} + \mathbb{1}_{Z_k^c} \\ &= \mathcal{I}(\bar{\Psi}_k, \bar{\Phi}_k) \mathbb{1}_{Z_k} + \mathbb{1}_{Z_k^c} \\ &\leq \mathcal{I}(\bar{\Psi}_k^{\Delta_k}, \bar{\Phi}_k^{\Delta_k}) \mathbb{1}_{Z_k} + \rho_k \end{aligned} \tag{18}$$

where, by Theorem 2 and the fact that $\|\delta_{k,0}\|_1 = O_p(1)$ (see equation (2.18) in [16]), as $k \rightarrow \infty$

$$\rho_k := 2 \frac{\bar{\Psi}_k(\Theta_k^c)}{\bar{\Psi}_k(\Theta_k)} \mathbb{1}_{Z_k} + 2 \frac{\bar{\Phi}_k(\Theta_k^c)}{\bar{\Phi}_k(\Theta_k)} \mathbb{1}_{Z_k} + \mathbb{1}_{Z_k^c} = o_p(1).$$

The final result is next obtained by showing that the first term on the right-hand side of (18) (third line) converges to zero in probability, if C_k is appropriately chosen.

8.3.2. Intermediate result

As an intermediate step, we show that, choosing Θ_k as above but now with $\epsilon_k = c/\sqrt{k}$, for any $c > 0$, we obtain

$$\mathcal{I}(\bar{\Psi}_k^{\Delta_k}, \bar{\Phi}_k^{\Delta_k}) \mathbb{1}_{Z_k} = o_p(1). \tag{19}$$

Note that, in this setup, $\Delta_k \subset \Delta := \{\delta \in \mathbb{R}^3 : \|\delta\|_1 \leq c\} \subset \sqrt{k}(\Theta - \bar{\theta}_0)$, for all sufficiently large k . Let \bar{p}_k and $\bar{\varphi}_k$ denote the Lebesgue densities of $\bar{\Pi}_k(\theta_0 + \cdot/\sqrt{k})$ and $\bar{\Phi}_k$, respectively. Moreover, for $\delta, \delta' \in \Delta$, let $e_k(\delta) = \exp\{\ell_k(\delta) - \ell_k(\mathbf{0})\}$ and

$$t_k(\delta, \delta') := \left(1 - \frac{\bar{\varphi}_k(\delta) e_k(\delta') \bar{p}_k(\delta')}{\bar{\varphi}_k(\delta') e_k(\delta) \bar{p}_k(\delta)} \right)_+,$$

and define the event $E_k := \{\sup_{\delta, \delta' \in \Delta} t_k(\delta, \delta') < \varepsilon\}$, for an arbitrarily small $\varepsilon > 0$. Then, we obtain the following inequality

$$\mathcal{I}(\bar{\Psi}_k^{\Delta_k}, \bar{\Phi}_k^{\Delta_k}) \mathbb{1}_{Z_k} \leq \mathcal{I}(\bar{\Psi}_k^{\Delta_k}, \bar{\Phi}_k^{\Delta_k}) \mathbb{1}_{Z_k \cap E_k} + \mathbb{1}_{Z_k \cap E_k^c}.$$

The expectation of $\mathbb{1}_{Z_k \cap E_k^c}$ is bounded from above by $\mathcal{Q}_k(E_k^c)$. By Condition 1, the expansion in formula (8) of Supplementary Material [31] and the fact that $\|\delta_{k,0}\|_1 = O_p(1)$, we have that $\sup_{\delta, \delta' \in \Delta} t_k(\delta, \delta')$ converges in probability to zero, thus $\mathcal{Q}_k(E_k^c) = o(1)$. Moreover, by similar arguments to those in the last display of page 359 of [24], we conclude that the expected value of the first term on the right-hand side in the above display is smaller than that of

$$2 \int_{\Delta_k} \int_{\Delta_k} t_k(\delta, \delta') \mathbb{1}_{Z_k \cap E_k} d\bar{\Phi}_k^{\Delta_k}(\delta) d\bar{\Psi}_k^{\Delta_k}(\delta'),$$

which, in turn, is smaller than ε . Since ε can be chosen arbitrarily small, by Markov inequality, we can now deduce the convergence result in (19).

8.3.3. Conclusion

In the previous steps, we have established that (19) holds true for any choice of Θ_k yielding a ball centered at $\mathbf{0}$ with fixed radius $\Delta_k = \{\delta \in \sqrt{k}(\Theta - \bar{\theta}_0) : \|\delta\|_1 < c\}$, for any $c > 0$. Similarly to line 11 on page 143 of [37] and lines 3–10 on page 360 of [24], we can now argue that such convergence result is still valid for some sequence $0 < C_k \uparrow \infty$ and set Θ_k yielding $\Delta_k = \{\delta \in \sqrt{k}(\Theta - \bar{\theta}_0) : \|\delta\|_1 < C_k\}$. The proof is now complete.

8.4. Proof of Corollary 1

The proof follows by similar arguments to those in the proof of Corollary 2 and is therefore omitted. We only point out that the marginal distributions of $\bar{\Psi}_k(\cdot - \delta_{k,0})$ and $\bar{\Phi}_k(\cdot - \delta_{k,0})$ have to be used herein in place of $\bar{\Omega}_k$ and Λ . Note that $\delta_{k,0}$ is as in the proof of Theorem 3.

8.5. Proof of Theorem 4

Triangular inequality and a few algebraic derivations allow to deduce that

$$\Omega_k(q \in \mathbb{R} : |q/Q_{F_0^m}(p) - 1| > \epsilon'_k) \leq \bar{\Psi}_k\left(\bar{\theta} \in \Theta : c_k d_p(\bar{\theta}) + v_k > \epsilon'_k\right),$$

where $d_p(\theta)$ is as in Definition 2(a) of Supplementary Material [31] and

$$c_k = \frac{a_{m,0}}{|Q_{F_0^m}(p)|}, \quad v_k = \left| \frac{a_{m,0} Q_{G_{\gamma_0}}(p) + \beta(m)}{Q_{F_0^m}(p)} - 1 \right|.$$

Under the considered assumptions, Theorem 2.1 in [11] guarantees that, as $k \rightarrow \infty$,

$$\frac{1}{c_k} = Q_{G_{\gamma_0}}(p) + \frac{V_0(m)}{a_{m,0}} + O(A_0(m))$$

and

$$v_k = c_k \left| Q_{G_{\gamma_0}}(p) - \frac{Q_{F_0^m}(p) - \beta(m)}{a_{m,0}} \right| = O(c_k A_0(m)).$$

Moreover, $a_{m,0}/V_0(m) \rightarrow \max(\gamma_0, 0)$ and, by hypothesis, $\sqrt{k}A_0(m) \rightarrow \lambda$. Thus, there exists $C > 0$ such that, for any $\epsilon > 0$,

$$\begin{aligned} \bar{\Psi}_k\left(\bar{\theta} \in \Theta : c_k d_p(\bar{\theta}) + v_k > \epsilon'_k\right) &\leq \bar{\Psi}_k\left(\bar{\theta} \in \Theta : d_p(\bar{\theta}) > C\epsilon_k\right) \\ &\leq \bar{\Psi}_k\left(\bar{\theta} \in U_\epsilon : d_p(\bar{\theta}) > C\epsilon_k\right) + \bar{\Psi}_k(U_\epsilon^c), \end{aligned}$$

where $\epsilon_k = C_k/\sqrt{k}$. By Lemma 22 of Supplementary Material [31], the second term on the right-hand side is of order $o_p(e^{-c'k})$, for a positive $c' > 0$. Moreover, a few adaptations to the proof of Lemma

22 of Supplementary Material [31], using Lemma 19 in place of Lemma 16 of Supplementary Material [31], lead to conclude that the first term on the right-hand side is of order $o_p(e^{-c''C_k^2})$, for a positive c'' . The final result then follows.

8.6. Proof of Theorem 5

8.6.1. Preliminaries

We start by setting up some notation and pointing out some mathematical properties used throughout the proof. Let $j(\gamma)$ be as in formula (19) of Supplementary Material [31] with $p_1 = 1 - p$. Note that the map $\gamma \mapsto Q_{G_\gamma}(p) =: \dot{q}_\gamma$ is monotone increasing, with positive and continuous derivative $j(\gamma)$ on $(\gamma_0 \pm \epsilon)$, for a suitably small $\epsilon > 0$. Consequently, the inverse map $\dot{q} \mapsto \gamma_{\dot{q}}$ is continuously differentiable over the interval $(\dot{q}_{\gamma_0 - \epsilon}, \dot{q}_{\gamma_0 + \epsilon})$, with derivative $(\partial/\partial \dot{q})\gamma_{\dot{q}} = \{j(\gamma_{\dot{q}})\}^{-1}$. By Lemma 12 of Supplementary Material [31] and the positivity of $j(\gamma)$, we also have the following Lipschitz continuity properties

$$\begin{aligned} |\gamma_{\dot{q}} - \gamma_{\dot{q}_0}| &\leq \frac{|\dot{q} - \dot{q}_0|}{\inf_{\tilde{\gamma} \in (\gamma_0 \pm \epsilon)} j(\tilde{\gamma})}, & \forall \dot{q} \in (\dot{q}_{\gamma_0 - \epsilon}, \dot{q}_{\gamma_0 + \epsilon}), \\ |\dot{q}_\gamma - \dot{q}_{\gamma_0}| &\leq \sup_{\tilde{\gamma} \in (\gamma_0 \pm \epsilon)} j(\tilde{\gamma}) |\gamma - \gamma_0|, & \forall \gamma \in (\gamma_0 \pm \epsilon). \end{aligned} \quad (20)$$

Next, we define the matrix $\dot{I}_0 := \mathbf{D} \mathbf{I}_0 \mathbf{D}$ and the random vector $\dot{\delta}_{k,0} := \dot{I}_0^{-1} \mathbf{D} S_{k, \bar{\theta}_0}$, where $\mathbf{D} = \text{diag}(\{j(\gamma_0)\}^{-1}, 1, 1)$. Finally, we define the set $\Xi := (\dot{q}_{-1}, \infty) \times \mathbb{R} \times (0, \infty)$ and the map $\mathcal{Q} : \Theta \mapsto \Xi : \bar{\theta} \mapsto (\dot{q}_{\bar{\theta}(1)}, \bar{\theta}^{(2)}, \bar{\theta}^{(3)})$. Then, for $\theta \in \Theta$, we introduce the reparametrisation

$$\dot{r}(\theta) := \mathcal{Q} \circ r(\theta) = (\dot{q}_\gamma, (b_m - b_{0,m})/a_{0,m}, a_m/a_{m,0})^\top,$$

yielding $\dot{r}(\theta_0) = (\dot{q}_{\gamma_0}, 0, 1)^\top =: \xi_0$. As an intermediate result, we show that the empirical Bayes posterior distribution of $\dot{r}(\theta)$ merges in total variation with the sequence of randomly centred normal probability measures $\dot{\Phi}_k := \mathcal{N}(\cdot; \xi_0 + \dot{\delta}_{k,0}/\sqrt{k}, k^{-1} \dot{I}_0^{-1})$ as $k \rightarrow \infty$. To do that, we adapt once more the arguments in the proofs of Theorem 10.1 in [37] and Theorem 2.1 in [24].

8.6.2. Intermediate result

By the property in (20), there exist $0 < c_1 < c_2 < \infty$ such that, for any sequence $\epsilon_k \downarrow 0$, defining

$$\Xi_k := \{\xi \in \Xi : \|\xi - \xi_0\|_1 < \epsilon_k\}$$

and, for $i = 1, 2$, $\epsilon_{k,i} = c_i \epsilon_k$, $\Theta_{k,i} := \{\bar{\theta} \in \Theta : \|\bar{\theta} - \bar{\theta}_0\|_1 < \epsilon_{k,i}\}$ and

$$\Xi_{k,i} := \{\xi \in \Xi : \mathcal{Q}^{-1}(\xi) \in \Theta_{k,i}\},$$

we have $\Xi_{k,1} \subset \Xi_k \subset \Xi_{k,2}$ for all large k . Consequently, denoting by $\dot{\Psi}_k = \Psi_k \circ \dot{r}^{-1}$ the empirical Bayes posterior of $\dot{r}(\theta)$ and letting $\dot{\Psi}_k^{\Xi_k}$ be its renormalised restriction to the set Ξ_k , provided that $\mathbb{1}_{Z_k} = 1$ with $Z_k := \{\dot{\Psi}_k(\Xi_k) > 0\}$, we have that

$$\dot{\Psi}_k(\Theta_{k,1}) = \dot{\Psi}_k(\Xi_{k,1}) \leq \dot{\Psi}_k(\Xi_k). \quad (21)$$

Let $\dot{\Phi}_k^{\Xi_k}$ be the renormalised restriction of $\dot{\Phi}_k$ on Ξ_k . Finally, define the probability measures $\dot{\Phi}_k := \dot{\Phi}_k(\sqrt{k}(\cdot - \xi_0))$ and $\dot{\Psi}_k := \dot{\Psi}_k(\sqrt{k}(\cdot - \xi_0))$, as well as their renormalised restrictions $\dot{\Phi}_k^{\dot{\Delta}_k}$ and $\dot{\Psi}_k^{\dot{\Delta}_k}$ over $\dot{\Delta}_k := \sqrt{k}(\Xi_k - \xi_0)$, provided that $\mathbb{1}_{Z_k} = 1$. Then,

$$\mathcal{I}(\dot{\Psi}_k, \dot{\Phi}_k) \leq \mathcal{I}(\dot{\Psi}_k^{\dot{\Delta}_k}, \dot{\Phi}_k^{\dot{\Delta}_k}) \mathbb{1}_{Z_k} + \rho_k, \quad (22)$$

where

$$\rho_k = 2 \frac{\dot{\Psi}_k(\Xi_k^c)}{\dot{\Psi}_k(\Xi_k)} \mathbb{1}_{Z_k} + 2 \frac{\dot{\Phi}_k(\Xi_k^c)}{\dot{\Phi}_k(\Xi_k)} \mathbb{1}_{Z_k} + \mathbb{1}_{Z_k^c}.$$

Choosing $\epsilon_k = C_k/\sqrt{k}$, for any sequence $0 < C_k \uparrow \infty$ such that $C_k = o(\sqrt{k})$, by Theorem 2 we have that $\dot{\Psi}_k(\Theta_{k,1}^c) = o_p(1)$. Thus, by the inequality in (21), we also have that $\dot{\Psi}_k(\Xi_k^c) = o_p(1)$. The latter result, together with $\|\dot{\delta}_{k,0}\|_1 = O_p(1)$, finally entails that $\rho_k = o_p(1)$. To further prove that

$$\mathcal{I}(\dot{\Psi}_k^{\dot{\Delta}_k}, \dot{\Phi}_k^{\dot{\Delta}_k}) \mathbb{1}_{Z_k} = o_p(1) \quad (23)$$

for $\epsilon_k = C_k/\sqrt{k}$ and at least one sequence C_k with the properties given above, it suffices to establish (23) for all the faster decaying sequences $\epsilon_k = C/\sqrt{k}$, with $C > 0$.

To this end, note that, by the inclusion $\Xi_k \subset \Xi_{k,2}$, for all large k we have

$$\dot{\Delta}_k \subset \dot{\Delta} := \{\dot{\delta} \in \mathbb{R}^3 : \|\dot{\delta}\|_1 \leq C\} \subset \sqrt{k}(\Xi_{k,2} - \xi_0).$$

Thus, for all $\dot{\delta} \in \dot{\Delta}$ and sufficiently large k , setting $\delta_{\dot{\delta}} := \sqrt{k}(\mathcal{Q}^{-1}(\xi_0 + \dot{\delta}/\sqrt{k}) - \bar{\theta}_0)^\top$, we have

$$\delta_{\dot{\delta}} \in \Delta := \{\delta \in \mathbb{R}^3 : \|\delta\|_1 \leq c_2 C\} \subset \sqrt{k}(\Theta - \bar{\theta}_0).$$

As a result, using the expansion in formula (8) of Supplementary Material [31], we obtain

$$\begin{aligned} \sup_{\dot{\delta} \in \dot{\Delta}} \left| \ell_k(\delta_{\dot{\delta}}) - \ell_k(\mathbf{0}) - \delta_{\dot{\delta}}^\top S_{k, \bar{\theta}_0} + \frac{\delta_{\dot{\delta}}^\top \mathbf{I}_0 \delta_{\dot{\delta}}}{2} \right| &\leq \sup_{\delta \in \Delta} \left| \ell_k(\delta) - \ell_k(\mathbf{0}) - \delta^\top S_{k, \bar{\theta}_0} + \frac{\delta^\top \mathbf{I}_0 \delta}{2} \right| \\ &= o_p(1). \end{aligned}$$

At the same time, the mean-value theorem guarantees that for all $\dot{\delta} \in \dot{\Delta}$ there exists \dot{q} between $\min(\dot{q}_{\gamma_0}, \dot{q}_{\gamma_0} + \dot{\delta}^{(1)}/\sqrt{k})$ and $\max(\dot{q}_{\gamma_0}, \dot{q}_{\gamma_0} + \dot{\delta}^{(1)}/\sqrt{k})$ such that

$$\|\delta_{\dot{\delta}} - \mathbf{D}\dot{\delta}\|_1 = |[\{j(\gamma_{\dot{q}})\}^{-1} - \{j(\gamma_0)\}^{-1}] \dot{\delta}^{(1)}|.$$

Therefore, by the continuity of $\dot{q} \mapsto j(\gamma_{\dot{q}})$ at \dot{q}_{γ_0} ,

$$\sup_{\dot{\delta} \in \dot{\Delta}} \left| \delta_{\dot{\delta}}^\top S_{k, \bar{\theta}_0} - \frac{\delta_{\dot{\delta}}^\top \mathbf{I}_0 \delta_{\dot{\delta}}}{2} - (\mathbf{D}\dot{\delta})^\top S_{k, \bar{\theta}_0} + \frac{(\mathbf{D}\dot{\delta})^\top \mathbf{I}_0 (\mathbf{D}\dot{\delta})}{2} \right| = o_p(1).$$

Consequently, using triangular inequality, we deduce that

$$\sup_{\dot{\delta} \in \dot{\Delta}} \left| \ell_k(\delta_{\dot{\delta}}) - \ell_k(\mathbf{0}) - (\mathbf{D}\dot{\delta})^\top S_{k, \bar{\theta}_0} + \frac{(\mathbf{D}\dot{\delta})^\top \mathbf{I}_0 (\mathbf{D}\dot{\delta})}{2} \right| = o_p(1). \quad (24)$$

Moreover, defining the data-dependent prior density of the local parameter $\sqrt{k}(\hat{r}(\boldsymbol{\theta}) - \boldsymbol{\xi}_0)$ at $\hat{\boldsymbol{\delta}} \in \hat{\Delta}$ by

$$\dot{p}_k(\hat{\boldsymbol{\delta}}) := \pi_k \circ r^{-1}(\hat{\boldsymbol{\theta}}_0 + \boldsymbol{\delta}_{\hat{\boldsymbol{\delta}}}/\sqrt{k}) \{ \sqrt{k} j(\gamma_{\dot{q}_{\gamma_0 + \hat{\boldsymbol{\delta}}^{(1)}/\sqrt{k}}}) \}^{-1} a_{m,0}^2,$$

by Condition 1 and the continuity of $q \mapsto j(\gamma_q)$ at q_{γ_0} , it holds that

$$\sup_{\hat{\boldsymbol{\delta}} \in \hat{\Delta}} \left| \frac{\dot{p}_k(\hat{\boldsymbol{\delta}})}{\dot{p}_k(\mathbf{0})} - 1 \right| = o_p(1). \quad (25)$$

Hence, letting $\dot{\varphi}_k$ be the Lebesgue density of $\dot{\Phi}_k$, setting $\dot{e}_k(\hat{\boldsymbol{\delta}}) = \exp\{\ell_k(\boldsymbol{\delta}_{\hat{\boldsymbol{\delta}}}) - \ell_k(\mathbf{0})\}$,

$$i_k(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\delta}}') := \left(1 - \frac{\dot{\varphi}_k(\hat{\boldsymbol{\delta}}) \dot{e}_k(\hat{\boldsymbol{\delta}}') \dot{p}_k(\hat{\boldsymbol{\delta}}')}{\dot{\varphi}_k(\hat{\boldsymbol{\delta}}') \dot{e}_k(\hat{\boldsymbol{\delta}}) \dot{p}_k(\hat{\boldsymbol{\delta}})} \right)_+$$

with $\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\delta}}' \in \hat{\Delta}$ and using (24)–(25) along with the equalities $(\mathbf{D}\hat{\boldsymbol{\delta}})^\top S_{k, \hat{\boldsymbol{\theta}}_0} = \hat{\boldsymbol{\delta}}^\top \dot{\mathbf{I}}_0 \hat{\boldsymbol{\delta}}_{k,0}$, $(\mathbf{D}\hat{\boldsymbol{\delta}})^\top \mathbf{I}_0(\mathbf{D}\hat{\boldsymbol{\delta}}) = \hat{\boldsymbol{\delta}}^\top \dot{\mathbf{I}}_0 \hat{\boldsymbol{\delta}}$ and $\|\hat{\boldsymbol{\delta}}_{k,0}\|_1 = O_p(1)$, we conclude that $\sup_{\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\delta}}' \in \hat{\Delta}} i_k(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\delta}}') = o_p(1)$. Consequently, defining the event $E_k := \{\sup_{\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\delta}}' \in \hat{\Delta}} i_k(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\delta}}') > \varepsilon\}$, for an arbitrarily small $\varepsilon > 0$, we have that the expected value of the term on the left-hand side of (23) is bounded from above by that of

$$2 \int_{\hat{\Delta}_k} \int_{\hat{\Delta}_k} i_k(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\delta}}') \mathbb{1}_{Z_k \cap E_k} d\dot{\Phi}_k^{\hat{\Delta}_k}(\hat{\boldsymbol{\delta}}) d\dot{\Psi}_k^{\hat{\Delta}_k}(\hat{\boldsymbol{\delta}}') + \mathbb{1}_{Z_k \cap E_k^c},$$

which in turn is smaller than 3ε , for all sufficiently large k .

We can now deduce that (23) holds true for all choices $\varepsilon_k = C/\sqrt{k}$. Thus, we can also deduce that there exists a sequence $0 < C_k \uparrow \infty$ satisfying $C_k = o(\sqrt{k})$ and such that (23) holds true for $\varepsilon_k = C_k/\sqrt{k}$. Using this result along with $\rho_k = o_p(1)$ and (22), we finally conclude that

$$\mathcal{T}(\dot{\Psi}_k, \dot{\Phi}_k) = \mathcal{T}(\dot{\Psi}_k, \dot{\Phi}_k) = o_p(1), \quad (26)$$

which is the desired intermediate result.

8.6.3. Conclusion

Define the sequences

$$c_k := \frac{Q_{G_{\boldsymbol{\theta}_0}}(p)}{Q_{F_0^m}(p)} \frac{\varrho_k}{\sqrt{k}} \frac{a_{m,0}}{b_{m,0} + a_{m,0} \dot{q}_{\gamma_0}}, \quad v_k := \varrho_k \left[\frac{Q_{G_{\boldsymbol{\theta}_0}}(p)}{Q_{F_0^m}(p)} - 1 \right]$$

and $d_k := c_k \dot{q}_{\gamma_0}$, then define the row vectors $\boldsymbol{\tau}_k = (c_k, c_k, d_k)$ and the maps

$$T_k : \sqrt{k}(\Xi - \boldsymbol{\xi}_0) \mapsto \mathbb{R} : \hat{\boldsymbol{\delta}} \mapsto c_k(\hat{\boldsymbol{\delta}}^{(1)} + \hat{\boldsymbol{\delta}}^{(2)} + \hat{\boldsymbol{\delta}}^{(1)} \hat{\boldsymbol{\delta}}^{(3)}/\sqrt{k}) + d_k \hat{\boldsymbol{\delta}}^{(3)} + v_k.$$

Note that $Q_{G_{\boldsymbol{\theta}_0}}(p) = a_{m,0} \dot{q}_{\gamma_0} + b_{m,0}$ and, therefore,

$$\varrho_k \left[\frac{Q_{G_{\boldsymbol{\theta}_0}}(p)}{Q_{F_0^m}(p)} - 1 \right] = T_k \left(\sqrt{k}(\hat{r}(\boldsymbol{\theta}) - \boldsymbol{\xi}_0) \right).$$

As a consequence, the empirical Bayes posterior distribution of the term on the left-hand side is $\Omega_k(Q_{F_0^m}(p) + \cdot Q_{F_0^m}(p))/\varrho_k = \dot{\Psi}_k \circ T_k^{-1}$. Next, observe that under the considered assumptions $\lim_{k \rightarrow \infty} a_{m,0}/V_0(m) = \max(\gamma_0, 0)$, entailing

$$c := \lim_{k \rightarrow \infty} c_k = \begin{cases} \frac{1}{1 + \gamma_0 \dot{q}_{\gamma_0}}, & \gamma_0 > 0, \\ 1, & \gamma_0 = 0, \\ 1, & \gamma_0 < 0 \text{ and } V_0(\infty) > 0, \\ -1, & \gamma_0 < 0 \text{ and } V_0(\infty) \leq 0, \end{cases}$$

and $d := \lim_{k \rightarrow \infty} d_k = \dot{q}_{\gamma_0} c$, while

$$v := \lim_{k \rightarrow \infty} v_k = c \lambda H_{\gamma_0, \rho} \left(-\frac{1}{\log(1-p)} \right).$$

Then, define $T(\dot{\delta}) = c(\dot{\delta}_1 + \dot{\delta}_2) + d\dot{\delta}_3 + v$ and the row vector $\tau = (c, c, d)$. We now have that, by triangular inequality,

$$\mathcal{I}(\dot{\Psi}_k \circ T_k^{-1}, \dot{\Phi}_k \circ T^{-1}) \leq \mathcal{I}(\dot{\Psi}_k \circ T_k^{-1}, \dot{\Phi}_k \circ T_k^{-1}) + \mathcal{I}(\dot{\Phi}_k \circ T_k^{-1}, \dot{\Phi}_k \circ T^{-1}).$$

On one hand, by the result in (26),

$$\mathcal{I}(\dot{\Psi}_k \circ T_k^{-1}, \dot{\Phi}_k \circ T_k^{-1}) \leq \mathcal{I}(\dot{\Psi}_k, \dot{\Phi}_k) = o_p(1).$$

On the other hand, since $\|\dot{\delta}_{k,0}\|_1 = O_p(1)$, $\dot{\Phi}_k \circ T_k^{-1}$ is the probability measure of a univariate Gaussian distribution whose mean and variance equal $\tau_k \dot{\delta}_{k,0} + v_k + O_p(k^{-1/2})$ and $\tau_k \dot{I}_0^{-1} \tau_k^\top + O_p(k^{-1/2})$, respectively, while $\dot{\Phi}_k \circ T^{-1} = \mathcal{N}(\cdot; \tau \dot{\delta}_{k,0} + v, \tau \dot{I}_0^{-1} \tau^\top)$. Thus, by Theorem 1.3 in [13],

$$\begin{aligned} \mathcal{I}(\dot{\Phi}_k \circ T_k^{-1}, \dot{\Phi}_k \circ T^{-1}) &\leq \frac{3|\tau_k \dot{I}_0^{-1} \tau_k^\top - \tau \dot{I}_0^{-1} \tau^\top| + O_p(k^{-1/2})}{2\tau \dot{I}_0^{-1} \tau^\top} \\ &\quad + \frac{|(\tau_k - \tau) \dot{\delta}_{k,0}| + |v_k - v| + O_p(k^{-1/2})}{2\sqrt{\tau \dot{I}_0^{-1} \tau^\top}} \\ &= o_p(1). \end{aligned}$$

Concluding, note that by defining the vector $\mathbf{D}_0^\top = \tau \mathbf{D}^{-1}$, we have that $\tau \dot{\delta}_{k,0} = \tau \mathbf{D}^{-1} \mathbf{I}_0^{-1} S_{k, \bar{\theta}_0} = \mathbf{D}_0^\top \delta_{k,0}$ and

$$\tau \dot{I}_0^{-1} \tau^\top = \tau \mathbf{D}^{-1} \mathbf{I}_0^{-1} \mathbf{D}^{-1} \tau^\top = \mathbf{D}_0^\top \mathbf{I}_0^{-1} \mathbf{D}_0.$$

The proof is now complete.

8.7. Proof of Corollary 2

Fix $\alpha \in (0, 1)$ and set $\dot{A}_k = \varrho_k(I_{k,\alpha}^A / \mathcal{Q}_{F_0^m}(p) - 1) - \mathbf{D}_0 \delta_{k,0}$, where $\delta_{k,0}$ is as in the proof of Theorem 3. Then, let

$$\dot{\Omega}_k = \Omega_k(\cdot \mathcal{Q}_{F_0^m}(p) / \varrho_k + \mathcal{Q}_{F_0^m}(p) - \mathbf{D}_0 \delta_{k,0})$$

and observe that

$$\dot{A}_k = \left(\mathcal{Q}_{\dot{\Omega}_k}(1 - \alpha/2), \mathcal{Q}_{\dot{\Omega}_k}(\alpha/2) \right).$$

Next, let $\Lambda = \mathcal{N}(\cdot; 0, \mathbf{D}_0^\top \mathbf{I}_0^{-1} \mathbf{D}_0)$. Thus, by Theorem 5, we have that $\mathcal{T}(\dot{\Omega}_k, \Lambda) = o_p(1)$ and therefore

$$\mathcal{Q}_{\dot{\Omega}_k}(1 - \alpha/2) = \mathcal{Q}_\Lambda(1 - \alpha/2) + o_p(1),$$

$$\mathcal{Q}_{\dot{\Omega}_k}(\alpha/2) = \mathcal{Q}_\Lambda(\alpha/2) + o_p(1).$$

As a consequence, for any $\epsilon \in (0, \min\{\alpha/2, (1 - \alpha)/2\})$, defining the intervals

$$\dot{A}_\epsilon^+ := (\mathcal{Q}_\Lambda(1 - \alpha/2 + \epsilon), \mathcal{Q}_\Lambda(\alpha/2 - \epsilon))$$

and

$$\dot{A}_\epsilon^- := (\mathcal{Q}_\Lambda(1 - \alpha/2 - \epsilon), \mathcal{Q}_\Lambda(\alpha/2 + \epsilon)),$$

we have that $\mathbf{Q}_k\{\dot{A}_k \subset \dot{A}_\epsilon^+\} = 1 + o(1)$ and $\mathbf{Q}_k\{\dot{A}_\epsilon^- \subset \dot{A}_k\} = 1 + o(1)$. In turn, we have that

$$\begin{aligned} \mathbf{Q}_k\{\mathcal{Q}_{F_0^m}(p) \in I_{k,\alpha}^A\} &= \mathbf{Q}_k\{-\mathbf{D}_0^\top \delta_{k,0} \in \dot{A}_k\} \\ &\leq \mathbf{Q}_k\{-\mathbf{D}_0^\top \delta_{k,0} \in \dot{A}_\epsilon^+\} + o(1) \\ &= 1 - \alpha + 2\epsilon + o(1), \end{aligned}$$

where the last line follows from $\mathbf{D}_0^\top \delta_{k,0} \xrightarrow{d} \mathcal{N}(0, \mathbf{D}_0^\top \mathbf{I}_0^{-1} \mathbf{D}_0)$. Similarly, we have that

$$\begin{aligned} \mathbf{Q}_k\{\mathcal{Q}_{F_0^m}(p) \in I_{k,\alpha}^A\} &\geq \mathbf{Q}_k\{-\mathbf{D}_0^\top \delta_{k,0} \in \dot{A}_\epsilon^-\} + o(1) \\ &= 1 - \alpha - 2\epsilon + o(1). \end{aligned}$$

The result now follows, since ϵ can be chosen arbitrarily small.

8.8. Proof of Corollary 3

The result in point (a) of Corollary 3 is a direct consequence of Proposition 3 of Supplementary Material [31] and Theorem 8.8 in [20]. The result in point (b) of Corollary 3 follows from that at point (a) and Lemma B.1(i) in [20]. The result in point (c) of Corollary 3 follows from that at point (b). We further have that

$$\|\hat{G}_k^{(m^*)} - F_0^{m^*}\|_\infty = O_p(\epsilon_k), \quad k \rightarrow \infty,$$

for all the sequences $\epsilon_k = C_k / \sqrt{k}$ satisfying the properties in the statement. As a result, with probability tending to 1, for any fixed p

$$1 - p - \epsilon_k \leq F_0^{m^*} \left(\mathcal{Q}_{\hat{G}_k^{(m^*)}}(p) \right) \leq 1 - p + \epsilon_k,$$

which can be re-expressed as

$$\frac{V_0\left(\frac{m^*}{-\log(1-p-\epsilon_k)}\right)}{V_0\left(\frac{m^*}{-\log(1-p)}\right)} - 1 \leq \frac{Q_{\hat{G}_k^{(m^*)}}(p)}{Q_{F_0^{m^*}}(p)} - 1 \leq \frac{V_0\left(\frac{m^*}{-\log(1-p-\epsilon_k)}\right)}{V_0\left(\frac{m^*}{-\log(1+p)}\right)} - 1. \quad (27)$$

The term on the left-hand side equals the product of

$$\chi_1(p) := \frac{V_0\left(\frac{m^*}{-\log(1-p-\epsilon_k)}\right) - V_0\left(\frac{m^*}{-\log(1-p)}\right)}{-\log(1-p) V_0'\left(\frac{m^*}{-\log(1-p)}\right)}$$

and

$$\chi_2(p) := \frac{-\log(1-p) V_0'\left(\frac{m^*}{-\log(1-p)}\right)}{V_0\left(\frac{m^*}{-\log(1-p)}\right)}.$$

Using bounds (2.19a)–(2.19b) in Theorem 2.3 of [11] and the properties of A_0 due to regular variation (see, e.g., [2,33]), along with a few algebraic derivations, we can deduce that for a constant $a_1 > 0$ as $k \rightarrow \infty$

$$\chi_1(p) \geq -a_1 \epsilon_k + o(k^{-1/2})$$

for all $p \in K$. Moreover, if $\gamma_0 > 0$, the function $xV_0'(x)/V_0(x)$ is slowly varying and satisfies $\lim_{x \rightarrow \infty} xV_0'(x)/V_0(x) = \gamma_0$. Thus, by Theorem 1.2.1 in [2]

$$\chi_2(p) = (1 + o(1)) \frac{m^* V_0'(m^*)}{V_0(m^*)} = (1 + o(1)) \frac{a_{m^*,0}}{|b_{m^*,0}|}$$

where the $o(1)$ terms are uniform in $p \in K$. If, instead, $\gamma_0 \leq 0$, the fact that $\lim_{x \rightarrow \infty} xV_0'(x)/V_0(x) = 0$ and the bounds (2.18) and (2.19a)–(2.19b) in Theorem 2.3 of [11] yield

$$\begin{aligned} |\chi_2(p)| &\leq (a_2 + o(1)) \frac{m^* V_0'(m^*)}{\left|V_0\left(\frac{m^*}{-\log(1-p)}\right)\right|} \\ &\leq \frac{a_2 + o(1)}{a_3 + o(1) + \frac{|V_0(m^*)|}{m^* V_0'(m^*)}} \\ &\leq a_4 \frac{a_{m^*,0}}{|b_{m^*,0}|} \end{aligned}$$

as $k \rightarrow \infty$, where a_2, a_3, a_4 are positive constants and the $o(1)$ terms are uniform in $p \in K$. As a consequence of all these results, there exists a constant $b > 0$ such that, as $k \rightarrow \infty$, the term on the left-hand side in (27) is bounded from below by $-b\epsilon_k a_{m^*,0}/|b_{m^*,0}|$, uniformly in $p \in K$. By similar arguments, there exists a constant $c > 0$ such that, as $k \rightarrow \infty$, the term on the right-hand side in (27) is bounded from above by $c\epsilon_k a_{m^*,0}/|b_{m^*,0}|$, uniformly in $p \in K$. The result in point (d) of Corollary 3 now follows, and the proof is complete.

Acknowledgements

The authors are grateful to Botond Szabo for his helpful suggestions. The authors are also grateful to the Editor, Associate Editor and two anonymous referees for their constructive suggestions, which have undoubtedly improved the presentation of this work. Simone Padoan is supported by the Bocconi Institute for Data Science and Analytics (BIDSA), Italy.

Supplementary Material

Supplementary material (insert doi). Supplement to “Empirical Bayes inference for the block maxima method” [31]. The supplementary material document contains a discussion on the validity of the technical conditions, further numerical results, a detailed discussion on the used theoretical framework, technical lemmas for the main proofs, further results on the real application.

References

- [1] BEIRLANT, J., GOEGBEUR, Y., SEGERS, J. and TEUGELS, J. L. (2004). *Statistics of Extremes: Theory and Applications*. Chichester: Wiley.
- [2] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1989). *Regular Variation*. Cambridge: Cambridge University Press.
- [3] BÜCHER, A. and SEGERS, J. (2017). On the maximum likelihood estimator for the Generalized Extreme-Value distribution. *Extremes* **20** 839–872.
- [4] CHOUDHURI, N., GHOSAL, S. and ROY, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.* **99** 1050–1059.
- [5] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Berlin: Springer Verlag.
- [6] COLES, S. and PERICCHI, L. (2003). Anticipating catastrophes through extreme value modelling. *J. R. Stat. Soc. Ser. C Appl. Stat.* **52** 405–416.
- [7] COLES, S. G. and POWELL, E. A. (1996). Bayesian methods in extreme value modelling: a review and new developments. *Int. Stat. Rev.* **64** 119–136.
- [8] COLES, S. G. and TAWN, J. A. (1996). A Bayesian analysis of extreme rainfall data. *J. R. Stat. Soc. Ser. C Appl. Stat.* **45** 463–478.
- [9] DE CARVALHO, M., PEREIRA, S., PEREIRA, P. and DE ZEA BERMUDEZ, P. (2021). An extreme value Bayesian Lasso for the conditional left and right tails. *J. Agric. Biol. Environ. Stat.* **27** 1–18.
- [10] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. New York: Springer.
- [11] DE HAAN, L. and RESNICK, S. (1996). Second-order regular variation and rates of convergence in extreme-value theory. *Ann. Prob.* **24** 97–124.
- [12] DE HAAN, L. and STADTMÜLLER, U. (1996). Generalized regular variation of second order. *J. Aust. Math. Soc.* **61** 381–395.
- [13] DEVROYE, L., MEHRABIAN, A. and REDDAD, T. (2020). The total variation distance between high-dimensional Gaussians. *ArXiv e-prints*: <https://arxiv.org/pdf/1810.08693.pdf>.
- [14] DIEBOLT, J., GUILLOU, A., NAVEAU, P. and RIBEREAU, P. (2008). Improving Probability-Weighted Moment methods for the Generalized Extreme Value distribution. *REVSTAT* **6** 33–50.
- [15] 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.
- [16] DOMBRY, C. and FERREIRA, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723.
- [17] FERREIRA, A. and DE HAAN, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298.

- [18] GARTHWAITE, P. H., FAN, Y. and SISSON, S. A. (2016). Adaptive optimal scaling of Metropolis–Hastings algorithms using the Robbins–Monro process. *Comm. Statist. Theory Methods* **45** 5098–5111.
- [19] GELMAN, A., GILKS, W. R. and ROBERTS, G. O. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120.
- [20] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge: Cambridge University Press.
- [21] HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242.
- [22] 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.
- [23] JENKINSON, A. (1969). Statistics of extremes. In *Estimation of maximum floods, WMO Tech. Note 98* 183–228.
- [24] KLEIJN, B. J. K. and VAN DER VAART, A. W. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electron. J. Stat.* **6** 354–381.
- [25] KNIGHT, K. (2000). *Mathematical Statistics*. London: Chapman & Hall/CRC.
- [26] KRÜGER, F., LERCH, S., THORARINSDOTTIR, T. and GNEITING, T. (2021). Predictive inference based on Markov Chain Monte Carlo output. *Int. Stat. Rev.* **89** 274–301.
- [27] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*. New York: Springer.
- [28] MARIBE, G., VERSTER, A. and BEIRLANT, J. (2016). Reducing MSE in estimation of heavy tails: a Bayesian approach. *arXiv preprint arXiv:1606.05687*.
- [29] NORTHROP, P. J. and ATTALIDES, N. (2016). Posterior propriety in Bayesian extreme value analyses using reference priors. *Statist. Sinica* **26** 721–743.
- [30] PADOAN, S. A. and RIZZELLI, S. (2022). Consistency of Bayesian inference for multivariate max-stable distributions. *Ann. Statist.* **50** 1490 – 1518.
- [31] PADOAN, S. A. and RIZZELLI, S. (2023). Supplement to “Empirical Bayes inference for the block maxima method”. <https://doi.org/insert>
- [32] PRESCOTT, P. and WALDEN, A. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67** 723–724.
- [33] RESNICK, S. I. (2007). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer.
- [34] ROBERT, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. New York: Springer.
- [35] SMITH, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72** 67–90.
- [36] STEPHENSON, A. and TAWN, J. (2004). Bayesian inference for extremes: accounting for the three extremal types. *Extremes* **7** 291–307.
- [37] VAN DER VAART, A. (2000). *Asymptotic Statistics*. Cambridge: Cambridge University Press.