Electronic Journal of Statistics
Vol. 18 (2024) 5258–5303
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2323

ANOVEX: ANalysis Of Variability for heavy-tailed EXtremes

Stéphane Girard[1], **Thomas Opitz**[2] **and Antoine Usseglio-Carleve**[3]

[1] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France,*
e-mail: stephane.girard@inria.fr

[2] *INRAE, Biostatistics and Spatial Processes, 84914 Avignon, France,*
e-mail: thomas.opitz@inrae.fr

[3] *Avignon Université, LMA UPR 2151, 84000 Avignon, France,*
e-mail: antoine.usseglio-carleve@univ-avignon.fr

Abstract: While analysis of variance (ANOVA) tests for differences in the means of independent samples, it is unsuitable for evaluating differences in tail behavior, especially when means do not exist or empirical estimation of means or higher moments is not consistent due to heavy-tailed distributions. Here, we propose an ANOVA-like decomposition to analyze tail variability, allowing for flexible representation of heavy tails through a set of user-defined extreme quantiles, possibly located outside the range of observations. Assuming regular variation, we introduce a test for significant tail differences across multiple independent samples and derive its asymptotic distribution. We investigate the theoretical behavior of the test statistics for the case of two samples, each following a Pareto distribution, and explore strategies for setting test hyperparameters. We conduct simulations that highlight generally reliable test behavior for a wide range of finite-sample situations. The test is applied to identify clusters of financial stock indices with similar extreme log-returns and to detect temporal changes in daily precipitation extremes in Germany.

MSC2020 subject classifications: 62G32, 62E20.
Keywords and phrases: Analysis of variance, extreme quantile, extreme-value theory, heavy tail, hypothesis test.

Received October 2023.

1. Introduction

Consider J samples of independent and identically distributed (i.i.d.) realizations of random variables with cumulative distribution functions $F_1, \dots, F_J$, respectively, where $J > 1$. If F_j has mean μ_j for $j = 1, \dots, J$, the classical ANalysis Of VAriance (ANOVA) tests the equality of the J means using a decomposition of the total variance into intra-class and inter-class variances. However, this approach to detect data heterogeneity suffers from some limitations. First, asymptotic properties of the test statistic are usually established under a Gaussian assumption, see [41, Chapter 10] for a possible relaxation of this assumption based on the central limit theorem. Moreover, testing the equality of means always requires the existence of the first moment, and even of the second moment in the ANOVA setting, whereas such conditions are not fulfilled by many distributions. Certain extensions relax these conditions by proposing, for instance, to test the equality of medians, as in [22, Chapter 6], or to use rank-transforms [30], or to generalize ANOVA by assuming only the existence of low-order moments [40]. All of these approaches focus on the central behavior of the distribution and are not suited to detect heterogeneity in the distribution tails. Distributions can share the same mean or median but show strongly contrasted tail behavior.

Many biological, environmental or physical phenomena, as well as financial and actuarial data, are known to be heavy-tailed and possess power-law behavior [6, 24, 31, 37, 38]. A common assumption is that the survival functions (s.f.) associated with $F_1, \dots, F_J$ are regularly varying, *i.e.*,

$$\overline{F}_j(x) = 1 - F_j(x) = x^{-1/\xi_j}\mathcal{L}_j(x), \quad \frac{\mathcal{L}_j(tx)}{\mathcal{L}_j(x)} \to 1, \ \ x \to \infty, \ t > 0, \ j = 1, \dots, J, \tag{1}$$

where $\xi_j > 0$ are the tail indices and $\mathcal{L}_j$ are slowly varying functions (as defined by the above limit). Condition (1) characterizes the class of distributions with positive tail index, also known as the Fréchet maximum domain of attraction [10, Theorem 1.2.1]. Such distributions only admit moments of order less than $1/\xi_j$, which limits classical ANOVA to moderately heavy-tailed distributions satisfying $\xi < 1/2$. We cite two counter-examples: claim amounts caused by tornadoes were found to have an estimated tail index close to one [7]; burnt areas of wildfires often have estimated tail index beyond one half [28, 36].

Certain test procedures focus on the tail indices $\xi_1, \dots, \xi_J$. For example, in [32] a test was developed for equal tail indices based on a nonparametric minimax principle; a test for comparing positive tail indices via an empirical likelihood ratio is introduced in [44]; a test for different tail indices in the setting of heteroscedastic extremes was proposed in [14, Section 3]. However, tests based on tail indices neglect the slowly varying functions $\mathcal{L}_j$, which could be modified by any positive factor without affecting either the tail index ξ_j or the behavior of classical tail-index estimators. Therefore, in practice the tail of a heavy-tailed distribution is usually described by both the tail index, also called shape parameter, and a scale parameter. In the Peaks-Over-Threshold (POT) method [9], positive excesses above a high threshold are modeled by a Generalized Pareto distribution with both shape and scale parameters, allowing for model-based tail comparison.

Here, we focus on extreme quantiles with probability level $\alpha = \alpha_n \to 1$ as the sample size n tends to infinity. Two heavy-tailed distributions have asymptotically equivalent extreme quantiles if and only if they share the same shape and scale parameters, and extreme quantiles are therefore interesting tools for testing tail heterogeneity. In [21], extreme quantiles were used to propose a test for the existence of a change point in a series of random variables, and the work in [11] concerns quantile-based tests for causal treatment effects. However, such quantile-based tests usually consider only a single quantile, which can be problematic since distributions with very different tail behavior can have the same quantile at a given probability level. The same issue occurs in [34], where the equality across samples of a single expectile (an alternative to the quantile) is tested. Moreover, if we focus only on a single parameter such as the shape, the scale or a single quantile, this can lead to relatively high uncertainty in statistical estimations and therefore relatively weak power of statistical tests for detecting differences in tails. If parametric assumptions are made about the tail distribution and lead to a tractable probability density function, then likelihood-based methods could be used to compare models having different parameters for different samples with models where some of the parameters are held equal across several samples (*e.g.* similar to the change-point analysis of [12, 13]), but we do not pursue such model-based approaches here.

Inspired by ANOVA, we propose to test for the equality of L extreme quantiles $q_1(\alpha_{\ell,n}),\dots,q_J(\alpha_{\ell,n})$, $\ell=1,\dots,L$, with $L>1$, across the J samples, by decomposing the variability of the corresponding extreme-quantile estimates. This puts focus on the most extreme possible events, for which statistical techniques that are directly based on the empirical distribution are not applicable. By jointly studying multiple quantiles with moderately large L (typically in the range of two to 100 depending on the application context), we obtain a precise characterization of tail behavior at extreme levels, and both the statistical uncertainty and the numerical cost of required computations remain relatively low. In practice, for a useful and straightforward choice of the L quantiles $\alpha_{\ell,n}$, $\ell=1,\dots,L$, when the sample size is n, we suggest using $\alpha_{\ell,n}=1-\ell/n$, such that the lowest considered extreme quantile at level $1-L/n$ is exceeded approximately L times in the sample, and the highest quantile at level $1-1/n$ is exceeded by events that occur approximately once with the given sample size. Nevertheless, our theoretical results are formulated for more general choices of $\alpha_{\ell,n}$. Our approach is simple, does not require the existence of any moment, and is based on the general heavy-tail assumption (1) that encompasses a wide class of distributions commonly used for risk management, such as the Pareto, Generalized Pareto, Fréchet or Student's t distributions.

The remainder of the paper is organized as follows. The new test statistic, called ANOVEX (*ANalysis Of Variability in EXtremes*), is introduced in Section 2, based on extreme quantile estimators. Section 3 then provides the asymptotic properties of this test statistic and precisely defines the ANOVEX test procedure. The statistical power of ANOVEX is discussed in Section 4 thanks to approximations of type I and type II errors for three examples involving Pareto distributions in the two-sample setting, *i.e.*, for $J=2$. The simulation study in Section 5 highlights the test performance in various simulation scenarios. Two real data examples (financial and environmental data) are considered in Section 6. The proofs of the results are postponed to Appendix A, while figures illustrating some additional simulation results are shown in Appendices B and C, respectively.

2. Setting and assumptions

We consider $J>1$ samples $E_j=\{X_i^{(j)},\, i=1,\dots,n_j\}$, $j=1,\dots,J$, with independence between samples and possibly different sample sizes $n_j>1$. Throughout, $X^{(j)}_{1,n_j} \le \ldots \le X^{(j)}_{n_j,n_j}$ denote the order statistics associated with E_j, $j = 1, \ldots, J$. We assume that the random variables in each E_j are identically distributed (independence within E_j is not required at this stage) according to a cumulative distribution function F_j satisfying condition (1). The latter property is denoted by $F_j \in \mathcal{C}_1(\xi_j)$ in the following definition.

Definition 1 (Class $\mathcal{C}_1(\xi)$)**.** *The cumulative distribution function F is said to belong to the class $\mathcal{C}_1(\xi)$, $\xi > 0$, if it satisfies condition* (1).

Equivalently,

$$\lim_{t\to\infty} \frac{\overline{F}(ty)}{\overline{F}(t)} = y^{-1/\xi} \text{ for all } y > 0.$$

For asymptotic results, we consider sample sizes n_j increasing at the same rate, *i.e.*, $n_j/n \to \lambda_j > 0$ as $n \to \infty$, for $j = 1, \ldots, J$. We denote by $q_j(\alpha) := \inf\{x \in \mathbb{R} : \alpha \le F_j(x)\}$ the quantile of distribution F_j at probability level $\alpha \in (0, 1)$. The null hypothesis to be tested is

(H_0) For all $(j, j') \in \{1, \ldots, J\}^2$ with $j \neq j'$, we have $q_{j'}(\alpha)/q_j(\alpha) \to 1$ as $\alpha \to 1$.

The tail behavior of the J samples is summarized through $J \times L$ estimators $\widetilde{q}_j(\alpha_{\ell,n})$ of extreme quantiles $q_j(\alpha_{\ell,n})$, where $\alpha_{\ell,n} \to 1$ such that $n(1 - \alpha_{\ell,n}) \to \tau_\ell \ge 0$ as $n \to \infty$, $\ell = 1, \ldots, L$ and $j = 1, \ldots, J$. The above conditions imply

$$\mathbb{P}\left(\max_{i=1,\ldots,n_j} X^{(j)}_i \le q_j(\alpha_{\ell,n})\right) = \alpha^{n_j}_{\ell,n} \to \exp\left(-\lambda_j \tau_\ell\right), \quad n \to \infty,$$

and therefore such quantiles can lie beyond the observation range when τ_ℓ is small. Direct empirical estimation of such quantiles leads to a high estimation variance and/or bias, so that semi-parametric estimation is preferable. Often, the Weissman estimator [43] is used:

$$\hat{q}^{\mathrm{W}}_j(\alpha_{\ell,n} \,|\, \beta_{j,n}) = \widehat{q}_j(\beta_{j,n}) \left(\frac{1 - \beta_{j,n}}{1 - \alpha_{\ell,n}}\right)^{\widehat{\xi}_j(\beta_{j,n})}, \tag{2}$$

where

- $\beta_{j,n}$ are intermediate probability levels, *i.e.*, $\beta_{j,n} \to 1$ and $n(1-\beta_{j,n}) \to \infty$ as $n \to \infty$, for $j = 1, \ldots, J$;
- $\widehat{q}_j(\beta_{j,n}) = X^{(j)}_{\lfloor \beta_j n_j \rfloor, n_j}$ is an estimator of the intermediate quantile of level $\beta_{j,n}$ in E_j, $j = 1, \ldots, J$, and
- $\widehat{\xi}_j(\beta_{j,n})$ is an estimator of ξ_j, such as the Hill estimator [20] based on the $k_{j,n} := \lfloor (1 - \beta_{j,n}) n_j \rfloor$ largest observations from sample E_j, $j = 1, \ldots, J$:

$$\hat{\xi}^{\mathrm{H}}_j(\beta_{j,n}) = \frac{1}{k_{j,n}} \sum_{i=0}^{k_{j,n}-1} \log X^{(j)}_{n_j - i, n_j} - \log X^{(j)}_{n_j - k_{j,n}, n_j}. \tag{3}$$

We propose to additively decompose the variance of quantiles into contributions from the factor (*i.e.*, the sample in our setting) and from the different quantile levels. Moreover, we work in log-scale to achieve statistically stable behavior of differences between log-quantiles and their averages in the heavy-tailed setting. We write the overall sum-of-squares as

$$\Delta_n = \frac{1}{JL}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\left(\log \widetilde{q}_j(\alpha_{\ell,n}) - \mu_{\alpha,n}\right)^2, \quad \text{with } \mu_{\alpha,n} = \frac{1}{JL}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\log \widetilde{q}_j(\alpha_{\ell,n}).$$

The new test statistic, called ANOVEX (*ANalysis of Variability in EXtremes*), is constructed based on the decomposition of this extreme log-quantile variance into two parts: variance $\Delta_{1,n}$ due to the J different samples, and variance $\Delta_{2,n}$ due to the L different quantile levels. This decomposition is formalized in the following lemma and illustrated in Figure 1. We construct the ANOVEX test statistic to reject (H_0) if $\Delta_{1,n}/\Delta_{2,n}$ exceeds an appropriately fixed threshold, *i.e.*, if the contribution of the inter-sample variance to the total variance is very strong. Therefore, the test statistic is straightforward to interpret and to compute.

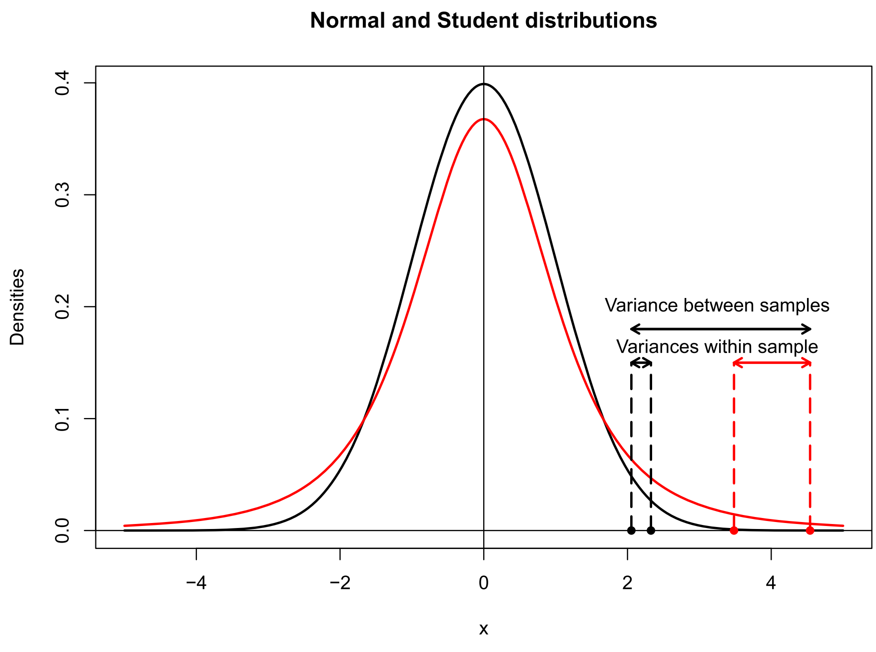

FIG 1. *Decomposition of extreme log-quantile variance for a standard normal (black curve) and a Student's t distribution* 3 *degrees of freedom (red curve), for two extreme quantiles at probability levels* $\{0.98, 0.99\}$.

Lemma 2.1. *The decomposition* $\Delta_n = \Delta_{1,n} + \Delta_{2,n}$ *holds where*

$$\Delta_{1,n} = \frac{1}{L}\sum_{\ell=1}^{L} \Delta_{1,\ell,n} \quad \textit{(variance due to the different samples)},$$

$$\Delta_{2,n} = \frac{1}{L}\sum_{\ell=1}^{L} \left(\mu_{\alpha,n}^{(\ell)} - \mu_{\alpha,n}\right)^2 \quad \textit{(variance due to the different quantile levels)},$$

with the components indexed by $\ell = 1, \dots, L$ *given as*

$$\mu_{\alpha,n}^{(\ell)} = \frac{1}{J}\sum_{j=1}^{J} \log \widetilde{q}_j(\alpha_{\ell,n}) \quad \textit{(mean estimation at level } \ell\textit{)},$$

$$\Delta_{1,\ell,n} = \frac{1}{J}\sum_{j=1}^{J} \left(\log \widetilde{q}_j(\alpha_{\ell,n}) - \mu_{\alpha,n}^{(\ell)}\right)^2 \quad \textit{(variance due to the different samples at level } \ell\textit{)}.$$

This lemma follows from the law of total variance (or Eve's law [4, Theorem 9.5.4]) applied to the finite probability space $\{1, \dots, J\} \times \{1, \dots, L\}$ equipped with the discrete uniform distribution.

3. Asymptotic distribution of the test statistic

We need the following notation defined for any positive $u_1, \dots, u_L$:

$$\text{smlog}(u_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L} (\log(u_\ell))^2, \text{ and}$$
$$\text{varlog}(u_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L} (\log(u_\ell))^2 - \left(\frac{1}{L}\sum_{\ell=1}^{L} \log(u_\ell)\right)^2 \tag{4}$$

which can be viewed respectively as the empirical second moment and variance of $(\log(u_1), \dots, \log(u_L))$. Our first main result establishes the asymptotic distribution of the test statistic under (H_0), which requires a deterministic rescaling factor.

Theorem 3.1. *Let* $E_1, \dots, E_J$ *be independent samples. Assume* $F_j \in \mathcal{C}_1(\xi_j)$ *for* $j = 1, \dots, J$ *and the following conditions:*

- *Extreme probability levels: For all* $\ell = 1, \dots, L$, $\alpha_{\ell,n} \to 1$ *as* $n \to \infty$.
- *Extreme quantile estimator:* $\widetilde{q}_j(\alpha_{\ell,n})$ *is an estimator of the extreme quantile* $q_j(\alpha_{\ell,n})$, *computed on the* n_j*-sample* E_j *where* $n_j/n \to \lambda_j > 0$ *as* $n \to \infty$, *and satisfies*

$$\bar{\sigma}_{\ell,n}^{-1} \left(\log\left(\widetilde{q}_j(\alpha_{\ell,n})\right) - \log\left(q_j(\alpha_{\ell,n})\right)\right) \xrightarrow{d} \xi_j Z_j, \tag{5}$$

where $Z_1, \dots, Z_J$ are independent standard Gaussian random variables and for deterministic normalizing sequences $\bar\sigma_{\ell,n} \to 0$ as $n \to \infty$, for all $j = 1, \dots, J$ and $\ell = 1, \dots, L$.

Then, under (H_0),

$$\frac{J\ \mathrm{varlog}((1-\alpha_n)_{1:L})}{\bar\sigma_n^2}\,\frac{\Delta_{1,n}}{\Delta_{2,n}} \xrightarrow{d} \chi^2_{J-1}, \tag{6}$$

with χ^2_{J-1} a chi-square random variable with $J-1$ degrees of freedom, where

$$\bar\sigma_n^2 = \frac{1}{L}\sum_{\ell=1}^{L} \bar\sigma_{\ell,n}^2.$$

If, moreover, $\alpha_{\ell,n} = 1 - \tau_\ell/n$ for $\ell = 1, \dots, L$, then (6) can be simplified as

$$\frac{J\ \mathrm{varlog}(\tau_{1:L})}{\bar\sigma_n^2}\,\frac{\Delta_{1,n}}{\Delta_{2,n}} \xrightarrow{d} \chi^2_{J-1}.$$

We emphasize that the independence of the limiting random variables Z_j is a consequence of the independence of the samples E_j, $j = 1, \dots, J$. Besides, the limiting distribution of most extreme quantile estimators is inherited from the limiting distribution of the tail-index estimator, see for instance [10, Theorem 4.3.8]. The assumption in (5) that the random variables Z_j do not depend on ℓ is thus non restrictive in general. Finally, for the results of Theorem 3.1 to hold we do not need independence of the variables in each of the samples E_j but only independence between the samples for different j. One can for instance refer to [23] for the asymptotic normality of tail-index estimators in the case of stationary sequences with short range dependence.

Remark 1 (Use of other risk measures)**.** *The result of Theorem 3.1 remains valid if the quantile is replaced by any other risk measure whose extreme estimator fulfills condition (5). Indeed, the Conditional Tail Expectation (also called Expected Shortfall, introduced in [1]), or the expectile [33], could be used, since they are asymptotically proportional to the extreme quantile [17, 15]. Here, we keep the focus on quantiles since they do not require a finite first moment.*

We next study an extreme quantile estimator satisfying the assumptions of Theorem 3.1. Condition (5) is fulfilled by combining the Weissman estimator (2) with the Hill estimator (3). However, a stronger second-order assumption, widely used in the extreme-value literature and satisfied by numerous distributions [3], is necessary. We denote it by $\mathcal{C}_2(\xi, \rho, A)$.

Definition 2 (Class $\mathcal{C}_2(\xi, \rho, A)$)**.** *The cumulative distribution function F belongs to the class $\mathcal{C}_2(\xi, \rho, A)$ with tail index $\xi > 0$ and second-order parameter $\rho < 0$, if there exists a measurable auxiliary function A with constant sign, satisfying $A(t) \to 0$ as $t \to \infty$, such that*

$$\lim_{t\to\infty} \frac{1}{A(1/\overline{F}(t))}\left(\frac{\overline{F}(ty)}{\overline{F}(t)} - y^{-1/\xi}\right) = y^{-1/\xi}\,\frac{y^{\rho/\xi}-1}{\xi\rho}, \quad \textit{for all } y > 0.$$

The function $|A|$ is regularly varying with index ρ. For small $\rho \le -1$, including the Pareto, Burr, Fréchet distributions, the distribution tail is very close to a Pareto tail, and the Weissman estimator is very accurate. For ρ near 0, as for the Generalized Pareto distribution with small $\xi = -\rho > 0$, the Weissman approximation remains valid but becomes accurate only relatively far in the tail.

From now on, we assume that the random variables in each E_j are mutually independent and identically distributed according to a cumulative distribution function F_j satisfying condition (1). Then, the combined Weissman-Hill estimator is asymptotically Gaussian under the second-order condition $\mathcal{C}_2(\xi, \rho, A)$, such that condition (5) of Theorem 3.1 holds (see the proof in Appendix), and we get the following result.

Corollary 3.1. *Let $E_1, \ldots, E_J$ be independent samples. Suppose that the random variables in each E_j are mutually independent and identically distributed according to a cumulative distribution function $F_j \in \mathcal{C}_2(\xi_j, \rho_j, A_j)$ for $j = 1, \ldots, J$ in Definition 2. Moreover, we assume the following conditions hold:*

- *Extreme probability levels: $\beta_n \to 1$ with $n(1-\beta_n) \to \infty$ and $\alpha_{\ell,n} \to 1$ for all $\ell = 1, \ldots, L$, as $n \to \infty$.*
- *Extreme quantile estimator: $\hat{q}_j^{W}(\alpha_{\ell,n} \,|\, \beta_{j,n})$ is the combined Weissman-Hill estimator (2), (3) of the extreme quantile $q_j(\alpha_{\ell,n})$ computed on the n_j-sample E_j such that, as $n \to \infty$ and for all $j = 1, \ldots, J$ and $\ell = 1, \ldots, L$,*

$$
\begin{aligned}
&n_j/n \to \lambda_j > 0,\\
&(1-\beta_{j,n})/(1-\beta_n) \to 1/\lambda_j,\\
&(1-\alpha_{\ell,n})/(1-\beta_n) \to 0,\\
&\sqrt{n(1-\beta_n)}/\log((1-\beta_n)/(1-\alpha_{\ell,n})) \to \infty.
\end{aligned}
$$

- *Second-order behavior: the auxiliary function A_j satisfies*

$$\sqrt{n(1-\beta_n)}A_j\left((1-\beta_n)^{-1}\right) \to 0, \quad n \to \infty, \quad \text{for all } j = 1, \ldots, J. \tag{7}$$

Then, under (H_0),

$$\frac{J\ \mathrm{varlog}((1-\alpha_n)_{1:L})\, n(1-\beta_n)}{\frac{1}{L}\sum_{\ell=1}^{L}\left(\log\left(\frac{1-\beta_n}{1-\alpha_{n,\ell}}\right)\right)^2}\,\frac{\Delta_{1,n}}{\Delta_{2,n}} \xrightarrow{d} \chi^2_{J-1}, \quad n \to \infty. \tag{8}$$

If, moreover, $\alpha_{\ell,n} = 1 - \tau_\ell/n$ for $\ell = 1, \ldots, L$, then (8) can be simplified as

$$\frac{J\,\mathrm{varlog}(\tau_{1:L})\, n(1-\beta_n)}{S_n(\beta_n, \tau_{1:L})}\,\frac{\Delta_{1,n}}{\Delta_{2,n}} \xrightarrow{d} \chi^2_{J-1}, \quad n \to \infty, \tag{9}$$

where

$$S_n(\beta_n, \tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L}\left(\log\left(\frac{n(1-\beta_n)}{\tau_\ell}\right)\right)^2. \tag{10}$$

Regarding the technical conditions involved in this result, we emphasize that the sample sizes $n_1, \dots, n_J$ must be asymptotically proportional. In practice, they should therefore not be too unbalanced, and n can be chosen as the average of n_j. The intermediate levels $\beta_{j,n}$ must be chosen accordingly, such that the number of order statistics $k_{j,n} = n_j(1-\beta_{j,n})$ remains asymptotically equivalent in all samples, which allows us to ensure that the normalizing constant $\bar{\sigma}_{\ell,n}^{-1}$ in (5) does not depend on the sample index j. The second-order condition (7) ensures that the Hill estimator is asymptotically unbiased. Bias-reduced versions of the Hill estimator with the same asymptotic variance exist [5], and this condition could then be dropped or replaced by a weaker condition on the auxiliary functions A_j.

The ANOVEX test rejects (H_0) with asymptotic level $\gamma \in (0,1)$ if

$$T_n := \frac{J \operatorname{varlog}(\tau_{1:L})\, n(1-\beta_n)}{S_n(\beta_n, \tau_{1:L})} \frac{\Delta_{1,n}}{\Delta_{2,n}} > \chi^2_{J-1,1-\gamma}, \tag{11}$$

where $\chi^2_{J-1,1-\gamma}$ denotes the quantile of level $1-\gamma$ of the chi-square distribution with $J-1$ degrees of freedom. The test is asymptotically equivalent to

$$\frac{J \operatorname{varlog}(\tau_{1:L})\, n(1-\beta_n)}{\left(\log\left(n(1-\beta_n)\right)\right)^2} \frac{\Delta_{1,n}}{\Delta_{2,n}} > \chi^2_{J-1,1-\gamma}.$$

For accuracy reasons, we adopt the first version (11) of T_n in the sequel.

4. Examples of type-I and type-II error approximations

4.1. Pareto-distribution setting

We analytically investigate the finite sample behavior of the test to study its power with data following the Pareto distribution, the prototype of regularly varying distributions. Numerical investigation into more general settings is carried out in the simulation study in Section 5. We leverage Rényi's representation [39] to provide accurate approximations of the distribution of the test statistic T_n, as well as type-I and type-II errors, for $J = 2$ independent samples. We write $\mathcal{P}(1/\xi)$ for the Pareto distribution with scale parameter 1 and shape parameter $1/\xi > 0$; its cumulative distribution function is $1 - x^{-1/\xi}$, $x > 1$. We consider three settings: identical Pareto distributions (Section 4.2); Pareto distributions with different scale parameters (Section 4.3); Pareto distributions with different shape parameters (Section 4.4). We treat the case where $\alpha_{\ell,n} = 1 - \tau_\ell/n$ for $\ell = 1, \dots, L$. In addition to the normalizing constant S_n in (10), we need the following constants:

$$s_n(\beta_n, \tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L} \sqrt{1 + \left(\log\left(\frac{n(1-\beta_n)}{\tau_\ell}\right)\right)^2},$$

$$\text{and } \mathfrak{s}_n(\beta_n, \tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L} \log\left(\frac{n}{\tau_\ell}\right) \sqrt{1 + \left(\log\left(\frac{n(1-\beta_n)}{\tau_\ell}\right)\right)^2}.$$

4.2. Identically distributed Pareto samples

If both samples follow the same Pareto distribution, the ANOVEX test is supposed to wrongly reject the null hypothesis with asymptotic probability γ (type-I error). We investigate the impact of test hyperparameters on this probability through a stochastic approximation of the test statistic.

Proposition 4.1. *Consider two independent samples $E_j = \{X_1^{(j)}, \dots, X_n^{(j)}\}$, $j = 1, 2$, of i.i.d. variables following the same Pareto distribution $\mathcal{P}(1/\xi)$, $\xi > 0$. Assume the following:*

- *(a) Probability levels: (β_n) is an intermediate probability level such that $(1 - \beta_n)\log(n) \to 0$ as $n \to 0$.*
- *(b) Extreme quantile estimator: $\hat{q}_j^{W}(\alpha_{\ell,n} \,|\, \beta_n)$ is the Weissman estimator (2) of the extreme quantile $q_j(\alpha_{\ell,n})$, based on the Hill estimator $\hat{\xi}_j^{H}(\beta_n)$ defined in (3), where $\alpha_{\ell,n} = 1 - \tau_\ell/n$, for $\ell = 1, \dots, L$ and $j = 1, 2$.*

Then, as $n \to \infty$,

$$T_n \stackrel{d}{=} \Gamma^2 \left(1 + \frac{1}{S_n(\beta_n, \tau_{1:L})}\right) \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(n(1-\beta_n))}\right)\right)$$

where Γ is a standard normal random variable.

This stochastic representation leads to a pretty accurate approximation of the rejection probability.

Remark 2. *Assume that conditions of Proposition 4.1 hold. The probability $\mathbb{P}_{H_0}\left(T_n > \chi^2_{1,1-\gamma}\right)$ to wrongly reject (H_0) with asymptotic level $\gamma \in (0, 1)$ is for large n approximately equal to*

$$p_n(\gamma) = 2\bar{\Phi}\left(\bar{\Phi}^{-1}(\gamma/2)\left(1 + \frac{1}{S_n(\beta_n, \tau_{1:L})}\right)^{-1/2}\right), \tag{12}$$

where $\overline{\Phi}(\cdot)$ is the standard Gaussian survival function.

According to (12), the behavior of $p_n(\gamma)$ is driven by $S_n(\beta_n, \tau_{1:L})$ but does not depend on the tail index ξ. Clearly, $p_n(\gamma) \xrightarrow{>} \gamma$ as $n \to \infty$ in view of $S_n(\beta_n, \tau_{1:L}) \sim (\log(n(1 - \beta_n)))^2$ as $n \to \infty$. Indeed, a first-order expansion yields

$$p_n(\gamma) = \gamma + \frac{\bar{\Phi}^{-1}(\gamma/2)\varphi(\bar{\Phi}^{-1}(\gamma/2))}{(\log(n(1-\beta_n)))^2}(1 + o(1)),$$

where $\varphi(\cdot)$ is the density of the standard Gaussian distribution. Therefore, we obtain a logarithmic rate of convergence with respect to $n(1-\beta_n)$, the expected number of observations above $q_1(\beta_n) = q_2(\beta_n)$. The accuracy of approximation (12) for Pareto distributions $\mathcal{P}(1/\xi = 4)$ is illustrated in the top left panel of Figure 2, for $n = 1{,}000$ and $L = 2, \dots, 30$. The first-order approximation (12) (solid blue curve) is compared to the empirical estimation of $p_n(\gamma = 0.05)$ based on $N = 10{,}000$ replications (dashed blue curve); see Section 5 for details. The approximation is fairly precise, especially for moderate values of L, with only a small bias of observed type-I errors with respect to the nominal level γ. As expected, $p_n(\gamma) \geq \gamma$ (horizontal black line for γ). The true $p_n(\gamma)$ does not seem to depend on ξ in these numerical experiments, in line with the analytical approximation (12). Finally, under the weak condition $\log(n(1-\beta_n)) \geq 3$, the quantity $S_n(\beta_n, \tau_{1:L})$ is a decreasing function of L, so that the approximation (12) is increasing with L if $\tau_\ell = \ell$. The practical choice of L is discussed in Section 5.

4.3. Pareto samples with different scale parameters

We consider two Pareto samples with the same shape parameter $1/\xi$, $\xi > 0$, but different scale parameters 1 and $\lambda_n > 0$, and we assume $\lambda_n \xrightarrow{\neq} 1$ as $n \to \infty$ to study the power of the ANOVEX test to reject (H_0). We provide an approximation of the distribution of the test statistic.

Proposition 4.2. *Suppose that the assumptions (a, b) of Proposition 4.1 hold. Consider two independent samples denoted by $E_1 = \{X_1^{(1)}, \dots, X_n^{(1)}\}$ and $E_2 = \{X_1^{(2)}, \dots, X_n^{(2)}\}$, where, for $i = 1, \dots, n$, the $X_i^{(1)}$s are i.i.d. from a Pareto distribution $\mathcal{P}(1/\xi)$, $\xi > 0$ and the $X_i^{(2)}$s are i.i.d. with $X_i^{(2)} \stackrel{d}{=} \lambda_n X_i^{(1)}$, $\lambda_n > 0$, $\lambda_n \to 1$ as $n \to \infty$ such that*

(i) $\frac{\log(n(1-\beta_n))}{(n(1-\beta_n))^{3/4}} \vee \sqrt{\frac{\log(n(1-\beta_n))}{n}} = o(\log(\lambda_n))$,

(ii) $\log(\lambda_n) = o(1/\sqrt{n(1-\beta_n)})$.

Then, as $n \to \infty$,

$$T_n \stackrel{d}{=} \left(\frac{(\log(\lambda_n))^2 n(1-\beta_n)}{2\xi^2 S_n(\beta_n, \tau_{1:L})} - \frac{\sqrt{2n(1-\beta_n)}\log(\lambda_n) s_n(\beta_n, \tau_{1:L})}{\xi S_n(\beta_n, \tau_{1:L})}\Gamma \right.$$
$$\left. + \frac{1 + S_n(\beta_n, \tau_{1:L})}{S_n(\beta_n, \tau_{1:L})}\Gamma^2 \right) \times \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(n(1-\beta_n))} \right) \right)$$

where Γ is a standard normal random variable.

This result yields an approximation of the type-II error when the alternative hypothesis is formulated as

$(H_{1,n})$ $X_i^{(1)} \sim \mathcal{P}(1/\xi)$ and $X_i^{(2)} \stackrel{d}{=} \lambda_n X_i^{(1)}$, $i = 1, \dots, n$ and $\lambda_n \to 1$ as $n \to \infty$.

Remark 3. *If assumption (i) of Proposition 4.2 is replaced by the slightly stronger condition*

(i') $(\log(n(1-\beta_n)))^2/(n(1-\beta_n))^{3/4} \vee \sqrt{\frac{(\log(n(1-\beta_n)))^3}{n}} = o(\log(\lambda_n))$,

then the probability $\mathbb{P}_{H_{1,n}}\left(T_n \leq \chi^2_{1,1-\gamma}\right)$ to accept (H_0) with asymptotic level $\gamma \in (0,1)$ is for large n approximately equal to

$$\bar{\Phi}\left(\Omega_{1,n} - \sqrt{\Omega_{2,n}}\right) - \bar{\Phi}\left(\Omega_{1,n} + \sqrt{\Omega_{2,n}}\right) \tag{13}$$

where

$$\Omega_{1,n} = \frac{\log(\lambda_n)\sqrt{n(1-\beta_n)}s_n(\beta_n,\tau_{1:L})}{\sqrt{2}\xi\left(1+S_n(\beta_n,\tau_{1:L})\right)},$$

$$\Omega_{2,n} = \frac{\left(\log(\lambda_n)\right)^2 n(1-\beta_n)}{2\xi^2}\frac{s_n(\beta_n,\tau_{1:L})^2-1-S_n(\beta_n,\tau_{1:L})}{\left(1+S_n(\beta_n,\tau_{1:L})\right)^2} + \frac{S_n(\beta_n,\tau_{1:L})}{1+S_n(\beta_n,\tau_{1:L})}\chi^2_{1,1-\gamma} > 0.$$

First, we have $\Omega_{1,n} \to 0$ and $\Omega_{2,n} \to \chi^2_{1,1-\gamma}$ as $n \to \infty$; see the proof of Remark 3 in the Appendix. Thus, $\mathbb{P}_{H_{1,n}}\left(T_n \leq \chi^2_{1,1-\gamma}\right) \to 1-\gamma$ as $n \to \infty$, and the asymptotic type-II error is $1-\gamma$ when $(H_{1,n})$ approaches (H_0). Second, considering $\lambda_n = 1$ yields

$$\begin{aligned}
&\bar{\Phi}\left(\Omega_{1,n} - \sqrt{\Omega_{2,n}}\right) - \bar{\Phi}\left(\Omega_{1,n} + \sqrt{\Omega_{2,n}}\right) \\
&= 1 - 2\bar{\Phi}\left(\bar{\Phi}^{-1}(\gamma/2)\left(1+\frac{1}{S_n(\beta_n,\tau_{1:L})}\right)^{-1/2}\right) \\
&= 1 - p_n(\gamma),
\end{aligned}$$

see (12), which is in accordance with the type-I error in Section 4.2. Third, in the converse case where λ_n is large, *i.e.* when the two Pareto distributions are very different, the approximation (13) tends to zero, and the ANOVEX test is likely to reject (H_0). The same reasoning may be applied to $\lambda_n^{1/\xi}$, which is the key quantity in $\Omega_{1,n}$ and $\Omega_{2,n}$. For a fixed value of λ_n, the approximated probability is a decreasing function of ξ: Heavy tails are thus more easily discerned. Adopting the classical choice $\beta_n = 1 - c/\sqrt{n}$, $c > 0$, conditions of Proposition 4.2 and Remark 3 imply that λ_n converges to 1 not faster than $n^{-3/8}$, up to a logarithmic factor. This may interpreted as the minimum gap between (H_0) and $(H_{1,n})$ that the ANOVEX test is able to discriminate.

The accuracy of the approximation given in Remark 3 is illustrated in the second row of Figure 2 in Section 5 for $\lambda_n = 1.2$ with $\xi \in \{0.15; 0.25; 0.35; 0.50\}$ (left panel), $\xi = 0.25$ with $\lambda_n \in \{1.1; 1.2; 1.3; 1.4\}$ (middle panel) and $\lambda_n = \lambda$ (fixed) $\in \{1.2; 1.5\}$ with $n \in \{500; 1,000; 5,000\}$ (right panel). The approximation (13) is remarkably accurate throughout. The numerical results also confirm the roles of λ_n and ξ in the type-II error.

4.4. Pareto samples with different shape parameters

For two Pareto samples with different shape parameters $1/\xi$ and $1/(\xi\theta_n)$, $\xi, \theta_n > 0$, we investigate the ability of the ANOVEX test to reject (H_0) when $\theta_n \xrightarrow{\neq} 1$ as $n \to \infty$. We provide an approximation of the test statistic distribution.

Proposition 4.3. *Suppose that the assumptions* (a,b) *of Proposition 4.1 hold with the additional condition that* $\log(n) = O(\log(n(1-\beta_n)))$. *Consider two* *independent samples* $E_1 = \{X_1^{(1)}, \dots, X_n^{(1)}\}$ *and* $E_2 = \{X_1^{(2)}, \dots, X_n^{(2)}\}$*, where, for* $i = 1, \dots, n$*, the* $X_i^{(1)}$*s are i.i.d. from a Pareto distribution* $\mathcal{P}(1/\xi)$*,* $\xi > 0$ *and the* $X_i^{(2)}$*s are i.i.d. with* $X_i^{(2)} \overset{d}{=} (X_i^{(1)})^{\theta_n}$*,* $\theta_n > 0$ *and* $\theta_n \to 1$ *as* $n \to \infty$*, where* θ_n *satisfies the same conditions (i) and (ii) as* λ_n *in Proposition 4.2. Then, as* $n \to \infty$*,*

$$\begin{aligned} T_n \overset{d}{=}\ & 2\left(\frac{n(1-\beta_n)(1-\theta_n)^2}{(1+\theta_n)^2}\frac{\operatorname{smlog}(n/\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}\right. \\ & \left. +2\frac{\sqrt{n(1-\beta_n)}\sqrt{1+\theta_n^2}(1-\theta_n)}{(1+\theta_n)^2}\frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}\Gamma + \frac{(1+\theta_n^2)}{(1+\theta_n)^2}\frac{1+S_n(\beta_n,\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}\Gamma^2\right) \\ & \times\left(1 + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(n(1-\beta_n))}\right) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)\right), \end{aligned}$$

where Γ *is a standard normal random variable, and* $\operatorname{smlog}(\cdot)$ *is defined in Equation (4).*

The same choices can be made for θ_n as for λ_n, leading to similar interpretations. This result allows approximating the type-II error associated with the ANOVEX test when the alternative hypothesis is

$(H'_{1,n})$ $X_i^{(1)} \sim \mathcal{P}(1/\xi)$ and $X_i^{(2)} \overset{d}{=} (X_i^{(1)})^{\theta_n}$, $i = 1, \dots, n$ and $\theta_n \to 1$ as $n \to \infty$.

Remark 4. *If assumption (i) in Proposition 4.2 (with* λ_n *instead of* θ_n*) is replaced by the slightly stronger condition (i') as in Remark 3, then the probability* $\mathbb{P}_{H'_{1,n}}\left(T_n \le \chi^2_{1,1-\gamma}\right)$ *to accept* (H_0) *with asymptotic level* $\gamma \in (0,1)$ *may be approximated for* n *large enough by*

$$\bar{\Phi}\left(\Psi_{1,n} - \sqrt{\Psi_{2,n}}\right) - \bar{\Phi}\left(\Psi_{1,n} + \sqrt{\Psi_{2,n}}\right) \tag{14}$$

where

$$\begin{aligned} \Psi_{1,n} &= \frac{\sqrt{n(1-\beta_n)}(\theta_n - 1)\mathfrak{s}_n(\beta_n,\tau_{1:L})}{\sqrt{1+\theta_n^2}\,(1+S_n(\beta_n,\tau_{1:L}))}, \\ \Psi_{2,n} &= \frac{(\theta_n-1)^2 n(1-\beta_n)}{(1+\theta_n^2)}\frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})^2 - (1+S_n(\beta_n,\tau_{1:L}))\operatorname{smlog}(n/\tau_{1:L})}{(1+S_n(\beta_n,\tau_{1:L}))^2} \\ &\quad + \frac{(1+\theta_n)^2}{(1+\theta_n^2)}\frac{S_n(\beta_n,\tau_{1:L})}{1+S_n(\beta_n,\tau_{1:L})}\frac{\chi^2_{1,1-\gamma}}{2} > 0. \end{aligned}$$

Again, $\mathbb{P}_{H'_{1,n}}\left(T_n \le \chi^2_{1,1-\gamma}\right)$ tends to $1-\gamma$ as $n \to \infty$. Moreover, taking $\theta_n = 1$ leads to $1 - p_n(\gamma)$. However, unlike the approximation of Proposition 4.2, this probability is not related to the tail index ξ. The accuracy of this approximation for several values of θ_n is illustrated on the top middle panel of Figure 2 in Section 5 for $\xi = 0.25$ and $\theta_n \in \{1.1; 1.2; 1.3; 1.4\}$. Another example with a fixed $\theta_n = \theta \in \{1.1; 1.3\}$ and a varying $n \in \{500; 1,000; 5,000\}$ is proposed (top right panel). Again, the curves associated with the approximated probability (14) are almost identical to the curves associated with the empirical type-II errors.

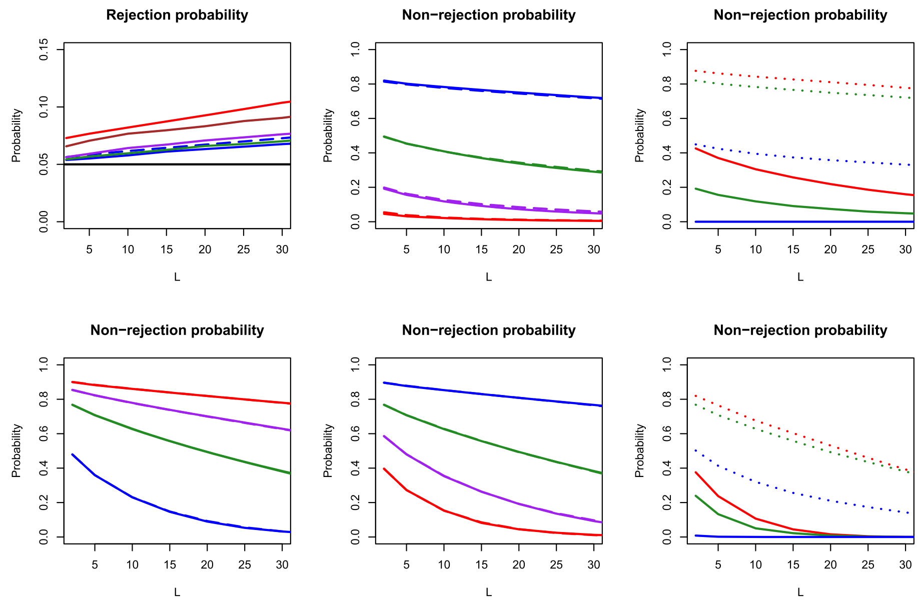

FIG 2. *Empirical (solid curves) and approximated (dashed curves) type I (top left panel) and type II (other panels) errors obtained for 10,000 replications, as shown as functions of L. Top left: Example ($\mathcal{P}$) with $\xi = 0.25$ and: $n_1 = n_2 = 1{,}000$ (blue), $n_1 = 1.5n_2 = 1,200$ (green), $n_1 \approx 2n_2 = 1,333$ (purple), $n_1 = 3n_2 = 1,500$ (brown) and $n_1 = 4n_2 = 1,600$ (red). Top middle: Example ($\mathcal{P}_\theta$) with $\xi = 0.25$ and $\theta_n = 1.1$ (green curves), $\theta_n = 1.2$ (brown curves), $\theta_n = 1.3$ (blue curves) and $\theta_n = 1.4$ (red curves). Top right: same setting with a fixed $\theta = 1.1$ (dotted curves) or 1.3 (solid curves), and $n = 500$ (red), 1,000 (green) or 5,000 (blue). Bottom left: Example ($\mathcal{P}_\lambda$) with $\lambda_n = 1.2$ and $\xi = 0.15$ (blue curves), $\xi = 0.25$ (green curves), $\xi = 0.35$ (purple curves) and $\xi = 0.5$ (red curves). Bottom middle: same setting with $\xi = 0.25$ and $\lambda_n = 1.1$ (blue curves), $\lambda_n = 1.2$ (green curves), $\lambda_n = 1.3$ (purple curves) and $\lambda_n = 1.4$ (red curves). Bottom right: same setting with a fixed $\lambda = 1.2$ (dotted curves) or 1.5 (solid curves), and $n = 500$ (red), 1,000 (green) or 5,000 (blue). In all examples, $n = 1{,}000$, $\beta_n = 0.9$, $\gamma = 0.05$ and $\alpha_{\ell,n} = 1 - \ell/n$, $\ell = 1, \ldots, L$.*

5. Simulation study

For all simulations in this section, the extreme quantile estimator we use is the Weissman estimator (2) combined with the Hill estimator (3). For each example, we simulate $N = 10{,}000$ times J samples of size $n = 1{,}000$, compute the test statistic (11) to assess (H_0) at confidence level $\gamma = 0.05$. We take $\alpha_{\ell,n} = 1 - \ell/n$ ($\tau_\ell = \ell$) for $\ell = 1, \ldots, L$, consider values of L ranging from two to 30, and report empirical rejection (or equivalently non-rejection) probabilities obtained through the N replications. Several heavy-tailed distributions with tail index $\xi > 0$ are considered:

- The Pareto distribution with s.f. $\overline{F}(x) = x^{-1/\xi}$ for $x > 1$;
- The Generalized Pareto distribution (GPD) with s.f. $\overline{F}(x) = (1 + \xi x)^{-1/\xi}$ for $x > 0$. This distribution fulfills Assumption $\mathcal{C}_2(\xi, \rho, A)$ with $\rho = -\xi$;

- The Fréchet distribution with s.f. $\overline{F}(x) = 1 - \exp(-x^{-1/\xi})$ for $x > 0$. Assumption $\mathcal{C}_2(\xi, \rho, A)$ is also satisfied with $\rho = -1$;
- The Burr(ρ) distribution with s.f. $\overline{F}(x) = (1 + x^{-\rho/\xi})^{1/\rho}$ for $x > 0$. In this case, $\mathcal{C}_2(\xi, \rho, A)$ is satisfied for any $\rho < 0$.
- The Student distribution with $\nu > 0$ degrees of freedom. Assumption $\mathcal{C}_2(\xi, \rho, A)$ is fulfilled with $\xi = 1/\nu$ and $\rho = -2/\nu$.

Firstly, we focus on the simplest case with $J = 2$ samples of random variables X_1 and X_2. In line with Sections 4.2, 4.3 and 4.4, we propose the following examples involving Pareto distributions:

($\mathcal{P}$) As in Proposition 4.1, $X_1 \overset{d}{=} X_2 = \mathcal{P}(1/\xi)$, *i.e.* both samples are i.i.d. replications of a Pareto distribution with $\xi = 0.25$. In this setting, we are supposed to reject (H_0) with probability $\gamma = 0.05$. We consider the cases of balanced $(n_1 = n_2)$ and unbalanced (several values of n_1/n_2 with fixed $n_1 + n_2 = 2n$) data. Note that in order to be in line with the conditions of Corollary 3.1, we take $1 - \beta_{j,n} = n(1 - \beta_n)/n_j$, $j \in \{1, 2\}$.

($\mathcal{P}_\lambda$) As in Proposition 4.2, X_1 follows a Pareto distribution, and $X_2 \overset{d}{=} \lambda_n X_1$, where (λ_n) fulfills the conditions of Proposition 4.2. This case is interesting (and pretty complicated) since the extreme quantiles of both distributions asymptotically coincide (as $n \to \infty$). However, in our finite sample size setting, $\lambda_n \neq 1$ and we expect the test to reject (H_0). The empirical non-rejection probabilities $\mathbb{P}_{H_{1,n}}\left(T_n \leq \chi^2_{1,0.95}\right)$ as well as their approximations calculated in Section 4.3 are computed for $\xi \in \{0.15; 0.25; 0.35; 0.5\}$, $\lambda_n = 1.2\ (= 1+2/n^{1/3})$ and $\xi = 0.25$, $\lambda_n \in \{1.1; 1.2; 1.3; 1.4\}$. We finally propose to fix $\lambda \in \{1.2; 1.5\}$ and compute the empirical non-rejection probabilities when the sample size $n \in \{500; 1,000; 5,000\}$ varies (with $\xi = 0.25$ and $\beta_n \equiv 0.9$).

($\mathcal{P}_\theta$) As in Proposition 4.3, X_1 follows a Pareto distribution, and $X_2 \overset{d}{=} X_1^{\theta_n}$, where (θ_n) fulfills the conditions of Proposition 4.3. Here also, the two distributions (and quantiles) are asymptotically the same, but are slightly different with $n = 1,000 < \infty$. The empirical non-rejection probabilities $\mathbb{P}_{H'_{1,n}}\left(T_n \leq \chi^2_{1,0.95}\right)$ as well as their approximations calculated in Section 4.4 are computed for $\xi = 0.25$ and $\theta_n = 1 + k/n^{1/3}$, $1 \leq k \leq 4$. We finally propose to fix $\theta \in \{1.1; 1.3\}$ and compute the empirical non-rejection probabilities when the sample size $n \in \{500; 1,000; 5,000\}$ varies (with $\xi = 0.25$ and $\beta_n \equiv 0.9$).

We limited ourselves to $\beta_n = 0.9$ ($\approx 1 - 3/\sqrt{n}$ when $n = 1{,}000$) everywhere for convenience, in order to fulfill the assumptions of Propositions 4.1, 4.2 and 4.3. The results are reported in Figure 2.

It clearly appears that the rejection rate is increasing with L. Moreover, the rejection rate tends to be slightly greater when the two sample sizes differ. For only moderately imbalanced samples, the results remain however very close to the balanced case (the imbalanced case is thus omitted in the following simulations). Finally, the type-I and type-II error approximations calculated in Propositions 4.1, 4.2 and 4.3 are remarkably accurate; see Figure 2. We now expand the framework of this simulation study by considering other heavy-tailed distributions (*i.e.*, belonging to the Fréchet MDA). For that purpose, we propose to study the following situations (still with $J = 2$ samples):

($\mathcal{F}_1$) $X_1 \stackrel{d}{=} X_2$, *i.e.*, both samples are i.i.d. replications of a heavy-tailed distribution. In this setting, we are supposed to reject (H_0) with probability $\gamma = 0.05$. Several distributions are considered: Burr (with $\rho = -1$), Fréchet, GPD and Student, with $\xi = 0.25$ and $\xi = 1$.

($\mathcal{F}_2$) X_1 follows a Fréchet distribution, and X_2 follows a Burr distribution with $\rho = -5$, -1 or -0.5, and $\xi = 0.25$ in both cases (the empirical results are not sensitive to the value of ξ here). In this context, the two extreme quantiles are asymptotically equivalent, and (H_0) should be rejected with probability $\gamma = 0.05$.

($\mathcal{F}_3$) X_1 and X_2 follow Fréchet distributions with different shape parameters. The null (H_0) thus has to be rejected.

($\mathcal{F}_4$) X_1 follows a GPD distribution, and X_2 follows a Student distribution. We choose among $\xi \in \{0.25, 0.5, 0.75, 1\}$ for both distributions, leading to a situation where the shapes are identical, but not the scales. The null (H_0) is thus supposed to be rejected.

($\mathcal{F}_5$) X_1 and X_2 follow Student distributions with different degrees of freedom. The tail indices are therefore different, and (H_0) has to be rejected.

The results are reported in Figure 3.
Interestingly, the type I error seems to be decreasing with ρ. Indeed, for small values of ρ (let us say $\rho \leq -1$), the choice of $L = 2$ seems to be the best calibrated one in terms of type I error. However, when ρ is close to 0, the rejection rate is too low if L is small, and a choice of a large L is thus more suited. Through additional simulations (considering $\tau_\ell = \ell$ and several Burr distributions with different values of ρ or Generalized Pareto distributions), we observed that for $n = 1{,}000$, a choice of $L \approx 20$ seems to be optimal when $\rho = -0.75$ (*i.e.* the rejection rate is around 5%). Similarly, $L \approx 50$ is tailored when $\rho = -0.5$, and $L \approx 80$ in the challenging case $\rho = -0.25$. The results are similar for Burr and Generalized Pareto distributions, and are apparently only sensitive to ρ, and not to the distribution itself. A prior estimation of ρ (using for instance the estimators of [19]) may thus be useful to select the parameter L yielding the best calibration of the type I error. Results for examples with $J > 2$ are reported in Appendix B.

6. Applications on real data

Throughout this section and as before, the extreme quantile estimator we use is the Weissman-Hill estimator.

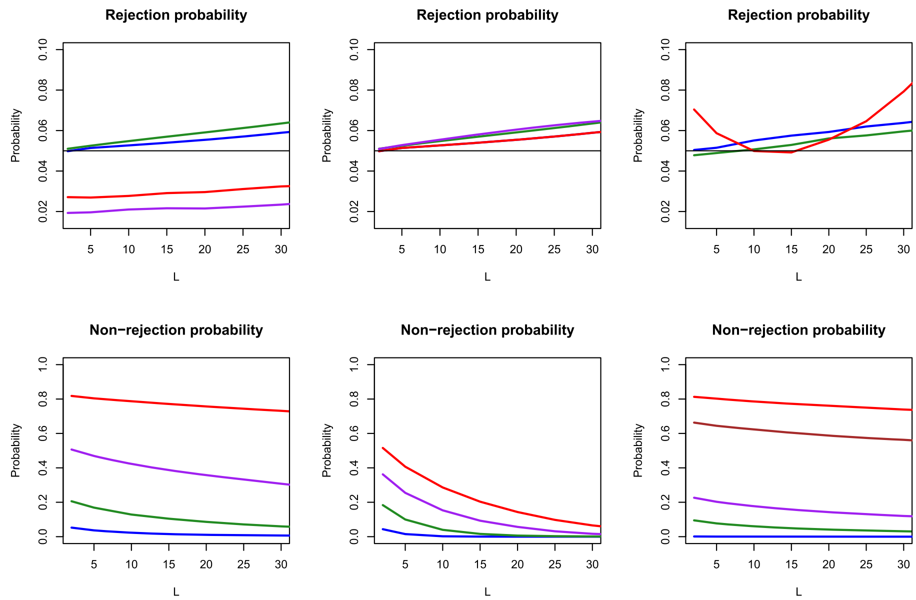

Fig 3. *Empirical type-I (top panels) and type-II (bottom panels) errors obtained for 10,000 replications, shown as functions of L. Top row, left and middle displays: Example ($\mathcal{F}_1$) with $\xi = 0.25$ (left) and 1 (middle). The underlying distribution is a Burr with $\rho = -1$ (blue), Fréchet (green), GPD (red) and Student (purple) distribution. Top row, right display: Example ($\mathcal{F}_2$) with $\rho = -5$ (blue), -1 (green) and -0.5 (red). Bottom row, left display: Example ($\mathcal{F}_3$) where the shape parameters are 0.25 vs 0.275 (red), 0.25 vs 0.3 (purple), 0.25 vs 0.325 (green) and 0.25 vs 0.35 (blue). Bottom row, middle display: Example ($\mathcal{F}_4$) with $\xi = 0.25$ (blue), 0.5 (green), 0.75 (purple) and 1 (red). Bottom row, right display: Example ($\mathcal{F}_5$), where the degrees of freedom are 2 vs 1 (blue), 2 vs 4 (green), 2 vs 4/3 (purple), 10/6 vs 4/3 (brown) and 2 vs 10/6 (red).*

6.1. Analysis of stock market indices

We consider negative daily log-returns for 11 emerging stock market indices, namely BIST 100 (Türkiye), IBOVESPA (Brazil), IPC Mexico, KOSPI Composite (South Korea), MOEX Russia, PSEi (Philippines), S&P BSE 500 (India), S&P MERVAL (Argentina), SSE Composite (China), TA-125 (Israel) and Tadawul All Shares (Saudi Arabia), to assess performance differences across different stock indices. We also added the European Euro Stoxx 50 in the study, leading to $J = 12$ samples. We have chosen these indices because, unlike the major stock indices (CAC 40, DAX, NASDAQ or Nikkei 225, for instance), they are only weakly cross-correlated (as shown in the sequel and in Figure 4). This justifies the use of the ANOVEX procedure, where the samples $E_1, \ldots, E_J$ are assumed to be independent. Similarly to [25, 32], we focus on the tail behavior of negative daily log-returns for different stock indices and apply the ANOVEX procedure to test whether their tails are equal. Data were collected on Yahoo Finance by taking the last 1,002 daily adjusted closing prices before

TABLE 1

Application to stock market indices. Columns report the empirical tail index and intermediate β_n^{th} quantile of the $n = 1{,}000$ negative log-returns (pre-filtered with an ARMA(1,1)-GARCH(1,1) model) for each financial index.

Stock index	$\widehat{\xi}_j^H(\beta_n)$	$\widehat{q}_j^H(\beta_n)$
BIST 100	0.325	1.181
Euro Stoxx 50	0.473	1.240
IBOVESPA	0.341	1.292
IPC Mexico	0.335	1.249
KOSPI Composite	0.327	1.216
MOEX Russia	0.436	1.160
PSEi	0.403	1.205
S&P BSE 500	0.309	1.221
S&P MERVAL	0.515	0.754
SSE Composite	0.390	1.238
TA-125	0.430	1.230
Tadawul All Shares	0.422	1.087

June 16, 2023 (included), leading to 12 samples $Y_1^{(j)}, \dots, Y_T^{(j)}$, $j = 1, \dots, 12$ of $T = 1,001$ log-returns. Obviously, such time series data are not serially independent. A solution to eliminate this dependence is to filter the time series with an ARMA(1,1)–GARCH(1,1) model:

$$Y_t^{(j)} = \mu^{(j)} + \phi^{(j)} Y_{t-1}^{(j)} + u_t^{(j)} + \theta^{(j)} u_{t-1}^{(j)},$$

where $u_t^{(j)} = \sigma_t^{(j)} \varepsilon_t^{(j)}$ is such that $\sigma_t^{(j)} = \sqrt{\mathfrak{c}^{(j)} + \mathfrak{a}^{(j)}(u_{t-1}^{(j)})^2 + \mathfrak{b}^{(j)}(\sigma_{t-1}^{(j)})^2}$, with $(\varepsilon_t^{(j)})$ an unobserved independent nondegenerate white noise sequence, *i.e.*, copies of a random variable ε with zero mean and unit variance, and the constants $\mu^{(j)}$, $\phi^{(j)}$, $\theta^{(j)}$, $\mathfrak{a}^{(j)}$, $\mathfrak{b}^{(j)}$ and $\mathfrak{c}^{(j)}$ the model parameters. We thus estimate these parameters using the `R` functions `garchFit` (available in the package `fGarch`, and its option `@residuals`, which provides the estimated values of $(u_t^{(j)})_t$) and `garch` (available in the package `tseries`), to estimate, with the option `$residuals`, the values of $(\varepsilon_t^{(j)})_t$ from the series of $(u_t^{(j)})_t$. This procedure finally leads to $J = 12$ (estimated) samples $\widehat{\varepsilon}_1^{(j)}, \dots, \widehat{\varepsilon}_n^{(j)}$ $(j = 1, \dots, J)$ of size $n = 1{,}000$. Data are summarized in Table 1, Figures 8 (Appendix) and 4 (where it is shown that the data is only weakly cross-correlated, the largest correlation coefficient being smaller than 0.067 in absolute value). Moreover, Figure 9 shows exponential QQ-plots of weighted log-spacings $i \log(\widehat{\varepsilon}_{n-i+1,n}^{(j)} / \widehat{\varepsilon}_{n-i,n}^{(j)})$, for $i = 1, \dots, 100$ and $j = 1, \dots, J$. The heavy-tail assumption seems valid here since all QQ-plots are close to a straight line with slope $\widehat{\xi}_j^H(\beta_n)$. We consider $\beta_n = 0.9$ and $\alpha_{\ell,n} = 1 - \ell/n$ for all $\ell = 1, \dots, L$ in the sequel.

We now study the proximity between the tails of these stock market log-returns. We compute the ANOVEX statistic for each pair of samples, with a suitable choice of L. Using the R function `mop` in the `Expectrem` package [8], the second-order parameter ρ is estimated at $\hat{\rho} \simeq -0.7$ in the twelve samples. Moreover, all tail indices ξ_j seem to lie around 0.4. We thus perform a simulation study with twelve samples of size 1,000 following a Burr distribution with $\xi = 0.4$

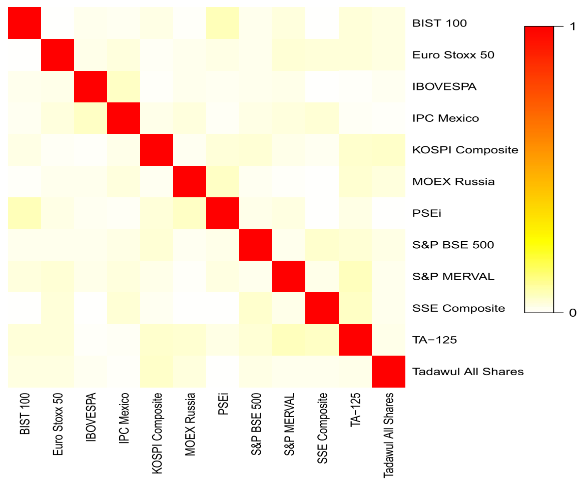

Fig 4. *Correlation matrix (absolute values) between 12 series of negative daily log-returns of stock market indices (pre-filtered by an ARMA(1,1)-GARCH(1,1) process).*

and $\rho = -0.7$, mimicking the estimated tail index and second-order parameter of the log-return samples. It appears that $L = 25$ provides a type I error around $\gamma = 0.05$, and we therefore set $L = 25$. Figure 5 gives an overview of the obtained results (the ANOVEX statistic being used as a dissimilarity measure to construct the tree). Finally, we can select the optimal number of groups by hierarchically applying the ANOVEX test (following the hierarchy illustrated in Figure 5 and with a given confidence level γ):

- Step 1: Compute all the ANOVEX tests between the neighboring samples (in the sense of the tree). When not rejected, they constitute subgroups at level 1. The potential subgroups at level 1 are thus SSE Composite/PSEi, Tadawul All Shares, TA-125/MOEX Russia, Euro Stoxx 50, IPC Mexico/IBOVESPA, S&P BSE 500/BIST 100, KOSPI Composite and S&P MERVAL.
- Step 2: Compute all the ANOVEX tests between the neighboring subgroups at level 1. When not rejected, they constitute subgroups at level 2. The potential subgroups at level 2 are SSE Composite/PSEi/ Tadawul All Shares, TA-125/MOEX Russia, Euro Stoxx 50, IPC Mexico/IBOVESPA, S&P BSE 500/BIST 100/KOSPI Composite and S&P MERVAL.
- ...iterate...
- Step 5: If the last two ANOVEX tests (between the two subgroups at level 4) are not rejected, then compute the ANOVEX test between all the samples. If the latter is rejected, we distinguish two groups (Group 1: SSE Composite, PSEi, Tadawul All Shares, TA-125, MOEX Russia and Euro Stoxx 50 and Group 2: IPC Mexico, IBOVESPA, S&P BSE 500, BIST 100, KOSPI Composite and S&P MERVAL).

For instance, a confidence level of 95% ($\gamma = 0.05$, and L fixed at 25 at each step) leads to the clustering into two groups, as mentioned in Step 5 above. It is interesting to notice this procedure tends to split the samples mainly into two groups: a first one containing the indices from Central and South America (and the Indian, Korean and Turkish indices as well), and a second one containing the other Eurasian indices (including the Euro Stoxx 50).

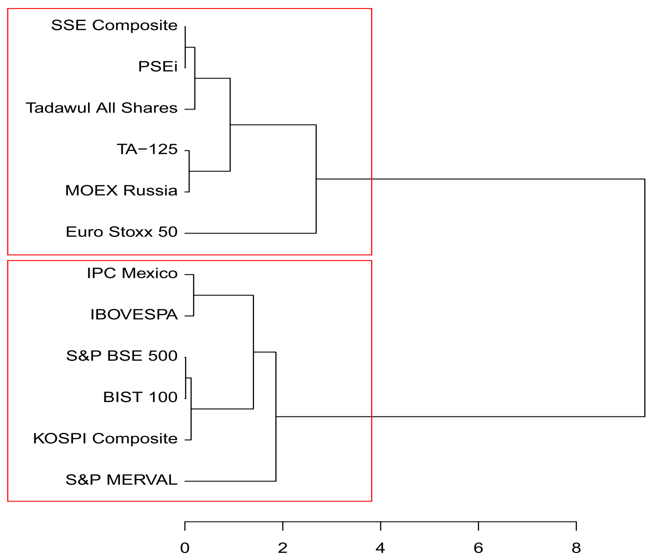

Fig 5. *Dendrogram of $J = 12$ market stock indices, where the dissimilarity measure is the ANOVEX test statistic (and the linkage method is the complete-linkage) applied to their negative daily log-returns (pre-filtered by an ARMA(1,1)–GARCH(1,1) process).*

6.2. Analysis of daily precipitation in Germany

To further illustrate the behavior of ANOVEX for $J > 2$ groups, we use the test to detect nonstationary behavior in extremes of daily accumulated precipitation across $J = 6$ decades for observations collected at weather stations in Germany. Heavy-tailedness is generally accepted for precipitation measurements at daily or higher temporal resolution [26, 2, 45, 35]. We consider two periods: 1961–2020 with only a few missing observations at 4342 weather stations, and 1901–1960 with sparser spatial coverage through 920 stations with a few missing observations. For each combination of an observation series and period with available data, the different samples consist of the observations for the different decades, either $(1901 + (j - 1) \times 10) - (1901 + (j \times 10))$ or $(1961 + (j - 1) \times 10) - (1961 + (j \times 10))$, with $j \in \{1, 2, \dots, 6\}$ and $J = 6$. Extremes of daily precipitation data show only very weak temporal correlation. Estimates of the extremal index obtained using the estimator of [42] are larger than 0.9 for more than 95% of the stations, implying that the average cluster length of exceedances is less than 1/0.9 and very close to one for those stations. Therefore, we consider that serial independence is a reasonable working assumption in this application.

We test (H_0) using the asymptotic χ^2-distribution under (H_0) with $\gamma = 0.05$ and $L = 10$, and $\beta_{j,n} = \beta_n$ is chosen to retain $k_{j,n} = k = 100$ extreme order statistics for each sample $j = 1, \dots, 6$. Figure 6 reports results where we highlight gauges with significant nonstationarity across decades, and we also report relatively low p-values larger than $\gamma = 0.05$. For the 1961–2020 period, the proportion of gauges with (H_0) rejected is around 5% and therefore of the order of γ, which corresponds to the expected number of type I errors. For 1901–1960, a higher proportion of 16% of stations has (H_0) rejected. Unreported results for other choices of k and L show relatively stable behavior of p-values over a range of k between 50 and 250 for fixed L, with slightly higher values for larger values of k for which estimation uncertainty is lower, thus indicating no strong sensitivity to the choice of k. p-values tend to be relatively lower when increasing L for fixed k, which is natural since the test then considers a larger number L of extreme quantiles, some of them estimated at lower probability levels, such that statistical uncertainty decreases and the power of the test increases; however, we would then test for differences in the distribution of events that may be less extreme than those considered with $L = 10$. Follow-up work will include expert interpretations by climatologists in terms of observation biases and existing knowledge about local climate systems and potential climate-change effects.

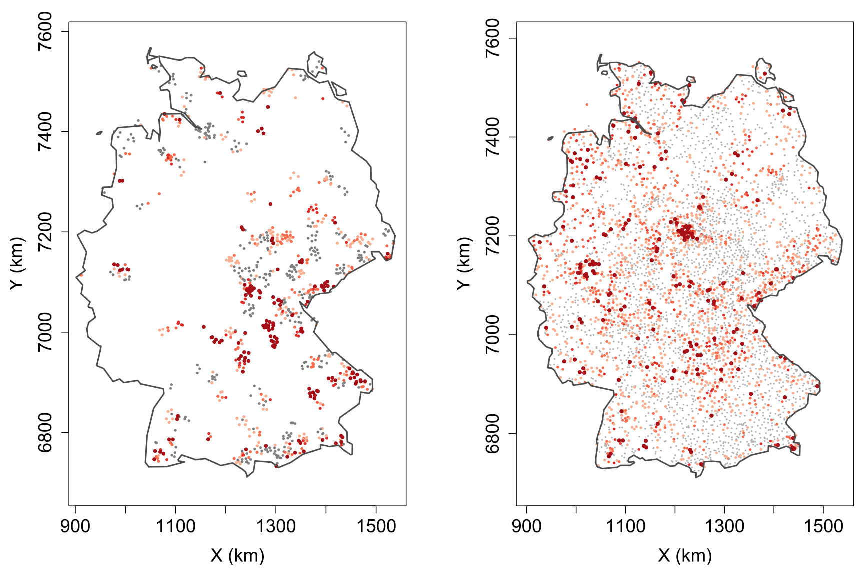

FIG 6. *p-values for German precipitation series. Left:* 1901–1960*; right:* 1961–2020*. p-value intervals are* $[0.0, 0.05]$ *(dark red, big points),* $(0.05, 0.5]$ *(lighter red, smaller points), and* $(0.5, 1]$ *(grey, small points).*

7. Conclusion

The ANOVEX test puts focus on differences in the distribution of the most extreme possible events, for which uncertainties are usually high, especially with naive approaches not making use of extreme-value statistics for extrapolation. The simulation study highlighted that our test is an easy-to-use, interpretable and efficient tool for detecting different tail behaviors. It is a valuable complement to existing tests that focus on more specific alternative hypotheses, such as differences in the tail index or, provided that tail indices coincide, in the tail scale. Our test is particularly powerful for detecting different tail scales in the case of small tail indices. The number and levels of extreme quantiles $\alpha_{\ell,n}$, $\ell = 1, \dots, L$, are important hyperparameters of the ANOVEX approach. In simulations and real-data analyses, we obtain reliable results when setting $\alpha_{\ell,n} = 1-\ell/n$, where the statistical power of the tests tends to increase strongest when increasing L from the lowest value 2 towards larger values around 10; for example, this can be seen from the relatively strong negative slopes in type-II errors for small values of L. The behavior of the test can be examined for a range of L-values in practice to better understand its sensitivity to this hyperparameter and take better-informed decisions.

In future work, we aim to investigate how an automatic choice of L could be achieved, for example by establishing an explicit link between L and the second-order parameter ρ. We also plan to relax certain assumptions, notably by considering dependent data (dependence across groups, and/or serial dependence within groups). In the current study, we have mainly considered samples with equal sizes in the error approximations and the simulation study. Although our theoretical result (Theorem 3.1) allows samples to have different sizes, biases may arise in practice, particularly when dealing with relatively small samples and strongly unbalanced sample sizes, resulting in type I error probabilities deviating significantly from the test level.

Furthermore, we plan to leverage the ANOVEX test statistic for change point detection. More generally, our test statistic provides a valuable tail dissimilarity measure for machine-learning tools, such as for making splits in regression trees and random forests. In recent related work of [29], estimation of a tree was proposed but with a focus solely on the tail index. The approach of [16] adopts a model-based approach for inferring trees by using the Generalized Pareto Distribution (GPD) for exceedances above a high threshold. In [18], GPD-based random forests were developed; however, the tree structure and likelihood weights are computed in a preliminary step using traditional quantile regression forests. To extend these existing approaches, an ANOVEX-based procedure could offer a rapid and robust likelihood-free method for tail prediction, applicable to both trees and random forests.

Appendix A: Proofs of the theoretical results

Some technical lemmas are collected in Section A.1 and will be useful to prove the main results in Section A.2.

A.1. Preliminary results

We first show the following: if all cumulative distribution functions $F_1, \dots, F_J$ are heavy-tailed under (H_0), then necessarily their tail-indices are the same.

Lemma A.1. *Assume (H_0) holds. If $F_j \in \mathcal{C}_1(\xi_j)$ for $j = 1, \dots, J$ then*

(i) $\xi_1 = \dots = \xi_J =: \xi$.
(ii) $\log q_j(\alpha) = -\xi \log(1-\alpha)(1+o(1))$ *for all* $j = 1, \dots, J$ *as* $\alpha \to 1$.

Proof of Lemma A.1. (i) Let us recall that $F_j \in \mathcal{C}_1(\xi_j)$ implies that there exists a slowly varying function L_j such that

$$q_j(\alpha) = (1-\alpha)^{-\xi_j} L_j(1/(1-\alpha)). \tag{15}$$

It straightforwardly follows that $q_{j_0}(\alpha)/q_j(\alpha) \to 0$ as $\alpha \to 1$ if $\xi_j > \xi_{j_0}$ and $q_{j_0}(\alpha)/q_j(\alpha) \to +\infty$ as $\alpha \to 1$ if $\xi_j < \xi_{j_0}$. The result (i) is thus proved.
(ii) From Lemma A.1(i), one has $\xi_1 = \dots = \xi_J =: \xi$ and therefore (15) can be rewritten as

$$\log(q_j(\alpha)) = -\xi \log(1-\alpha) + \log(L_j(1/(1-\alpha))) = -\xi \log(1-\alpha)(1+o(1)),$$

which proves the result. □

The second lemma provides an asymptotic equivalent of $\Delta_{1,n}$ (defined in Lemma 2.1) as $n \to \infty$.

Lemma A.2. *Let $E_1, \dots, E_J$ be independent samples. Assume $F_j \in \mathcal{C}_1(\xi_j)$ for $j = 1, \dots, J$.*

- *For all $n \geq 1$ and $\ell = 1, \dots, L$, let $\alpha_{\ell,n} \in (0,1)$ such that $\alpha_{\ell,n} \to 1$ as $n \to \infty$.*
- *Let $\widetilde{q}_j(\alpha_{\ell,n})$ be an estimator of the extreme quantile $q_j(\alpha_{\ell,n})$ computed on the n_j-sample E_j such that $n_j \to \infty$ as $n \to \infty$ and*

$$\sigma_{j,\ell,n}^{-1} \left(\log \widetilde{q}_j(\alpha_{\ell,n}) - \log q_j(\alpha_{\ell,n}) \right) \xrightarrow{d} \xi_j Z_j,$$

where $Z_1, \dots, Z_J$ are independent standard Gaussian random variables and for some $\sigma_{j,\ell,n} \to \infty$ as $n \to \infty$, for all $j = 1, \dots, J$ and $\ell = 1, \dots, L$.

Then, under (H_0),

$$\Delta_{1,n} = \frac{\xi^2}{L} \sum_{\ell=1}^{L} \left[\frac{1}{J} \sum_{j=1}^{J} \sigma_{j,\ell,n}^2 Z_j^2 - \left(\frac{1}{J} \sum_{k=1}^{J} \sigma_{k,\ell,n} Z_k \right)^2 \right] (1 + o_{\mathbb{P}}(1)),$$

Proof of Lemma A.2. Let us first recall that for all $\ell = 1, \dots, L$,

$$\Delta_{1,\ell,n} = \frac{1}{J} \sum_{j=1}^{J} \left(\log \widetilde{q}_j(\alpha_{\ell,n}) - \mu_{\alpha,n}^{(\ell)} \right)^2 \text{ with } \mu_{\alpha,n}^{(\ell)} = \frac{1}{J} \sum_{j=1}^{J} \log \widetilde{q}_j(\alpha_{\ell,n}).$$

Under (H_0), we have in view of Lemma A.1(i):

$$\log \widetilde{q}_j(\alpha_{\ell,n}) = \log q_j(\alpha_{\ell,n}) + \sigma_{j,\ell,n}\xi Z_j(1+o_{\mathbb{P}}(1)),$$

where $Z_1,\dots,Z_J$ follow independent standard Gaussian distributions.

Lemma A.1(ii) entails

$$\log \widetilde{q}_j(\alpha_{\ell,n}) = -\xi\log(1-\alpha_{\ell,n})(1+o(1)) + \sigma_{j,\ell,n}\xi Z_j(1+o_{\mathbb{P}}(1)),$$

and therefore,

$$\left(\log \widetilde{q}_j(\alpha_{\ell,n}) - \mu_{\alpha,n}^{(\ell)}\right)^2 = \xi^2\left(\sum_{k=1}^J \kappa_{k,j}\sigma_{k,\ell,n}Z_k\right)^2(1+o_{\mathbb{P}}(1)),$$

with $\kappa_{j,j}=1-1/J$ and $\kappa_{k,j}=-1/J$ if $k\neq j$, so that

$$\begin{aligned}\Delta_{1,\ell,n} &= \frac{J-1}{J^2}\xi^2\sum_{j=1}^J \sigma_{j,\ell,n}^2 Z_j^2(1+o_{\mathbb{P}}(1))\\ &\quad -\frac{2}{J^2}\xi^2\sum_{1\le j<k\le J}\sigma_{j,\ell,n}\sigma_{k,\ell,n}Z_jZ_k(1+o_{\mathbb{P}}(1))\\ &= \xi^2\left[\frac{1}{J}\sum_{j=1}^J\sigma_{j,\ell,n}^2Z_j^2 - \left(\frac{1}{J}\sum_{k=1}^J\sigma_{k,\ell,n}Z_k\right)^2\right](1+o_{\mathbb{P}}(1)).\end{aligned}$$

Finally, without any further assumption, the term $\Delta_{1,n}$ may be written as follows:

$$\Delta_{1,n} = \frac{\xi^2}{L}\sum_{\ell=1}^L\left[\frac{1}{J}\sum_{j=1}^J\sigma_{j,\ell,n}^2Z_j^2 - \left(\frac{1}{J}\sum_{k=1}^J\sigma_{k,\ell,n}Z_k\right)^2\right](1+o_{\mathbb{P}}(1)),$$

which is the desired result. □

Similarly, the third lemma provides an asymptotic equivalent of $\Delta_{2,n}$ (defined in Lemma 2.1) as $n\to\infty$.

Lemma A.3. *Let $E_1,\dots,E_J$ be J samples. Assume $F_j\in\mathcal{C}_1(\xi_j)$ for $j=1,\dots,J$.*

- *For all $n\ge 1$ and $\ell=1,\dots,L$, let $\alpha_{\ell,n}\in(0,1)$ such that $\alpha_{\ell,n}\to 1$ as $n\to\infty$.*
- *Let $\widetilde{q}_j(\alpha_{\ell,n})$ be an estimator of the extreme quantile $q_j(\alpha_{\ell,n})$ computed on the n_j-sample E_j such that $n_j\to\infty>0$ as $n\to\infty$ and*

$$\log\widetilde{q}_j(\alpha_{\ell,n})/\log q_j(\alpha_{\ell,n}) \xrightarrow{\mathbb{P}} 1, \tag{16}$$

for all $j=1,\dots,J$ and $\ell=1,\dots,L$.

Then, under (H_0),

$$\Delta_{2,n} = \xi^2 \ \mathrm{varlog}((1-\alpha_n)_{1:L})(1+o_{\mathbb{P}}(1)).$$

Proof of Lemma A.3. Combining (16) with Lemma A.1(*ii*) yields

$$\log \widetilde{q}_j(\alpha_{\ell,n}) = \left(\log q_j(\alpha_{\ell,n})\right)(1+o_{\mathbb{P}}(1)) = -\xi \log(1-\alpha_{\ell,n})(1+o_{\mathbb{P}}(1))$$

so that

$$\mu_{\alpha,n}^{(\ell)} = \frac{1}{J}\sum_{j=1}^{J} \log \widetilde{q}_j(\alpha_{\ell,n}) = -\xi \log(1-\alpha_{\ell,n})(1+o_{\mathbb{P}}(1)),$$

for all $\ell = 1,\dots,L$, and consequently,

$$\mu_{\alpha,n} - \mu_{\alpha,n}^{(\ell)} = -\xi \ \frac{1}{L}\sum_{k=1}^{L} \widetilde{\kappa}_{k,\ell} \log(1-\alpha_{k,n})(1+o_{\mathbb{P}}(1)),$$

where $\widetilde{\kappa}_{\ell,\ell} = 1-L$ and $\widetilde{\kappa}_{k,\ell} = 1$ if $k \neq \ell$. Some straightforward calculations lead to

$$\begin{aligned}
\Delta_{2,n} &= \frac{1}{L}\sum_{\ell=1}^{L}\left(\mu_{\alpha,n} - \mu_{\alpha,n}^{(\ell)}\right)^2 \\
&= \xi^2 \left(\frac{L-1}{L^2}\sum_{\ell=1}^{L} \left(\log\left(1-\alpha_{\ell,n}\right)\right)^2 - \frac{2}{L^2}\sum_{1\le \ell < \ell' \le L} \log\left(1-\alpha_{\ell,n}\right)\log\left(1-\alpha_{\ell',n}\right)\right) \\
&\quad (1+o_{\mathbb{P}}(1)) \\
&= \xi^2 \left(\frac{1}{L}\sum_{\ell=1}^{L} \left(\log\left(1-\alpha_{\ell,n}\right)\right)^2 - \frac{1}{L^2}\sum_{\ell=1}^{L}\sum_{\ell'=1}^{L} \log\left(1-\alpha_{\ell,n}\right)\log\left(1-\alpha_{\ell',n}\right)\right) \\
&\quad (1+o_{\mathbb{P}}(1)) \\
&= \xi^2 \left[\frac{1}{L}\sum_{\ell=1}^{L} \left(\log\left(1-\alpha_{\ell,n}\right)\right)^2 - \left(\frac{1}{L}\sum_{\ell=1}^{L} \left(\log\left(1-\alpha_{\ell,n}\right)\right)^2\right)\right](1+o_{\mathbb{P}}(1)),
\end{aligned}$$

which is the expected result. □

The next lemma provides precise asymptotic representations associated with Hill estimators and intermediate quantiles computed on Pareto samples.

Lemma A.4. *Consider two independent samples* $E_1 = \{X_1^{(1)},\dots,X_n^{(1)}\}$ *and* $E_2 = \{X_1^{(2)},\dots,X_n^{(2)}\}$, *both distributed from a Pareto distribution* $\mathcal{P}(1/\xi)$, $\xi > 0$. *Let* (β_n) *be an intermediate probability level,* $\widehat{q}_j(\beta_n) := X^{(j)}_{\lfloor n(1-\beta_n)\rfloor,n}$ *and let* $\hat{\xi}^H_j(\beta_n)$ *be the associated Hill estimators (3),* $j = 1,2$. *Then, the following asymptotic representations hold:*

$$\hat{\xi}^H_j(\beta_n) \stackrel{d}{=} \xi + \frac{\xi}{\sqrt{n(1-\beta_n)}}\Gamma_{\xi,j}\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)\right), \quad (17)$$

$$\log\left(\frac{\widehat{q}_j(\beta_n)}{q_j(\beta_n)}\right) \stackrel{d}{=} \frac{\xi}{\sqrt{n(1-\beta_n)}}\Gamma_{q,j} \tag{18}$$

$$\times\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)+O_{\mathbb{P}}(1-\beta_n)\right), \tag{19}$$

$$\hat{\xi}_1^{H}(\beta_n)-\hat{\xi}_2^{H}(\beta_n) \stackrel{d}{=} \frac{\sqrt{2}\xi}{\sqrt{n(1-\beta_n)}}\Gamma_\xi\left(1+O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right)\right), \tag{20}$$

$$\log\left(\frac{\widehat{q}_1(\beta_n)q_2(\beta_n)}{q_1(\beta_n)\widehat{q}_2(\beta_n)}\right) \stackrel{d}{=} \frac{\sqrt{2}\xi}{\sqrt{n(1-\beta_n)}}\Gamma_q\left(1+O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right)+O_{\mathbb{P}}(1-\beta_n)\right), \tag{21}$$

where $\Gamma_{q,j}$, $\Gamma_{\xi,j}$, Γ_q and Γ_ξ are standard Gaussian random variables, $j=1,2$. Moreover, Equations (20) and (21) also hold for two Pareto samples with different scale parameters: $X_i^{(2)} \stackrel{d}{=} \lambda X_i^{(1)}$, $\lambda>0$, $i=1,\dots,n$.

Proof of Lemma A.4. Let us introduce $k_n=\lfloor n(1-\beta_n)\rfloor$ to simplify the notation. Let $j\in\{1,2\}$. First, Rényi's representation entails that the log-spacings $\left(\log(X_{n-i,n}^{(j)})-\log(X_{n-k_n,n}^{(j)})\right)$, $i=0,\dots,k_n-1$ are independent and exponentially distributed. Hill estimators are thus Gamma distributed:

$$\hat{\xi}_j^{\mathrm{H}}(\beta_n) \stackrel{d}{=} \frac{1}{k_n}\sum_{i=1}^{k_n}\mathcal{E}_i^{(j)},$$

where $\{\mathcal{E}_1^{(j)},\dots,\mathcal{E}_{k_n}^{(j)}\}$ are i.i.d. realizations of an exponential distribution with mean ξ. Berry-Esseen Theorem thus yields

$$\hat{\xi}_j^{\mathrm{H}}(\beta_n)-\xi \stackrel{d}{=} \frac{\xi}{\sqrt{k_n}}\Gamma_{\xi,j}\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right),$$

and (17) is proved. Second, one has, in view of Rényi's representation:

$$\log\left(\frac{\widehat{q}_j(\beta_n)}{q_j(\beta_n)}\right) \stackrel{d}{=} \mathcal{E}_{n-k_n,n}^{(j)}-\xi\log(n/k_n) \stackrel{d}{=} \sum_{i=k_n+1}^{n}\frac{\mathcal{E}_i^{(j)}}{i}-\xi\log(n/k_n),$$

and thus, introducing $Y_i^{(j)}=(\mathcal{E}_i^{(j)}-\xi)/(\xi i)$, the following expansion holds:

$$\sqrt{k_n}\log\left(\frac{\widehat{q}_j(\beta_n)}{q_j(\beta_n)}\right) \stackrel{d}{=} \xi\sqrt{k_n}\sum_{i=k_n+1}^{n}Y_i^{(j)} + \xi\sqrt{k_n}\left(\sum_{i=k_n+1}^{n}\frac{1}{i}-\log(n/k_n)\right) =: \xi(A_n^{(j)}+B_n).$$

The well-known formula

$$\sum_{i=1}^{n}\frac{1}{i}=\log(n)+\gamma-\frac{1}{2n}(1+o(1)),$$

(where γ is Euler's constant) entails that the non-random term can be controlled as $B_n = O(1/\sqrt{k_n})$. Letting $\sigma_i^2 = \mathbb{E}((Y_i^{(j)})^2)$ and $\rho_i = \mathbb{E}(|Y_i^{(j)}|^3)$, Berry-Esseen Theorem for non identically distributed random variables shows that

$$\frac{\sum_{i=k_n+1}^n Y_i^{(j)}}{\sqrt{\sum_{i=k_n+1}^n \sigma_i^2}} = \Gamma_{q,j} + O_{\mathbb{P}}\left(\frac{\max_{i=k_n+1,\dots,n} \rho_i/\sigma_i^2}{\sqrt{\sum_{i=k_n+1}^n \sigma_i^2}}\right),$$

or equivalently,

$$A_n^{(j)} = \sqrt{k_n \sum_{i=k_n+1}^n \sigma_i^2}\, \Gamma_{q,j} + O_{\mathbb{P}}\left(\sqrt{k_n} \max_{i=k_n+1,\dots,n} \rho_i/\sigma_i^2\right).$$

Moreover, $\sigma_i^2 = 1/i^2$, $\rho_i = c/i^3$ with $c > 0$ so that

$$\begin{aligned}\sum_{i=k_n+1}^n \sigma_i^2 &= \frac{1}{k_n}\left(1 + O\left(\frac{1}{k_n}\right)\right) - \frac{1}{n}\left(1 + O\left(\frac{1}{n}\right)\right) \\ &= \frac{1}{k_n}\left(1 + O\left(\frac{1}{k_n}\right) + O\left(\frac{k_n}{n}\right)\right), \\ \max_{i=k_n+1,\dots,n} \rho_i/\sigma_i^2 &= O\left(\frac{1}{k_n}\right),\end{aligned}$$

and therefore $A_n^{(j)} = \Gamma_{q,j} + O_{\mathbb{P}}(1/\sqrt{k_n}) + O_{\mathbb{P}}(k_n/n)$. All in all,

$$\sqrt{k_n}\log\left(\frac{\widehat{q}_j(\beta_n)}{q_j(\beta_n)}\right) \stackrel{d}{=} \xi\Gamma_{q,j} + O(1/\sqrt{k_n}) + O_{\mathbb{P}}(k_n/n),$$

and (18) is proved. Moreover, one has

$$\hat{\xi}_1^{\mathrm{H}}(\beta_n) - \hat{\xi}_2^{\mathrm{H}}(\beta_n) \stackrel{d}{=} \frac{(\mathcal{E}_1^{(1)} - \mathcal{E}_1^{(2)}) + \ldots + (\mathcal{E}_{k_n}^{(1)} - \mathcal{E}_{k_n}^{(2)})}{k_n} \stackrel{d}{=} \frac{\mathcal{L}_1 + \ldots + \mathcal{L}_{k_n}}{k_n},$$

where $\{\mathcal{L}_1, \ldots, \mathcal{L}_{k_n}\}$ are i.i.d. realizations of a centered Laplace distribution with variance $2\xi^2$. Since the Laplace distribution is log-concave, centered and symmetric, [27, Theorem 1] may be applied to refine the Berry-Esseen bound with k_n (or equivalently $n(1-\beta_n)$) instead of $\sqrt{k_n}$ (or equivalently $\sqrt{n(1-\beta_n)}$), hence the third result (20). Similarly,

$$\begin{aligned}\log\left(\frac{\widehat{q}_1(\beta_n)}{\widehat{q}_2(\beta_n)}\right) &= \log\left(\frac{\widehat{q}_1(\beta_n)}{q_1(\beta_n)}\right) - \log\left(\frac{\widehat{q}_2(\beta_n)}{q_2(\beta_n)}\right) \\ &\stackrel{d}{=} \sum_{i=k_n+1}^n \frac{(\mathcal{E}_i^{(1)} - \mathcal{E}_i^{(2)})}{i} \stackrel{d}{=} \sum_{i=k_n+1}^n \frac{\mathcal{L}_i}{i},\end{aligned}$$

and the second result of [27, Theorem 1] can be used to establish the Berry-Esseen bound. Rewriting

$$\sqrt{k_n}\log\left(\frac{\widehat{q}_1(\beta_n)}{\widehat{q}_2(\beta_n)}\right) \stackrel{d}{=} \sqrt{k_n}\sqrt{\sum_{i=k_n+1}^{n}\frac{1}{i^2}}\sum_{i=k_n+1}^{n}\theta_{i,n}\mathcal{L}_i,$$

with $\theta_{i,n} = \frac{1}{i\sqrt{\sum_{j=k_n+1}^{n} 1/j^2}}$ and $\sum_{i=k_n+1}^{n}\theta_{i,n}^2 = 1$, it thus follows:

$$\sum_{i=k_n+1}^{n}\theta_{i,n}\mathcal{L}_i = \sqrt{2}\xi\Gamma_q\left(1+O_{\mathbb{P}}\left(\sum_{i=k_n+1}^{n}\theta_{i,n}^4\right)\right).$$

Straightforward calculations on Riemann series lead to

$$\begin{aligned}
\sum_{i=k_n+1}^{n}\theta_{i,n}^4 &= \frac{\sum\limits_{i=k_n+1}^{n}\frac{1}{i^4}}{\left(\sum\limits_{i=k_n+1}^{n}\frac{1}{i^2}\right)^2} = \frac{\sum\limits_{i=k_n+1}^{\infty}\frac{1}{i^4} - \sum\limits_{i=n+1}^{\infty}\frac{1}{i^4}}{\left(\sum\limits_{i=k_n+1}^{\infty}\frac{1}{i^2} - \sum\limits_{i=n+1}^{\infty}\frac{1}{i^2}\right)^2} \\
&= \frac{O\left(\frac{1}{k_n^3}\right)}{\left(\frac{1}{k_n}+O\left(\frac{1}{k_n^2}\right)+O\left(\frac{1}{n}\right)\right)^2} = O\left(\frac{1}{k_n}\right),
\end{aligned}$$

$$\sqrt{k_n\sum_{i=k_n+1}^{n}\frac{1}{i^2}} = \sqrt{k_n}\sqrt{\frac{1}{k_n}+O\left(\frac{1}{k_n^2}\right)+O\left(\frac{1}{n}\right)} = 1+O\left(\frac{1}{k_n}\right)+O\left(\frac{k_n}{n}\right).$$

Combining the previous expansions yields the expected result (21). To conclude, since the Hill estimator is scale invariant, Equation (20) also holds if $X_i^{(2)} \sim \lambda\mathcal{P}(1/\xi)$, for all $\lambda > 0$ and $i = 1,\dots,n$. By noticing that, in this case, $\widehat{q}_2(\beta_n)/(\lambda q_1(\beta_n)) \stackrel{d}{=} \widehat{q}_1(\beta_n)/q_1(\beta_n)$, Equation (21) holds true as well. □

A.2. Proofs of main results

Proof of Theorem 3.1. First, (5) and Lemma A.2 entail

$$\begin{aligned}
\Delta_{1,n} &= \frac{\xi^2}{L}\sum_{\ell=1}^{L}\bar{\sigma}_{\ell,n}^2\left[\frac{1}{J}\sum_{j=1}^{J}Z_j^2 - \left(\frac{1}{J}\sum_{k=1}^{J}Z_k\right)^2\right](1+o_{\mathbb{P}}(1)) \\
&\stackrel{d}{=} \frac{\xi^2}{J}\left(\frac{1}{L}\sum_{\ell=1}^{L}\bar{\sigma}_{\ell,n}^2\right)\chi_{J-1}^2(1+o_{\mathbb{P}}(1)) \\
&= \frac{\xi^2}{J}\bar{\sigma}_n^2\chi_{J-1}^2(1+o_{\mathbb{P}}(1)),
\end{aligned}$$

where $Z_1, \dots, Z_J$ follow independent standard Gaussian distributions. Second, remark that (5) and $\bar\sigma_{\ell,n} \to 0$ imply (16) so that Lemma A.3 yields $\Delta_{2,n} = \xi^2 \operatorname{varlog}((1-\alpha_n)_{1:L})(1+o_{\mathbb{P}}(1))$, and consequently,

$$\frac{J \operatorname{varlog}((1-\alpha_n)_{1:L})}{\bar\sigma_n^2} \frac{\Delta_{1,n}}{\Delta_{2,n}} \overset{d}{=} \chi^2_{J-1}(1+o_{\mathbb{P}}(1)),$$

which proves the result. □

Proof of Corollary 3.1. Under (H_0), [10, Theorem 4.3.8] entails the following representation for all $j = 1, \dots, J$ and $\ell = 1, \dots, L$:

$$\log\left(\hat q_j^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_{j,n})\right) = \log\left(q_j(\alpha_{\ell,n})\right) + \frac{\log\left(\frac{1-\beta_{j,n}}{1-\alpha_{\ell,n}}\right)}{\sqrt{n_j(1-\beta_{j,n})}} \xi_j Z_j (1+o_{\mathbb{P}}(1)), \tag{22}$$

where $Z_1, \dots, Z_J$ are independent standard Gaussian random variables. Moreover, taking account of $\xi_j = \xi$ from Lemma A.1(i), $n_j/n \to \lambda_j$ and $(1-\beta_{j,n})/(1-\beta_n) \to 1/\lambda_j$ as $n \to \infty$ for all $j = 1, \dots, J$ yields

$$\log\left(\hat q_j^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_{j,n})\right) = \log\left(q_j(\alpha_{\ell,n})\right) + \frac{\log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)}{\sqrt{n(1-\beta_n)}} \xi Z_j (1+o_{\mathbb{P}}(1)),$$

since $\log\left(\frac{1-\beta_{j,n}}{1-\alpha_{\ell,n}}\right) \sim \log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)$ as $n \to \infty$. As a consequence,

$$\frac{\sqrt{n(1-\beta_n)}}{\log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)} \left(\log\left(\hat q_j^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_{j,n})\right) - \log\left(q_j(\alpha_{\ell,n})\right)\right) = \xi Z_j (1+o_{\mathbb{P}}(1)),$$

and (5) holds with $\bar\sigma_{\ell,n} = \frac{\log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)}{\sqrt{n(1-\beta_n)}}$. The result follows from Theorem 3.1. □

Proof of Proposition 4.1. In the case where $J = 2$, $\Delta_{1,n}$ can be simplified as

$$\Delta_{1,n} = \frac{1}{4L} \sum_{\ell=1}^{L} \left(\log \widetilde q_1(\alpha_{\ell,n}) - \log \widetilde q_2(\alpha_{\ell,n})\right)^2.$$

Using the Weissman estimator (2) and the Hill estimator (3), we have, for $j = 1, 2$:

$$\log \hat q_j^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) = \log q_j(\alpha_{\ell,n}) + \log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)\left(\hat\xi_j^{\mathrm{H}}(\beta_n) - \xi\right) + \log\left(\frac{\widehat q_j(\beta_n)}{q_j(\beta_n)}\right), \tag{23}$$

and therefore

$$\log \hat q_1^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) - \log \hat q_2^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) = \log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)\left(\hat\xi_1^{\mathrm{H}}(\beta_n) - \hat\xi_2^{\mathrm{H}}(\beta_n)\right) + \log\left(\frac{\widehat q_1(\beta_n)}{q_1(\beta_n)}\right) - \log\left(\frac{\widehat q_2(\beta_n)}{q_2(\beta_n)}\right). \tag{24}$$

Since $X_i^{(1)}$ and $X_i^{(2)}$ are Pareto distributed for $i = 1, \dots, n$, Equations (20) and (21) of Lemma A.4 yield

$$\hat{\xi}_1^{\text{H}}(\beta_n) - \hat{\xi}_2^{\text{H}}(\beta_n) \overset{d}{=} \frac{\sqrt{2}\xi}{\sqrt{n(1-\beta_n)}}\Gamma_\xi \left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right)\right) \quad (25)$$

$$\log\left(\frac{\widehat{q}_1(\beta_n)}{q_1(\beta_n)}\right) - \log\left(\frac{\widehat{q}_2(\beta_n)}{q_2(\beta_n)}\right) \overset{d}{=} \quad (26)$$

$$\frac{\sqrt{2}\xi}{\sqrt{n(1-\beta_n)}}\Gamma_q \times \left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right) + O_{\mathbb{P}}(1-\beta_n)\right) \quad (27)$$

where Γ_q and Γ_ξ are two independent standard Gaussian random variables. Hence, plugging (25) and (26) into (24), it follows that

$$\begin{aligned}
&\log \hat{q}_1^{\text{W}}(\alpha_{\ell,n} \mid \beta_n) - \log \hat{q}_2^{\text{W}}(\alpha_{\ell,n} \mid \beta_n) \\
&\quad \overset{d}{=} \frac{\log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)}{\sqrt{n(1-\beta_n)}}\sqrt{2}\xi\Gamma_\xi \left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right)\right) \\
&\qquad + \frac{\sqrt{2}\xi}{\sqrt{n(1-\beta_n)}}\Gamma_q \left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right) + O_{\mathbb{P}}(1-\beta_n)\right).
\end{aligned}$$

Taking account of $\alpha_{\ell,n} = 1 - \tau_\ell/n$, the above equality can be rewritten as

$$\begin{aligned}
&\log \hat{q}_1^{\text{W}}(\alpha_{\ell,n} \mid \beta_n) - \log \hat{q}_2^{\text{W}}(\alpha_{\ell,n} \mid \beta_n) \\
&\overset{d}{=} \frac{1}{\sqrt{n(1-\beta_n)}} \left\{ \log\left(\frac{n(1-\beta_n)}{\tau_\ell}\right)\sqrt{2}\xi\,\Gamma_\xi \left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right)\right) + \sqrt{2}\xi\Gamma_q \right. \\
&\qquad\qquad \left. + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right) + O_{\mathbb{P}}(1-\beta_n)\right\} \\
&\overset{d}{=} \frac{1}{\sqrt{n(1-\beta_n)}} \left\{ \sqrt{1 + \left(\log\left(\frac{n(1-\beta_n)}{\tau_\ell}\right)\right)^2}\sqrt{2}\xi\Gamma \right. \\
&\qquad\qquad \left. + O_{\mathbb{P}}\left(\frac{\log(n(1-\beta_n))}{n(1-\beta_n)}\right) + O_{\mathbb{P}}(1-\beta_n)\right\},
\end{aligned}$$

where Γ is a standard Gaussian random variable, and consequently,

$$\begin{aligned}
\Delta_{1,n} &\overset{d}{=} \frac{1}{4L}\sum_{\ell=1}^{L}\left[\frac{\sqrt{2}\xi}{\sqrt{n(1-\beta_n)}}\sqrt{1 + \left(\log\left(\frac{n(1-\beta_n)}{\tau_\ell}\right)\right)^2}\Gamma\right]^2 \\
&\qquad \left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(n(1-\beta_n))}\right)\right) \\
&\overset{d}{=} \frac{\xi^2\Gamma^2}{2}\frac{1 + S_n(\beta_n, \tau_{1:L})}{n(1-\beta_n)}\left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(n(1-\beta_n))}\right)\right).
\end{aligned}$$

Similarly, one can easily prove, thanks to the choice $\alpha_{\ell,n} = 1 - \tau_\ell/n$, that

$$\begin{aligned}\Delta_{2,n} &= \frac{1}{L^2}\left[L \sum_{\ell=1}^{L} \left(\frac{\log \hat{q}_1^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) + \log \hat{q}_2^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n)}{2} \right) \right.\\ &\qquad \left. - \left(\sum_{\ell=1}^{L} \frac{\log \hat{q}_1^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) + \log \hat{q}_2^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n)}{2} \right)^2 \right]\\ &= \operatorname{varlog}(\tau_{1:L}) \left(\frac{\hat{\xi}_1^{\mathrm{H}}(\beta_n) + \hat{\xi}_2^{\mathrm{H}}(\beta_n)}{2} \right)^2\\ &= \xi^2 \operatorname{varlog}(\tau_{1:L}) \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right) \right). \end{aligned} \tag{28}$$

Combining the two previous results yields the first-order approximation:

$$\begin{aligned}\frac{\Delta_{1,n}}{\Delta_{2,n}} &\stackrel{d}{=} \frac{\Gamma^2}{2\operatorname{varlog}(\tau_{1:L})} \frac{1 + S_n(\beta_n, \tau_{1:L})}{n(1-\beta_n)}\\ &\qquad \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(n(1-\beta_n))} \right) \right),\end{aligned}$$

and, in view of (9) the test statistic T_n becomes, under the condition $(1 - \beta_n)\log(n) \to 0$ as $n \to \infty$:

$$T_n = \Gamma^2 \left(1 + \frac{1}{S_n(\beta_n, \tau_{1:L})} \right) \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(n(1-\beta_n))} \right) \right)$$

and the result is proved. □

Proof of Proposition 4.2. As a consequence of (23), one has

$$\begin{aligned}\log \hat{q}_1^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) - \log \hat{q}_2^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) &= \log\left(\frac{q_1(\alpha_{\ell,n})}{q_2(\alpha_{\ell,n})} \right)\\ &\quad + \log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}} \right) \left(\hat{\xi}_1^{\mathrm{H}}(\beta_n) - \hat{\xi}_2^{\mathrm{H}}(\beta_n) \right)\\ &\quad + \log\left(\frac{\widehat{q}_1(\beta_n)}{q_1(\beta_n)} \right) - \log\left(\frac{\widehat{q}_2(\beta_n)}{q_2(\beta_n)} \right).\end{aligned}$$

Besides, $q_2(\alpha_{\ell,n}) = \lambda_n q_1(\alpha_{\ell,n})$ and Equations (20), (21) in Lemma A.4 entail:

$$\hat{\xi}_1^{\mathrm{H}}(\beta_n) - \hat{\xi}_2^{\mathrm{H}}(\beta_n) \stackrel{d}{=} \frac{\sqrt{2}\xi}{\sqrt{n(1-\beta_n)}} \Gamma_\xi \left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)} \right) \right),$$

$$\begin{aligned}\log\left(\frac{\widehat{q}_1(\beta_n)}{q_1(\beta_n)} \right) - \log\left(\frac{\widehat{q}_2(\beta_n)}{q_2(\beta_n)} \right) &\stackrel{d}{=} \frac{\sqrt{2}\xi}{\sqrt{n(1-\beta_n)}} \Gamma_q \left(1 + O_{\mathbb{P}}\left(\frac{1}{n(1-\beta_n)} \right) \right.\\ &\qquad \left. + O_{\mathbb{P}}(1-\beta_n) \right),\end{aligned}$$

where Γ_ξ and Γ_q are standard Gaussian random variables. Taking account of $\alpha_{\ell,n} = 1 - \tau_\ell/n$ and introducing $k_n = n(1-\beta_n)$ to simplify the notation yields

$$\begin{aligned}
&\log \hat{q}_1^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) - \log \hat{q}_2^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) + \log(\lambda_n) \\
=&\frac{\sqrt{2}\xi \log(k_n/\tau_\ell)}{\sqrt{k_n}}\Gamma_\xi\left(1 + O_{\mathbb{P}}\left(\frac{1}{k_n}\right)\right) + \frac{\sqrt{2}\xi}{\sqrt{k_n}}\Gamma_q\left(1 + O_{\mathbb{P}}\left(\frac{1}{k_n}\right) + O_{\mathbb{P}}(1-\beta_n)\right) \\
=&\frac{\sqrt{2}\xi\sqrt{1+\log(k_n/\tau_\ell)^2}}{k_n}\Gamma\left(1 + O_{\mathbb{P}}\left(\frac{1}{k_n}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right),
\end{aligned}$$

where Γ is a standard Gaussian random variable. Then, recalling that

$$S_n(\beta_n, \tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L}\left(\log\left(\frac{k_n}{\tau_\ell}\right)\right)^2,\ s_n(\beta_n, \tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L}\sqrt{1+\left(\log\left(\frac{k_n}{\tau_\ell}\right)\right)^2},$$

it follows:

$$\begin{aligned}
\Delta_{1,n} &= \frac{1}{4L}\sum_{\ell=1}^{L}\left(\log \hat{q}_1^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n) - \log \hat{q}_2^{\mathrm{W}}(\alpha_{\ell,n} \,|\, \beta_n)\right)^2 \\
&= \frac{(\log(\lambda_n))^2}{4} - \frac{\xi}{\sqrt{2}}\frac{\log(\lambda_n)\, s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}}\Gamma\left(1 + O_{\mathbb{P}}\left(\frac{1}{k_n}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right) \\
&\quad + \frac{\xi^2}{2}\frac{(1+S_n(\beta_n, \tau_{1:L}))}{k_n}\Gamma^2\left(1 + O_{\mathbb{P}}\left(\frac{1}{k_n}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right).
\end{aligned}$$

Assumption (i) entails $(\log(k_n))^2/k_n^2 \vee \log(k_n)/n = o((\log(\lambda_n))^2)$ while condition $\log(n)(1-\beta_n) \to 0$ implies $\log(k_n)/n = o(1/k_n)$. Besides, remarking that $S_n(\beta_n, \tau_{1:L}) \sim (\log(k_n))^2$ and $s_n(\beta_n, \tau_{1:L}) \sim \log(k_n)$ as $n \to \infty$ yields

$$\begin{aligned}
\Delta_{1,n} = &\frac{(\log(\lambda_n))^2}{4} - \frac{\xi}{\sqrt{2}}\frac{\log(\lambda_n)\, s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}}\Gamma + \frac{\xi^2}{2}\frac{(1+S_n(\beta_n, \tau_{1:L}))}{k_n}\Gamma^2 \\
&+ O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^2}\right) + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right) + O_{\mathbb{P}}\left(\frac{\log(\lambda_n)\log(k_n)}{k_n^{3/2}}\right) \\
&+ O_{\mathbb{P}}\left(\log(\lambda_n)\sqrt{k_n}\right).
\end{aligned}$$

Assumption (i) implies in particular $\log(\lambda_n) = O(\log(k_n)/\sqrt{k_n})$ which, in turn, entails that the third and fourth $O_{\mathbb{P}}(\cdot)$ are respectively bounded above by the first and second ones. The above expansion can thus be simplified as

$$\begin{aligned}
\Delta_{1,n} = &\frac{(\log(\lambda_n))^2}{4} - \frac{\xi}{\sqrt{2}}\frac{\log(\lambda_n)\, s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}}\Gamma + \frac{\xi^2}{2}\frac{(1+S_n(\beta_n, \tau_{1:L}))}{k_n}\Gamma^2 \\
&+ O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^2}\right) + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right).
\end{aligned}$$

Similarly to (28) in the proof of Proposition 4.1, one has

$$\Delta_{2,n} = \xi^2 \operatorname{varlog}(\tau_{1:L}) \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right).$$

Combining the previous two results, and since $(\log k_n)^2/k_n^{3/2} = o((\log(\lambda_n))^2)$, in view of (i), we get the following asymptotic expansion of $\Delta_{1,n}/\Delta_{2,n}$:

$$\begin{aligned}
\frac{\Delta_{1,n}}{\Delta_{2,n}} &= \frac{(\log(\lambda_n))^2}{4\xi^2 \operatorname{varlog}(\tau_{1:L})}\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right) \\
&\quad - \frac{1}{\sqrt{2}\xi}\frac{\log(\lambda_n)\, s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}\operatorname{varlog}(\tau_{1:L})}\Gamma\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right) \\
&\quad + \frac{1}{2}\frac{(1 + S_n(\beta_n, \tau_{1:L}))}{k_n \operatorname{varlog}(\tau_{1:L})}\Gamma^2\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right) \\
&\quad + O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^2}\right) + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right) \\
&= \frac{(\log(\lambda_n))^2}{4\xi^2 \operatorname{varlog}(\tau_{1:L})} - \frac{1}{\sqrt{2}\xi}\frac{\log(\lambda_n)\, s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}\operatorname{varlog}(\tau_{1:L})}\Gamma + \frac{1}{2}\frac{(1 + S_n(\beta_n, \tau_{1:L}))}{k_n \operatorname{varlog}(\tau_{1:L})}\Gamma^2 \\
&\quad + O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^{3/2}}\right) + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right) + O_{\mathbb{P}}\left(\frac{(\log(\lambda_n))^2}{\sqrt{k_n}}\right) \\
&\quad + O_{\mathbb{P}}\left(\frac{\log(\lambda_n)\log(k_n)}{k_n}\right).
\end{aligned}$$

The third and fourth $O_{\mathbb{P}}(\cdot)$ are bounded above by the first one since $\log(\lambda_n) = O(\log(k_n)/\sqrt{k_n})$ in view of condition (ii), and consequently,

$$\begin{aligned}
&\xi^2 \operatorname{varlog}(\tau_{1:L})\frac{\Delta_{1,n}}{\Delta_{2,n}} \\
&= \frac{(\log(\lambda_n))^2}{4} - \frac{\xi}{\sqrt{2}}\frac{\log(\lambda_n)\, s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}}\Gamma + \frac{\xi^2}{2}\frac{(1 + S_n(\beta_n, \tau_{1:L}))}{k_n}\Gamma^2 \\
&+ O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^{3/2}}\right) + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right) \\
&= \left(\frac{(\log(\lambda_n))^2}{4} - \frac{\xi}{\sqrt{2}}\frac{\log(\lambda_n) s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}}\Gamma + \frac{\xi^2}{2}\frac{(1 + S_n(\beta_n, \tau_{1:L}))}{k_n}\Gamma^2\right) \\
&\times \left(1 + \frac{O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^{3/2}}\right) + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right)}{\frac{(\log(\lambda_n))^2}{4} - \frac{\xi}{\sqrt{2}}\frac{\log(\lambda_n) s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}}\Gamma + \frac{\xi^2}{2}\frac{(1+S_n(\beta_n, \tau_{1:L}))}{k_n}\Gamma^2}\right).
\end{aligned}$$

Focusing on the denominator of the above term, condition (ii) shows that $\sqrt{k_n}\log(\lambda_n) \to 0$ as $n \to \infty$ leading to

$$\frac{(\log(\lambda_n))^2}{4} - \frac{\xi}{\sqrt{2}}\frac{\log(\lambda_n) s_n(\beta_n, \tau_{1:L})}{\sqrt{k_n}}\Gamma + \frac{\xi^2}{2}\frac{(1 + S_n(\beta_n, \tau_{1:L}))}{k_n}\Gamma^2$$

$$= \frac{\xi^2}{2}\frac{(\log(k_n))^2}{k_n}\Gamma^2(1+o_{\mathbb{P}}(1)),$$

and thus

$$\frac{\Delta_{1,n}}{\Delta_{2,n}} = \frac{1}{\xi^2\,\mathrm{varlog}(\tau_{1:L})}\left(\frac{(\log(\lambda_n))^2}{4} - \frac{\xi}{\sqrt{2}}\frac{\log(\lambda_n)s_n(\beta_n,\tau_{1:L})}{\sqrt{k_n}}\Gamma \right.$$
$$\left. + \frac{\xi^2}{2}\frac{(1+S_n(\beta_n,\tau_{1:L}))}{k_n}\Gamma^2\right)\times\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)+O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right),$$

or equivalently,

$$\begin{aligned}T_n &= \frac{2\,\mathrm{varlog}(\tau_{1:L})\,k_n}{S_n(\beta_n,\tau_{1:L})}\,\frac{\Delta_{1,n}}{\Delta_{2,n}}\\ &= \left(\frac{(\log(\lambda_n))^2k_n}{2\xi^2S_n(\beta_n,\tau_{1:L})} - \frac{\sqrt{2k_n}\log(\lambda_n)s_n(\beta_n,\tau_{1:L})}{\xi S_n(\beta_n,\tau_{1:L})}\Gamma + \frac{1+S_n(\beta_n,\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}\Gamma^2\right)\\ &\times\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)+O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right).\end{aligned}$$

The result follows. □

Proof of Remark 3. From Proposition 4.2, one has

$$\mathbb{P}\left(T_n \le \chi^2_{1,1-\gamma}\right) = \mathbb{P}\left(a_n\Gamma^2 + b_n\Gamma + C_n \le 0\right)$$

where

$$\begin{aligned}a_n &= \frac{1+S_n(\beta_n,\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})},\\ b_n &= -\frac{\sqrt{2k_n}\log(\lambda_n)s_n(\beta_n,\tau_{1:L})}{\xi S_n(\beta_n,\tau_{1:L})}\end{aligned}$$

are two non random sequences and

$$\begin{aligned}C_n &= \frac{(\log(\lambda_n))^2k_n}{2\xi^2S_n(\beta_n,\tau_{1:L})} - \chi^2_{1,1-\gamma}(1+\varepsilon_n),\\ \varepsilon_n &= O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)+O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\end{aligned}$$

are two random variables. It straightforwardly follows that

$$\Omega_{1,n} := -\frac{b_n}{2a_n} = \frac{\sqrt{k_n}\log(\lambda_n)s_n(\beta_n,\tau_{1:L})}{\sqrt{2}\xi\left(1+S_n(\beta_n,\tau_{1:L})\right)} \text{ and } \frac{b_n^2}{4a_n^2} = \frac{k_n(\log(\lambda_n))^2s_n(\beta_n,\tau_{1:L})^2}{2\xi^2\left(1+S_n(\beta_n,\tau_{1:L})\right)^2}.$$

Moreover, in view of (i) and $\log(n)(1-\beta_n)\to 0$, one has:

$$-\frac{C_n}{a_n} = \chi^2_{1,1-\gamma}\frac{S_n(\beta_n,\tau_{1:L})}{1+S_n(\beta_n,\tau_{1:L})} - \frac{(\log(\lambda_n))^2k_n}{2\xi^2(1+S_n(\beta_n,\tau_{1:L}))}$$

$$+ O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right).$$

Besides, (i') is equivalent to $(\log(k_n))^2/k_n^{3/4} \vee \sqrt{(\log(k_n))^3/n} = o(\log(\lambda_n))$ and thus

$$\begin{aligned}\frac{b_n^2}{4a_n^2} - \frac{C_n}{a_n} &= \frac{k_n(\log(\lambda_n))^2}{2\xi^2\left(1+S_n(\beta_n,\tau_{1:L})\right)^2}\left(s_n(\beta_n,\tau_{1:L})^2 - 1 - S_n(\beta_n,\tau_{1:L})\right)\\ &\quad + \chi^2_{1,1-\gamma}\frac{S_n(\beta_n,\tau_{1:L})}{1+S_n(\beta_n,\tau_{1:L})} + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\\ &= \Omega_{2,n}\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right),\end{aligned}$$

where

$$\Omega_{2,n} = \frac{(\log(\lambda_n))^2 k_n}{2\xi^2}\frac{s_n(\beta_n,\tau_{1:L})^2 - 1 - S_n(\beta_n,\tau_{1:L})}{\left(1+S_n(\beta_n,\tau_{1:L})\right)^2} + \frac{S_n(\beta_n,\tau_{1:L})}{1+S_n(\beta_n,\tau_{1:L})}\chi^2_{1,1-\gamma}.$$

A second-order Taylor expansion shows that $s_n(\beta_n,\tau_{1:L})^2 - 1 - S_n(\beta_n,\tau_{1:L}) \to 0$ as $n\to\infty$. Besides, since $S_n(\beta_n,\tau_{1:L}) \sim (\log k_n)^2$ as $n\to\infty$, it follows from (ii) that $\Omega_{2,n} \to \chi^2_{1,1-\gamma}$ as $n\to\infty$. As a consequence, Ω_n is positive for n large enough. Hence,

$$\begin{aligned}\mathbb{P}\left(T_n \le \chi^2_{1,1-\gamma}\right) &= \mathbb{P}\left(\Gamma \in \left[-\frac{b_n}{2a_n} - \sqrt{\frac{b_n^2}{4a_n^2} - \frac{C_n}{a_n}}, -\frac{b_n}{2a_n} + \sqrt{\frac{b_n^2}{4a_n^2} - \frac{C_n}{a_n}}\right]\right)\\ &= \mathbb{P}\left(\Gamma \in \left[\Omega_{1,n} \pm \sqrt{\Omega_{2,n}}\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right)\right]\right),\end{aligned}$$

and the proposed approximation follows. □

Proof of Proposition 4.3. Following the steps of the proof of Proposition 4.2, one has:

$$\begin{aligned}\log\left(\frac{\hat{q}_1^{\mathrm{W}}(\alpha_{\ell,n}\,|\,\beta_n)}{\hat{q}_2^{\mathrm{W}}(\alpha_{\ell,n}\,|\,\beta_n)}\right) &= \log\left(\frac{q_1(\alpha_{\ell,n})}{q_2(\alpha_{\ell,n})}\right) + \log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)\left(\hat{\xi}_1^{\mathrm{H}}(\beta_n) - \hat{\xi}_2^{\mathrm{H}}(\beta_n)\right)\\ &\quad + \log\left(\frac{\widehat{q}_1(\beta_n)}{q_1(\beta_n)}\frac{q_2(\beta_n)}{\widehat{q}_2(\beta_n)}\right) + \xi\log\left(\frac{1-\beta_n}{1-\alpha_{\ell,n}}\right)(\theta_n - 1). \quad (29)\end{aligned}$$

First, we establish a result similar to those of Lemma A.4 adapted to our setting. Equations (17) and (18) yield:

$$\hat{\xi}_2^{\mathrm{H}}(\beta_n) \stackrel{d}{=} \frac{\theta_n}{k_n}\sum_{i=1}^{k_n}\mathcal{E}_i^{(2)} \stackrel{d}{=} \theta_n\xi + \frac{\theta_n\xi}{\sqrt{n(1-\beta_n)}}\Gamma_{\xi,2}\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)\right),$$

since $\{\mathcal{E}_1^{(2)},\ldots,\mathcal{E}_{k_n}^{(2)}\}$ are i.i.d. realizations of an exponential distribution with mean ξ. It thus comes

$$\hat{\xi}_1^{\mathrm{H}}(\beta_n) - \hat{\xi}_2^{\mathrm{H}}(\beta_n) \stackrel{d}{=} (1-\theta_n)\xi + \frac{\sqrt{1+\theta_n^2}\,\xi}{\sqrt{n(1-\beta_n)}}\Gamma_\xi\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)\right).$$

Similarly,

$$\log\left(\frac{\widehat{q}_2(\beta_n)}{q_2(\beta_n)}\right) \overset{d}{=} \theta_n\left\{\mathcal{E}^{(2)}_{n-k_n,n} - \xi\log(n/k_n)\right\}$$
$$\overset{d}{=} \frac{\theta_n\xi}{\sqrt{n(1-\beta_n)}}\Gamma_{q,j}\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)+O_{\mathbb{P}}(1-\beta_n)\right),$$

and one thus has

$$\log\left(\frac{\widehat{q}_1(\beta_n)}{q_1(\beta_n)}\right) - \log\left(\frac{\widehat{q}_2(\beta_n)}{q_2(\beta_n)}\right) \overset{d}{=} \frac{\sqrt{1+\theta_n^2}\xi}{\sqrt{n(1-\beta_n)}}\Gamma_q$$
$$\times\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)+O_{\mathbb{P}}(1-\beta_n)\right).$$

Moreover, $q_2(\alpha_{\ell,n}) = q_1(\alpha_{\ell,n})^{\theta_n}$ and $q_1(\alpha_{\ell,n}) = (\tau_\ell/n)^{-\xi}$, hence

$$\log\left(q_1(\alpha_{\ell,n})/q_2(\alpha_{\ell,n})\right) = \xi(\theta_n-1)\log(\tau_\ell/n)$$

and replacing in (29) yields

$$\log\left(\frac{\hat{q}_1^{\mathrm{W}}(\alpha_{\ell,n}\,|\,\beta_n)}{\hat{q}_2^{\mathrm{W}}(\alpha_{\ell,n}\,|\,\beta_n)}\right) = \xi(1-\theta_n)\log\left(\frac{n}{\tau_\ell}\right)$$
$$+\frac{\sqrt{1+\log\left(\frac{k_n}{\tau_\ell}\right)^2}}{\sqrt{k_n}}\xi\sqrt{1+\theta_n^2}\Gamma\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)+O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right),$$

where we have introduced $k_n = n(1-\beta_n)$. Using (i), one has $\log(k_n)/k_n^{3/4} \vee \sqrt{\log(k_n)/n} = o(\log(\theta_n))$ and taking account of $\log(n)(1-\beta_n) \to 0$ as $n\to\infty$, it follows

$$\Delta_{1,n} = \frac{\xi^2(1-\theta_n)^2\,\mathrm{smlog}(n/\tau_{1:L})}{4} + \frac{\sqrt{1+\theta_n^2}\xi^2(1-\theta_n)\mathfrak{s}_n(\beta_n,\tau_{1:L})\Gamma}{2\sqrt{k_n}}$$
$$+\frac{(1+S_n(\beta_n,\tau_{1:L}))\xi^2(1+\theta_n^2)\Gamma^2}{4k_n} + O_{\mathbb{P}}\left(\frac{\log(\theta_n)(\log(n))^2)}{k_n}\right)$$
$$+O_{\mathbb{P}}\left(\log(\theta_n)\log(n)\sqrt{\frac{1-\beta_n}{n}}\right)+O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right)+O_{\mathbb{P}}\left(\frac{(\log(k_n))^2)}{k_n^{3/2}}\right),$$

where it is recalled that

$$\mathfrak{s}_n(\beta_n,\tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L}\log\left(\frac{n}{\tau_\ell}\right)\sqrt{1+\log\left(\frac{k_n}{\tau_\ell}\right)^2}$$

and

$$\mathrm{smlog}(n/\tau_{1:L}) = \frac{1}{L}\sum_{\ell=1}^{L}\left(\log\left(\frac{n}{\tau_\ell}\right)\right)^2.$$

The condition $\log(\theta_n) = o(1/\sqrt{k_n})$ in *(ii)* ensures that the first and second $O_{\mathbb{P}}$ are bounded from above by the two others, hence

$$\Delta_{1,n} = \frac{\xi^2(1-\theta_n)^2\,\mathrm{smlog}(n/\tau_{1:L})}{4} + \frac{\sqrt{1+\theta_n^2}\xi^2(1-\theta_n)\mathfrak{s}_n(\beta_n,\tau_{1:L})\Gamma}{2\sqrt{k_n}} + \frac{(1+S_n(\beta_n,\tau_{1:L}))\xi^2(1+\theta_n^2)\Gamma^2}{4k_n} + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right) + O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^{3/2}}\right),$$

In addition, straightforward calculations lead to

$$\begin{aligned}\Delta_{2,n} &= \mathrm{varlog}(\tau_{1:L})\left(\frac{\hat{\xi}_1^{\mathrm{H}}(\beta_n)+\hat{\xi}_2^{\mathrm{H}}(\beta_n)}{2}\right)^2\\ &= \xi^2\,\mathrm{varlog}(\tau_{1:L})\left(\frac{1+\theta_n}{2}\right)^2\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right).\end{aligned}$$

Combining the previous two results yields the following asymptotic expansion

$$\begin{aligned}\frac{\Delta_{1,n}}{\Delta_{2,n}} &= \frac{\mathrm{smlog}(n/\tau_{1:L})(1-\theta_n)^2}{\mathrm{varlog}(\tau_{1:L})(1+\theta_n)^2}\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right)\\ &+ \frac{2\sqrt{1+\theta_n^2}(1-\theta_n)}{\sqrt{k_n}(1+\theta_n)^2}\frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})}{\mathrm{varlog}(\tau_{1:L})}\Gamma\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right)\\ &+ \frac{1+\theta_n^2}{(1+\theta_n)^2}\frac{1+S_n(\beta_n,\tau_{1:L})}{\mathrm{varlog}(\tau_{1:L})k_n}\Gamma^2\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right)\\ &+ O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right) + O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^{3/2}}\right).\end{aligned}$$

Condition *(ii)* implies $\log(\theta_n)^2(\log(n))^2/\sqrt{k_n} = o((\log(k_n))^2/k_n^{3/2})$ leading to

$$\begin{aligned}\frac{\Delta_{1,n}}{\Delta_{2,n}} &= \frac{\mathrm{smlog}(n/\tau_{1:L})(1-\theta_n)^2}{\mathrm{varlog}(\tau_{1:L})(1+\theta_n)^2} + \frac{2\sqrt{1+\theta_n^2}(1-\theta_n)}{\sqrt{k_n}(1+\theta_n)^2}\frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})}{\mathrm{varlog}(\tau_{1:L})}\Gamma\\ &+ \frac{1+\theta_n^2}{(1+\theta_n)^2}\frac{1+S_n(\beta_n,\tau_{1:L})}{\mathrm{varlog}(\tau_{1:L})k_n}\Gamma^2 + O_{\mathbb{P}}\left(\frac{(\log(\theta_n))^2(\log(n))^2}{\sqrt{k_n}}\right)\\ &+ O_{\mathbb{P}}\left(\frac{\log(\theta_n)\log(n)\log(k_n)}{k_n}\right) + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right) + O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^{3/2}}\right)\\ &= \frac{\mathrm{smlog}(n/\tau_{1:L})(1-\theta_n)^2}{\mathrm{varlog}(\tau_{1:L})(1+\theta_n)^2} + \frac{2\sqrt{1+\theta_n^2}(1-\theta_n)}{\sqrt{k_n}(1+\theta_n)^2}\frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})}{\mathrm{varlog}(\tau_{1:L})}\Gamma\\ &+ \frac{1+\theta_n^2}{(1+\theta_n)^2}\frac{1+S_n(\beta_n,\tau_{1:L})}{\mathrm{varlog}(\tau_{1:L})k_n}\Gamma^2 + O_{\mathbb{P}}\left(\frac{\log(k_n)}{n}\right) + O_{\mathbb{P}}\left(\frac{(\log(k_n))^2}{k_n^{3/2}}\right)\\ &= \frac{\mathrm{smlog}(n/\tau_{1:L})(1-\theta_n)^2}{\mathrm{varlog}(\tau_{1:L})(1+\theta_n)^2} + \frac{2\sqrt{1+\theta_n^2}(1-\theta_n)}{\sqrt{k_n}(1+\theta_n)^2}\frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})}{\mathrm{varlog}(\tau_{1:L})}\Gamma\\ &+ \frac{1+\theta_n^2}{(1+\theta_n)^2}\frac{1+S_n(\beta_n,\tau_{1:L})}{\mathrm{varlog}(\tau_{1:L})k_n}\Gamma^2\left(1+O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right).\end{aligned}$$

Therefore, the test statistics can be written as

$$\begin{aligned}
T_n &= \frac{2\,\mathrm{varlog}(\tau_{1:L})\,k_n}{S_n(\beta_n,\tau_{1:L})}\,\frac{\Delta_{1,n}}{\Delta_{2,n}} \\
&= \frac{2\,\mathrm{smlog}(n/\tau_{1:L})k_n(1-\theta_n)^2}{S_n(\beta_n,\tau_{1:L})(1+\theta_n)^2} + \frac{4\mathfrak{s}_n(\beta_n,\tau_{1:L})\sqrt{k_n}\sqrt{1+\theta_n^2}(1-\theta_n)}{S_n(\beta_n,\tau_{1:L})(1+\theta_n)^2}\Gamma \\
&\quad + \frac{2(1+\theta_n^2)}{(1+\theta_n)^2}\frac{1+S_n(\beta_n,\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}\Gamma^2\left(1+O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right),
\end{aligned}$$

or, under the condition $\log(n)(1-\beta_n)\to 0$,

$$\begin{aligned}
T_n &= 2\left(\frac{\mathrm{smlog}(n/\tau_{1:L})k_n(1-\theta_n)^2}{S_n(\beta_n,\tau_{1:L})(1+\theta_n)^2} + 2\frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})\sqrt{k_n}\sqrt{1+\theta_n^2}(1-\theta_n)}{S_n(\beta_n,\tau_{1:L})(1+\theta_n)^2}\Gamma\right. \\
&\quad \left. + \frac{(1+\theta_n^2)}{(1+\theta_n)^2}\frac{1+S_n(\beta_n,\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}\Gamma^2\right) \\
&\quad \times\left(1+O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)\right),
\end{aligned}$$

and the result is proved. □

Proof of Remark 4. Let us introduce the two non-random sequences

$$\begin{aligned}
a_n &= 2\frac{(1+\theta_n^2)}{(1+\theta_n)^2}\frac{1+S_n(\beta_n,\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}, \\
b_n &= 4\sqrt{k_n}\frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}\frac{\sqrt{1+\theta_n^2}(1-\theta_n)}{(1+\theta_n)^2},
\end{aligned}$$

as well as the two random variables

$$\begin{aligned}
C_n &= 2k_n\frac{\mathrm{smlog}(n/\tau_{1:L})}{S_n(\beta_n,\tau_{1:L})}\frac{(1-\theta_n)^2}{(1+\theta_n)^2} - \chi^2_{1,1-\gamma}(1+\varepsilon_n), \\
\varepsilon_n &= O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right)+O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right),
\end{aligned}$$

so that $\mathbb{P}\left(T_n\le\chi^2_{1,1-\gamma}\right)=\mathbb{P}\left(a_n\Gamma^2+b_n\Gamma+C_n\le 0\right)$. First,

$$\Psi_{1,n} := -\frac{b_n}{2a_n} = \frac{\sqrt{k_n}(\theta_n-1)\mathfrak{s}_n(\beta_n,\tau_{1:L})}{\sqrt{1+\theta_n^2}\,(1+S_n(\beta_n,\tau_{1:L}))}$$

with

$$\begin{aligned}
\frac{b_n^2}{4a_n^2} &= \frac{k_n(\theta_n-1)^2\mathfrak{s}_n^2(\beta_n,\tau_{1:L})}{(1+\theta_n^2)\,(1+S_n(\beta_n,\tau_{1:L}))^2}, \\
-\frac{C_n}{a_n} &= \frac{\chi^2_{1,1-\gamma}}{2}\frac{(1+\theta_n)^2S_n(\beta_n,\tau_{1:L})}{(1+\theta_n^2)(1+S_n(\beta_n,\tau_{1:L}))}
\end{aligned}$$

$$- \frac{(\theta_n - 1)^2 k_n \,\mathrm{smlog}(n/\tau_{1:L})}{(1+\theta_n^2)(1+S_n(\beta_n,\tau_{1:L}))} + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right),$$

hence, under (i'), one has $(\log(k_n))^2/k_n^{3/4} \vee \sqrt{(\log(k_n))^3/n} = o(\log(\theta_n))$, and therefore

$$\frac{b_n^2}{4a_n^2} - \frac{C_n}{a_n} = \Psi_{2,n}\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k_n}}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(k_n)}\right)\right),$$

where

$$\begin{aligned}\Psi_{2,n} =& \frac{(\theta_n-1)^2 k_n}{(1+\theta_n^2)} \frac{\mathfrak{s}_n(\beta_n,\tau_{1:L})^2 - (1+S_n(\beta_n,\tau_{1:L}))\,\mathrm{smlog}(n/\tau_{1:L})}{(1+S_n(\beta_n,\tau_{1:L}))^2} \\ &+ \frac{(1+\theta_n)^2}{(1+\theta_n^2)} \frac{S_n(\beta_n,\tau_{1:L})}{1+S_n(\beta_n,\tau_{1:L})} \frac{\chi^2_{1,1-\gamma}}{2}.\end{aligned}$$

It appears that $\Psi_{2,n} \to \chi^2_{1,1-\gamma}$ as $n \to \infty$ under the condition $\log(n) = O(\log(k_n))$ and thus $\Psi_{2,n} > 0$ for n large enough. As a consequence, one has

$$\begin{aligned}\mathbb{P}\left(T_n \le \chi^2_{1,1-\gamma}\right) &= \mathbb{P}\left(\Gamma \in \left[-\frac{b_n}{2a_n} - \sqrt{\frac{b_n^2}{4a_n^2} - \frac{C_n}{a_n}}, -\frac{b_n}{2a_n} + \sqrt{\frac{b_n^2}{4a_n^2} - \frac{C_n}{a_n}}\right]\right) \\ &= \mathbb{P}\left(\Gamma \in [I_-, I_+]\right), \text{ where} \\ I_\pm = \Psi_{1,n} &\pm \sqrt{\Psi_{2,n}}\left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right) + O_{\mathbb{P}}\left(\frac{1-\beta_n}{\log(n(1-\beta_n))}\right)\right),\end{aligned}$$

leading to the proposed approximation. □

Appendix B: Simulation results for $J > 2$ samples

In extension of the simulation study in Section 5, we consider some examples where $J > 2$, more specifically $J \in \{5, 10, 15\}$:

- $(\mathcal{MF})$ $X_1, \ldots, X_J$ all follow an unit $(\xi = 1)$ Fréchet distribution. We are thus supposed to reject (H_0) with probability 0.05. This distribution has no mean, and the classical ANOVA is thus not applicable.
- $(\mathcal{MM})$ For all $j = 1, \ldots, J$, X_j follows an unit Pareto distribution if $j = 4k+1$, an unit Fréchet distribution if $j = 4k+2$, an unit Burr(-1) distribution if $j = 4k+3$ and an unit GPD if $j = 4(k+1)$ $(k \in \mathbb{N})$. The extreme quantiles of all these distributions are asymptotically equivalent and (H_0) is thus satisfied.
- $(\mathcal{MP})$ $X_1, \ldots, X_{J-1}$ all follow an unit Pareto distribution, and X_J follows a Pareto distribution with tail index θ. If $\theta \ne 1$, then (H_0) obviously has to be rejected. We propose to consider the cases $\theta = 0.8$ and $\theta = 1.2$.
- $(\mathcal{CP})$ $X_1, \ldots, X_{J-1}$ all follow an unit Burr distribution (with $\rho = -1$), and $X_J \overset{d}{=} \pi X_1 + (1-\pi)\sqrt{X_1}$ is a contaminated Burr distribution

($\pi \in [0,1]$). When $\pi < 1$, the quantiles of X_J are asymptotically equivalent to $\pi(1-\tau)^{-1}$ and differ from those of $X_1, \dots, X_{J-1}$ which are asymptotically equivalent to $(1-\tau)^{-1}$, and we thus hope to reject (H_0). We consider the cases $\pi \in \{0.2, 0.5, 0.8\}$.

The results are reported in Figure 7. Unsurprisingly, the error rates are increasing with the number J of samples. Comparing the first two examples, it seems that the type I error is not significantly sensitive to the underlying distribution (for two distributions with the same second-order parameter ρ, confirming the observation made in the previous paragraph). Example $(\mathcal{MP})$ shows that, in more than 65% of the replications, the ANOVEX procedure is able detect whether a sample over five has a slightly lower tail index (0.8 vs 1 for the four other samples). The result drops at 50% when a sample has a slightly

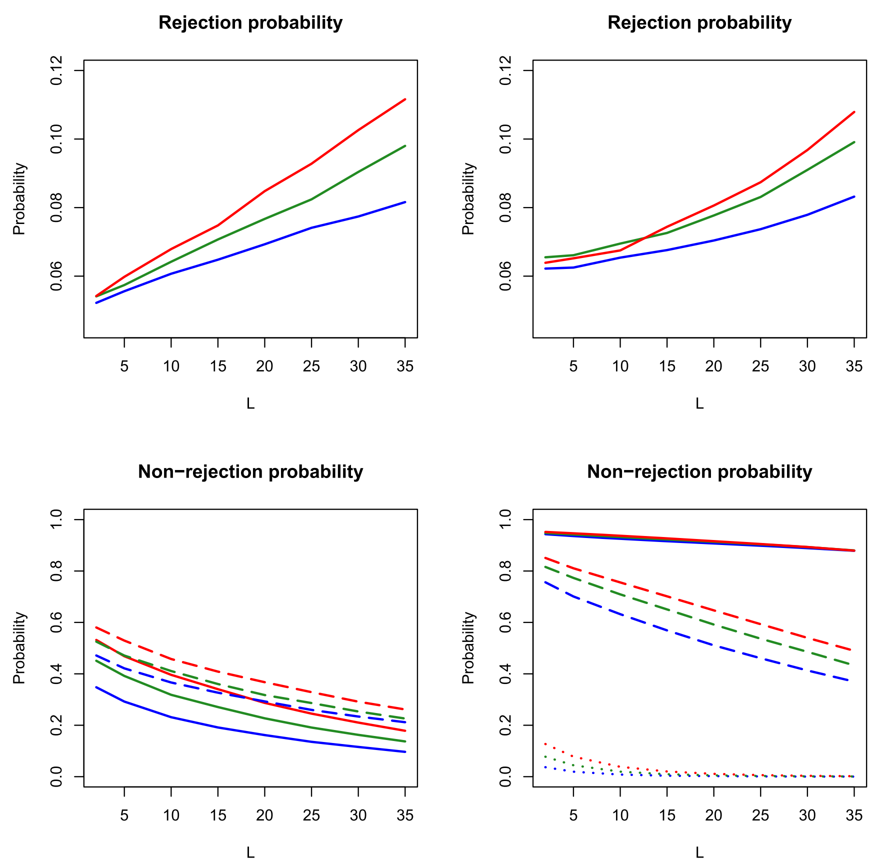

FIG 7. *From left to right, top to bottom: examples* $(\mathcal{MF})$, $(\mathcal{MM})$, $(\mathcal{MP})$ *with* $\theta = 0.8$ *(solid curves) and* $\theta = 1.2$ *(dashed curves), and* $(\mathcal{CP})$ *with* $\pi = 0.2$ *(dotted curves),* $\pi = 0.5$ *(dashed curves) and* $\pi = 0.8$ *(solid curves). In all examples,* $n = 1{,}000$, $\beta_n = 0.9$ *and* $J = 5$ *(blue curves),* $J = 10$ *(green curves) and* $J = 15$ *(red curves).*

greater tail index (1.2 vs 1). The ANOVEX test is thus efficient to discriminate samples with different tail indices. However, the latter is less efficient when all the samples share the same tail index, but have different scale parameters. Indeed, example ($\mathcal{CP}$) shows that, when the mixture parameter π is slightly lower than 1 (0.8), the test is rejected with a rate of only 10% (when $J = 5$, 10 or 15). This rate is obviously much better when $\pi = 0.5$ (more than 25 % when $J = 5$) and $\pi = 0.2$ (almost 100%).

Appendix C: Additional figures for real-data application

Figures 8 and 9 present additional information related to the financial data analysed in Section 6.

Acknowledgments

The authors would like to warmly thank two anonymous referees, an Associate Editor and the Editor for their numerous very constructive comments that helped to improve the content and quality of this paper.

Funding

This work is part of the ANOVEX project funded by the French national research institutes INRAE and Inria within the scope of their collaborations on the topic of environmental risks. This work is further supported by the French National Research Agency (ANR) in the framework of the Investissements d'Avenir Program (ANR-15-IDEX-02). The first author acknowledges the support of the Chair "Stress Test, Risk Management and Financial Steering", led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas.

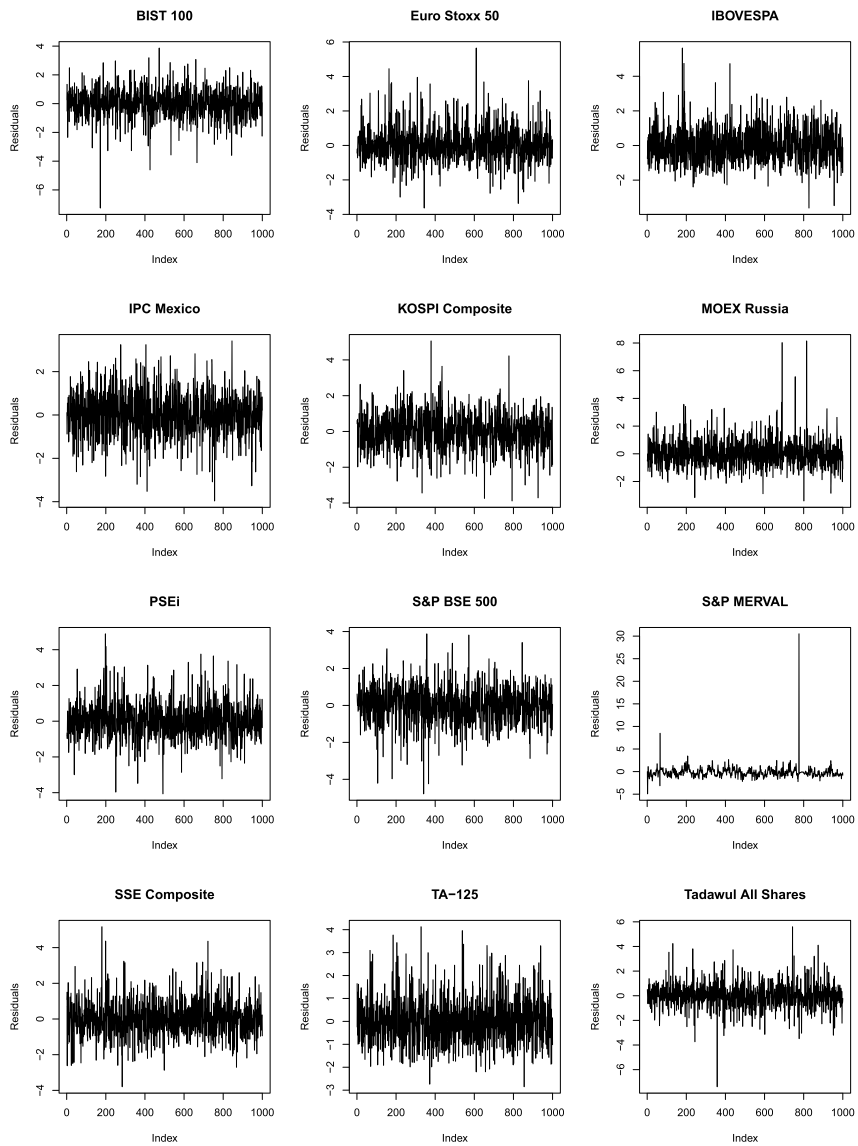

FIG 8. $n = 1{,}000$ *ARMA(1,1)-GARCH(1,1) residuals associated with* $J = 12$ *times series of negative daily log-returns of stock market indices.*

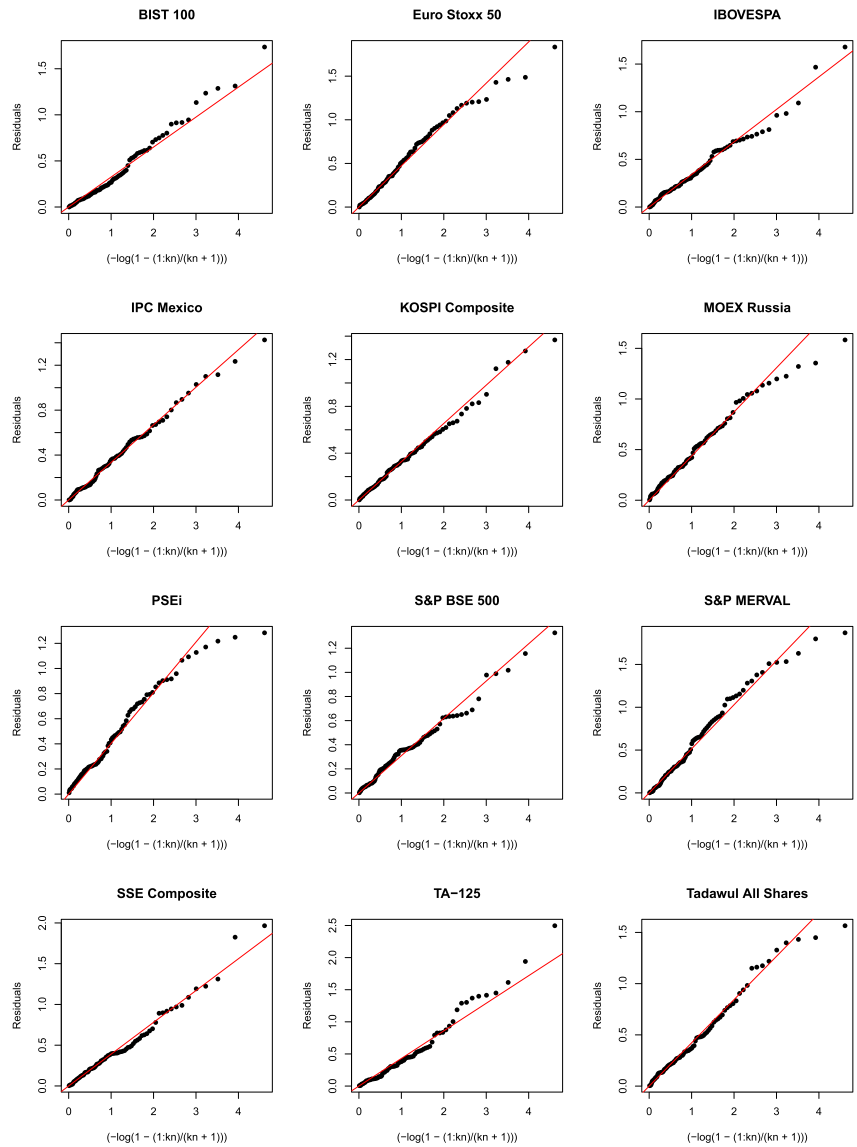

FIG 9. *Exponential QQ-plots of the weighted log-spacings associated with the daily log-returns computed from* $J = 12$ *stock market indices.*

References

[1] Artzner, P., Delbaen, F., Eber, J. M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9** 203–228. MR1850791

[2] Bacro, J.-N., Gaetan, C., Opitz, T. and Toulemonde, G. (2020). Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association* **115** 555–569. MR4107657

[3] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications* **558**. John Wiley & Sons. MR2108013

[4] Blitzstein, J. and Hwang, J. (2019). *Introduction to Probability*. Chapman and Hall/CRC. MR3929729

[5] Caeiro, F., Gomes, M. I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *REVSTAT-Statistical Journal* **3** 113–136. MR2259357

[6] Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review* **51** 661–703. MR2563829

[7] Daouia, A., Stupfler, G. and Usseglio-Carleve, A. (2023). Inference for extremal regression with dependent heavy-tailed data. *The Annals of Statistics* **51** 2040–2066. MR4678795

[8] Daouia, A., Stupfler, G. and Usseglio-Carleve, A. (2024). An expectile computation cookbook. *Statistics and Computing* **34** 103. MR4721861

[9] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **52** 393–425. MR1086795

[10] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction* **21**. Springer. MR2234156

[11] Deuber, D., Li, J., Engelke, S. and Maathuis, M. H. (2024). Estimation and inference of extremal quantile treatment effects for heavy-tailed distributions. *Journal of the American Statistical Association* **119** 2206–2216. MR4797934

[12] Dierckx, G. and Teugels, J. L. (2010). Change point analysis of extreme values. *Environmetrics* **21** 661–686. MR2838438

[13] Dupuis, D., Sun, Y. and Wang, H. J. (2015). Detecting change-points in extremes. *Statistics and Its Interface* **8** 19–31. MR3320386

[14] Einmahl, J. H. J., Haan, L. and Zhou, C. (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **78** 31–51. MR3453645

[15] El-Methni, J., Gardes, L. and Girard, S. (2014). Nonparametric estimation of extreme risks from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics* **41** 988–1012. MR3277034

[16] Farkas, S., Heranval, A., Lopez, O. and Thomas, M. (2024). Generalized Pareto Regression Trees for extreme events analysis. *Extremes* **27** 437–477. MR4778953

[17] Girard, S., Stupfler, G. and Usseglio-Carleve, A. (2021). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *The Annals of Statistics* **49** 3358–3382. MR4352533

[18] Gnecco, N., Terefe, E. M. and Engelke, S. (2024). Extremal random forests. *Journal of the American Statistical Association*. In press. https://doi.org/10.1080/01621459.2023.2300522.

[19] Gomes, M. I. and Martins, M. J. (2002). “Asymptotically unbiased” estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5** 5–31. MR1947785

[20] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3** 1163–1174. MR0378204

[21] Hoga, Y. (2017). Testing for changes in (extreme) VaR. *The Econometrics Journal* **20** 23–51. MR3636960

[22] Hollander, M., Wolfe, D. A. and Chicken, E. (2013). *Nonparametric Statistical Methods*. John Wiley & Sons. MR3221959

[23] Hsing, T. (1991). On tail index estimation using dependent data. *The Annals of Statistics* **19** 1547–1569. MR1126337

[24] Jondeau, E., Poon, S. H. and Rockinger, M. (2007). *Financial Modeling Under Non-Gaussian Distributions*. Springer Science & Business Media. MR2260802

[25] Jondeau, E. and Rockinger, M. (2003). Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance* **10** 559–581.

[26] Katz, R., Parlange, M. and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources* **25** 1287–1304.

[27] Klartag, B. (2009). A Berry-Esseen type inequality for convex bodies with an unconditional basis. *Probability Theory and Related Fields* **145** 1–33. MR2520120

[28] Koh, J., Pimont, F., Dupuy, J. L. and Opitz, T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *The Annals of Applied Statistics* **17** 560–582. MR4539044

[29] Maillart, A. and Robert, C. Y. (2023). Tail index partition-based rules extraction with application to tornado damage insurance. *ASTIN Bulletin: The Journal of the IAA* **53** 258–284. MR4594688

[30] Marden, J. and Muyot, M. (1995). Rank tests for main and interaction effects in analysis of variance. *Journal of the American Statistical Association* **90** 1388–1398. MR1379482

[31] Mikosch, T. (1997). Heavy-tailed modelling in insurance. *Communications in Statistics. Stochastic Models* **13** 799–815. MR1482293

[32] Mougeot, M. and Tribouley, K. (2010). Procedure of test to compare the tail indices. *Annals of the Institute of Statistical Mathematics* **62** 383–412. MR2592104

[33] Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55** 819–847. MR0906565

[34] Padoan, S. and Stupfler, G. (2022). Joint inference on extreme expectiles for multivariate heavy-tailed distributions. *Bernoulli* **28** 1021–1048. MR4388928

[35] Peng Zhong, T. O. Manuela Brunner and Huser, R. (2024). Spatial modeling and future projection of extreme precipitation extents. *Journal of the American Statistical Association* 1–22. https://doi.org/10.1080/01621459.2024.2408045.

[36] Pereira, J. M. C. and Turkman, K. F. (2019). Statistical models of vegetation fires: Spatial and temporal patterns. In *Handbook of Environmental and Ecological Statistics* 401–420. Chapman and Hall/CRC. MR3889904

[37] Pisarenko, V. and Rodkin, M. (2010). *Heavy-tailed Distributions in Disaster Analysis* **30**. Springer Science & Business Media.

[38] Rachev, S. (2003). *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance, Book 1.* Elsevier.

[39] Rényi, A. (1953). On the theory of order statistics. *Acta Mathematica Academiae Scientiarum Hungarica* **4** 191–231. MR0061792

[40] Rizzo, M. and Székely, G. (2010). DISCO analysis: A nonparametric extension of analysis of variance. *The Annals of Applied Statistics* **4** 1034 – 1055. MR2758432

[41] Scheffe, H. (1999). *The Analysis of Variance* **72**. John Wiley & Sons. MR1673563

[42] Süveges, M. and Davison, A. C. (2010). Model misspecification in peaks over threshold analysis. *The Annals of Applied Statistics* **4** 203–221. MR2758170

[43] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association* **73** 812–815. MR0521329

[44] Worms, J. and Worms, R. (2015). A test for comparing tail indices for heavy-tailed distributions via empirical likelihood. *Communications in Statistics-Theory and Methods* **44** 3289–3302. MR3384076

[45] Yadav, R., Huser, R. and Opitz, T. (2021). Spatial hierarchical modeling of threshold exceedances using rate mixtures. *Environmetrics* **32** e2662. MR4248739