Optimal weighted pooling for inference about the tail index and extreme quantiles

ABDELAATI DAOUIA [a], SIMONE A. PADOAN [b] and GILLES STUPFLER [c]

[1]*Toulouse School of Economics, University of Toulouse Capitole, France,* [a]*abdelaati.daouia@tse-fr.eu*
[2]*Department of Decision Sciences, Bocconi University, via Roentgen 1, 20136 Milano, Italy,* [b]*simone.padoan@unibocconi.it*
[3]*Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France,* [c]*gilles.stupfler@univ-angers.fr*

This paper investigates pooling strategies for tail index and extreme quantile estimation from heavy-tailed data. To fully exploit the information contained in several samples, we present general weighted pooled Hill estimators of the tail index and weighted pooled Weissman estimators of extreme quantiles calculated through a nonstandard geometric averaging scheme. We develop their large-sample asymptotic theory across a fixed number of samples, covering the general framework of heterogeneous sample sizes with different and asymptotically dependent distributions. Our results include optimal choices of pooling weights based on asymptotic variance and MSE minimization. In the important application of distributed inference, we prove that the variance-optimal distributed estimators are asymptotically equivalent to the benchmark Hill and Weissman estimators based on the unfeasible combination of subsamples, while the AMSE-optimal distributed estimators enjoy a smaller AMSE than the benchmarks in the case of large bias. We consider additional scenarios where the number of subsamples grows with the total sample size and effective subsample sizes can be low. We extend our methodology to handle serial dependence and the presence of covariates. Simulations confirm the statistical inferential theory of our pooled estimators. Two applications to real weather and insurance data are showcased.

Keywords: Extreme values; Heavy tails; Inference; Pooling; Testing

1. Introduction

The question of how to accurately model extreme events arises in many fields of statistical applications, such as insurance and environmental science, where the former is concerned with very large claims detrimental to a company's solvability, and the latter focuses on extreme events that affect the Earth's natural processes. Studying these events typically involves analyzing a well-chosen random variable Y, such as claim amounts in insurance or rainfall levels in environmental science. In applications of this nature, the data is often collected at different locations in a geographical region: insurance companies compile portfolios divided across states, regions, or cities, and weather agencies collect data from several weather stations. An important and contemporary problem in extreme value analysis is then to find a statistically efficient way to combine estimates calculated separately from different data samples.

In this article we focus on the question of inference in this multi-sample context for the tail index and extreme quantiles of heavy-tailed data. Our approach is based on the use of a general terminology and theory of pooling applied to different samples whose distributions have a common target parameter, given in our setup by the tail index or an extreme quantile. Instead of naively averaging the subsample estimators, it is of interest, both from a theoretical and a practical perspective, to construct a general class of weighted pooled estimators and to establish a fully data-driven inferential procedure integrating the optimal choice of weights. In particular, we develop the asymptotic theory of the optimally pooled tail index and extreme quantile estimators, under weak technical conditions, covering both scenarios where the number of available subsamples is bounded or growing with the total sample size, as well as the general situation where the observed data can be dependent within and/or across subsamples. Pooling has a rich history dating back to [4] for the estimation of the common mean of several samples. The idea of combining pooling and extreme value techniques has, as far as we know, originally been suggested by [14] in climate science problems.

The first contribution of this paper is a joint asymptotic normality result for Hill estimators [13] calculated from a fixed number m of samples of heavy-tailed data. In particular, we allow for different sample sizes, effective sample sizes, marginal distributions, and for dependence across samples. We apply this general result to design optimal pooling strategies of subsample Hill estimators for tail index estimation. We consider optimal weights that minimize either the asymptotic variance or the Asymptotic Mean Squared Error (AMSE) of the pooled estimator. These developments rely on a very general theory that we derive for a generic weighted pooled estimator built from m subsample estimators for a common unknown parameter. This theory comes into play when the subsample estimators are jointly asymptotically normal, and can be biased and correlated. To the best of our knowledge, no such unrestricted approach has been fully investigated. We also construct bias-reduced versions of the proposed pooled tail index estimators. Then we discuss the fundamental extreme value problem of estimating extreme quantiles either locally for each sample in the tail homogeneous setting of equal marginal tail indices, where tail quantiles are possibly only asymptotically proportional across subsamples, or globally by pooling subsample extreme quantile Weissman estimators [22] in the more restrictive tail homoskedastic setting of asymptotically equivalent marginal tail quantiles. Our approach relies on a specific weighted geometric pooling scheme, particularly relevant for extreme quantiles, as opposed to arithmetic averaging naturally used for pooling ordinary quantiles [15]. Moreover, we explore inferential aspects of pooling for extreme values by constructing likelihood ratio-type tests for either tail homogeneity or tail homoskedasticity, as well as Gaussian asymptotic confidence intervals for the tail index and extreme quantiles. Taking into account the dependence existing between samples is especially crucial for environmental applications, in which dependence across sites, representing spatial dependence, should be properly handled. Moreover, our proposed inferential methodology avoids having to resort to bootstrap, whose calibration can be difficult in heavy-tailed settings.

We also consider distributed inference as an important application of our general theory. In this case, due to computational costs or privacy restrictions, the data in sample number j can only be processed by the jth machine having collected the data, with very restricted or no communication allowed between the m machines, before an end user operating from a central machine conducts distributed inference from limited information transmitted by each machine. Under this setup, we examine and compare the asymptotic theory of the distributed tail index and extreme quantile estimators to the behavior of their respective benchmark Hill and Weissman estimators based on the unfeasible direct combination of subsamples. We extend this theory further by considering first the case when effective sample sizes are highly unbalanced among machines, and then the case of a growing number of machines $m = m(n) \to \infty$ with the total sample size n. Finally, we tackle the problem of serial dependence within the data in the presence of covariates, showing how appropriately filtering the observations allows to recover the asymptotic theory from independent observations. This is highly relevant to inference in environmental and financial applications, in which the individual data samples are typically time series.

The paper is organized as follows. Section 2 develops our general pooling theory for tail index and extreme quantile estimation, while Section 3 focuses on the special framework of distributed inference. Section 4 extends our methodology to handle serial dependence and the presence of covariates through filtering. Section 5 illustrates the usefulness of the proposed methods through a simulation study and applications to insurance and weather data. Section 6 concludes. The supplement to this article contains additional theoretical results and all the proofs, with further details on our simulation study. Our methods and data have been incorporated into the open-source `R` package `ExtremeRisks`.

2. Pooling extreme value estimators

2.1. Pooled Hill estimators of the tail index

Let $\boldsymbol{X} = (X_1, \ldots, X_m)^\top$ denote an $m-$dimensional random vector, and $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,m})^\top$ $(i \geq 1)$ denote independent copies of $\boldsymbol{X}$. We assume that the available data consists of the $X_{i,j}$, for $1 \leq j \leq m$ and $1 \leq i \leq n_j = n_j(n) \to \infty$ as $n \to \infty$, with $n = \sum_{j=1}^m n_j$ being the total number of univariate data points available across all samples. This setup allows for the very frequent situation where subsample sizes are not equal and/or the data points at each site have not been recorded over the same time period. We focus on the general framework where the components X_j of the random vector $\boldsymbol{X}$ have continuous, right heavy-tailed distribution functions F_j, with associated survival functions $\overline{F}_j = 1 - F_j$ and tail quantile functions $U_j : t \mapsto \inf\{x \in \mathbb{R} \,|\, 1/\overline{F}_j(x) \geq t\}$ that satisfy

$C_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{A})$ For any $j \in \{1, \ldots, m\}$, the function U_j satisfies the second-order condition:

$C_2(\gamma_j, \rho_j, A_j)$ U_j is second-order regularly varying in a neighborhood of $+\infty$ with index $\gamma_j > 0$, second-order parameter $\rho_j \leq 0$ and an auxiliary function A_j having constant sign and converging to 0 at infinity, that is,

$$\forall x > 0, \lim_{t\to\infty} \frac{1}{A_j(t)} \left[\frac{U_j(tx)}{U_j(t)} - x^{\gamma_j} \right] = \begin{cases} x^{\gamma_j} \dfrac{x^{\rho_j} - 1}{\rho_j} & \text{if } \rho_j < 0, \\ x^{\gamma_j} \log x & \text{if } \rho_j = 0. \end{cases}$$

In this condition $|A_j|$ is regularly varying with index ρ_j [by Theorems 2.3.3 and 2.3.9 in 5], meaning that the larger $|\rho_j|$ is, the smaller the gap between the right tail of U_j and a purely Pareto tail. All usual heavy-tailed distributions satisfy these conditions, see Table 2.1 on p.59 of [1] for a list of examples.

To incorporate the dependence between samples into the inference procedure, we assume an appropriate pairwise tail dependence structure based on the functions $\overline{C}_{j,\ell}(u, v) = \mathbb{P}(\overline{F}_j(X_j) \leq u, \overline{F}_\ell(X_\ell) \leq v)$ $(u, v \in [0, 1])$ that are essentially the bivariate survival copulae of $\boldsymbol{X}$, namely:

$\mathcal{J}(\boldsymbol{R})$ For any (j, ℓ) with $j \neq \ell$, there is a function $R_{j,\ell}$ such that $\lim_{s\to\infty} s\,\overline{C}_{j,\ell}(x_j/s, x_\ell/s) = R_{j,\ell}(x_j, x_\ell)$ for any $(x_j, x_\ell) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}$.

This condition imposes the existence of a limiting dependence structure in the joint right tail of X_j and X_ℓ, given by the *tail copula* $R_{j,\ell}$ (see [18]). It can be viewed as a minimal assumption when it comes to assessing the dependence structure between extreme value estimators.

After ordering the data in the jth sample as $X_{1:n_j,j} \leq X_{2:n_j,j} \leq \cdots \leq X_{n_j:n_j,j}$, we introduce the marginal Hill estimators $\widehat{\gamma}_j(k_j) = k_j^{-1} \sum_{i=1}^{k_j} \log(X_{n_j-i+1:n_j,j}/X_{n_j-k_j:n_j,j})$ which involve the top $(k_j + 1)$ highest order statistics in each sample, for $k_j = k_j(n) \geq 1$. The integer k_j is the *effective sample size* in sample j, and we set $k = \sum_{j=1}^m k_j$ to be the *total effective sample size* in the vector of estimators $\widehat{\boldsymbol{\gamma}}_n = \widehat{\boldsymbol{\gamma}}_n(k_1, \ldots, k_m) = (\widehat{\gamma}_1(k_1), \ldots, \widehat{\gamma}_m(k_m))^\top$. Our ultimate interest is in the case $\boldsymbol{\gamma} = \gamma\mathbf{1}$ where the γ_j are equal to a common γ estimated by $\widehat{\gamma}_n(\boldsymbol{\omega}) = \widehat{\gamma}_n(\omega_1, \ldots, \omega_m) = \sum_{j=1}^m \omega_j \widehat{\gamma}_j(k_j) = \boldsymbol{\omega}^\top \widehat{\boldsymbol{\gamma}}_n$, with $\boldsymbol{\omega}^\top \mathbf{1} = 1$. The asymptotic distribution of any element within this class of estimators is stated in the following theorem, along with the joint asymptotic normality of $\widehat{\gamma}_j(k_j)$, for $j \in \{1, \ldots, m\}$. Here and throughout the article the symbols $\xrightarrow{d}$ and $\xrightarrow{\mathbb{P}}$ respectively stand for convergence in distribution and in probability, and all convergences of sequences should naturally be understood as holding when $n \to \infty$.

Theorem 1. *Assume that conditions $C_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{A})$ and $\mathcal{J}(\boldsymbol{R})$ hold. Suppose that $k_j = k_j(n) \to \infty$ with $k_j/n_j \to 0$, $n_1/n_j \to b_j \in (0, \infty)$, $k_1/k_j \to c_j \in (0, \infty)$ (with $b_1 = c_1 = 1$) and $\sqrt{k_j} A_j(n_j/k_j) \to \lambda_j \in$*

$\mathbb{R}$ for any $j \in \{1,\dots,m\}$. Let the weight vector $\boldsymbol{\omega} = (\omega_1,\dots,\omega_m)^\top$ be such that $\boldsymbol{\omega}^\top \mathbf{1} = 1$ and define a vector $\boldsymbol{B}$ and symmetric matrix $\mathbf{V}$ by

$$\boldsymbol{B} = \left(\frac{\lambda_1}{1-\rho_1},\dots,\frac{\lambda_m}{1-\rho_m}\right)^\top \text{ and } \mathbf{V}_{j,\ell} = \begin{cases} \gamma_j^2 & \text{if } j=\ell, \\ \gamma_j\gamma_\ell \dfrac{R_{j,\ell}(b_j c_\ell, b_\ell c_j)}{\max(b_j,b_\ell)\sqrt{c_j}\sqrt{c_\ell}} & \text{if } j<\ell. \end{cases}$$

Then $(\sqrt{k_1}(\widehat{\gamma}_1(k_1)-\gamma_1),\dots,\sqrt{k_m}(\widehat{\gamma}_m(k_m)-\gamma_m))^\top \xrightarrow{d} \mathcal{N}(\boldsymbol{B},\mathbf{V})$. In particular, if $\boldsymbol{\gamma} = \gamma\mathbf{1}$, then $\sqrt{k}(\widehat{\gamma}_n(\boldsymbol{\omega})-\gamma) \xrightarrow{d} \mathcal{N}(\boldsymbol{\omega}^\top\boldsymbol{B_c}, \boldsymbol{\omega}^\top\mathbf{V_c}\boldsymbol{\omega})$, with

$$\boldsymbol{B_c} = \sqrt{\sum_{i=1}^m c_i^{-1}}\left(\sqrt{c_1}\frac{\lambda_1}{1-\rho_1},\dots,\sqrt{c_m}\frac{\lambda_m}{1-\rho_m}\right)^\top$$

$$\text{and } [\mathbf{V_c}]_{j,\ell} = \left(\sum_{i=1}^m c_i^{-1}\right)\begin{cases} \gamma^2 c_j & \text{if } j=\ell, \\ \gamma^2\dfrac{R_{j,\ell}(b_j c_\ell, b_\ell c_j)}{\max(b_j,b_\ell)} = \gamma^2\dfrac{b_j c_j}{\max(b_j,b_\ell)} R_{j,\ell}(c_\ell/c_j, b_\ell/b_j) & \text{if } j<\ell. \end{cases}$$

The matrix $\mathbf{V}$ is positive definite if and only if $\mathbf{V_c}$ is so, and hence we have the following results on optimal weight choices:

1. *(Variance-optimal weights) There is a unique solution to the minimization problem of $\boldsymbol{\omega}^\top\mathbf{V_c}\boldsymbol{\omega}$ subject to the constraint $\boldsymbol{\omega}^\top\mathbf{1}=1$, which is*

$$\boldsymbol{\omega}^{(\mathrm{Var})} = \frac{\mathbf{V_c}^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{V_c}^{-1}\mathbf{1}}, \text{ and then } \sqrt{k}(\widehat{\gamma}_n(\boldsymbol{\omega}^{(\mathrm{Var})})-\gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\mathbf{1}^\top\mathbf{V_c}^{-1}\boldsymbol{B_c}}{\mathbf{1}^\top\mathbf{V_c}^{-1}\mathbf{1}}, \frac{1}{\mathbf{1}^\top\mathbf{V_c}^{-1}\mathbf{1}}\right).$$

2. *(AMSE-optimal weights) There is a unique solution to the minimization problem of $\mathrm{AMSE}(\boldsymbol{\omega}) = k^{-1}[(\boldsymbol{\omega}^\top\boldsymbol{B_c})^2 + \boldsymbol{\omega}^\top\mathbf{V_c}\boldsymbol{\omega}]$ subject to the constraint $\boldsymbol{\omega}^\top\mathbf{1}=1$, which is*

$$\boldsymbol{\omega}^{(\mathrm{AMSE})} = \frac{(1+\boldsymbol{B_c}^\top\mathbf{V_c}^{-1}\boldsymbol{B_c})\mathbf{V_c}^{-1}\mathbf{1} - (\mathbf{1}^\top\mathbf{V_c}^{-1}\boldsymbol{B_c})\mathbf{V_c}^{-1}\boldsymbol{B_c}}{(1+\boldsymbol{B_c}^\top\mathbf{V_c}^{-1}\boldsymbol{B_c})(\mathbf{1}^\top\mathbf{V_c}^{-1}\mathbf{1}) - (\mathbf{1}^\top\mathbf{V_c}^{-1}\boldsymbol{B_c})^2}.$$

The optimal value of $\mathrm{AMSE}(\boldsymbol{\omega})$ is

$$\mathrm{AMSE}(\boldsymbol{\omega}^{(\mathrm{AMSE})}) = \frac{1}{k}\times\frac{1+\boldsymbol{B_c}^\top\mathbf{V_c}^{-1}\boldsymbol{B_c}}{(1+\boldsymbol{B_c}^\top\mathbf{V_c}^{-1}\boldsymbol{B_c})(\mathbf{1}^\top\mathbf{V_c}^{-1}\mathbf{1}) - (\mathbf{1}^\top\mathbf{V_c}^{-1}\boldsymbol{B_c})^2}.$$

Finally, if $\widehat{\boldsymbol{\omega}}_n^\top\mathbf{1}=1$ with $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$, then the composite estimator $\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n)$ is $\sqrt{k}$–asymptotically equivalent to $\widehat{\gamma}_n(\boldsymbol{\omega})$ in the sense that $\sqrt{k}(\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n)-\widehat{\gamma}_n(\boldsymbol{\omega})) = \mathrm{o}_{\mathbb{P}}(1)$.

The proof requires applying a very general pooling result (Theorem A.1, that we state and prove in Section A of the Supplementary Material document) in conjunction with the assumption that the sample sizes n_j are asymptotically proportional (possibly unbalanced) and so are the effective sample sizes k_j. This ensures that none of the $\widehat{\gamma}_j(k_j)$ imposes its limiting distribution to the others. Section A of the Supplementary Material document provides further interpretation and properties of weighted pooled estimators, including the connection between asymptotic variance-optimal weights and pseudo-maximum likelihood estimation, a connection between AMSE-optimal weights and the regularization of bias-optimal weights, sensitivity to uncertainty in weight estimation, and gains in asymptotic variance compared to naive pooling (*i.e.* when $\omega_j = 1/m$ for $1 \le j \le m$).

2.2. Optimal choices of weights

Since the optimal values of weights in Theorem 1 depend on the asymptotic bias and variance components, these should be estimated first. It is then convenient to assume that $\rho_j < 0$ and $A_j(t) = \gamma\beta_j t^{\rho_j}$ for some constant β_j. Under this mild assumption (see Table 1 in [12] for a list of examples), consistent estimators $\widehat{\beta}_j$ and $\widehat{\rho}_j$ of β_j and ρ_j are available and implemented, for example, in the R function `mop` from the package `evt0` (see Section C.1 of the Supplementary Material document for further details). This yields an estimator of $\lambda_j = \lim_{n\to\infty}\sqrt{k_j}A_j(n_j/k_j)$ as $\widehat{\lambda}_j = \sqrt{k_j}\times\widehat{\gamma}_n(\boldsymbol{\omega})\widehat{\beta}_j(n_j/k_j)^{\widehat{\rho}_j}$. Here the choice of $\boldsymbol{\omega}$ is arbitrary; without any prior knowledge of the asymptotic dependence structure between the X_j, the use of naive weights $\boldsymbol{\omega} = (1/m,\ldots,1/m)^\top$ seems sensible.

To estimate the covariance matrix, let $n_{j,\ell} = \min(n_j, n_\ell)$ and $k_{j,\ell} = k_j$ if $n_j < n_\ell$ and k_ℓ otherwise, and consider the estimator of the tail copula function $R_{j,\ell}$ defined as

$$\widehat{R}_{j,\ell}(u,v) = \widehat{R}_{j,\ell}(u,v;k_{j,\ell}) = \frac{1}{k_{j,\ell}}\sum_{i=1}^{n_{j,\ell}}\mathbb{1}\left\{\frac{n_{j,\ell}+1-r_{n_{j,\ell},i,j}}{k_{j,\ell}(n_{j,\ell}+1)/n_{j,\ell}} \le u,\ \frac{n_{j,\ell}+1-r_{n_{j,\ell},i,\ell}}{k_{j,\ell}(n_{j,\ell}+1)/n_{j,\ell}} \le v\right\}.$$

[Here $r_{n_{j,\ell},i,j}$ (resp. $r_{n_{j,\ell},i,\ell}$) stands for the rank of $X_{i,j}$ (resp. $X_{i,\ell}$) among the observations $X_{1,j}, X_{2,j},\ldots,X_{n_{j,\ell},j}$ (resp. $X_{1,\ell}, X_{2,\ell},\ldots,X_{n_{j,\ell},\ell}$), namely, the first $n_{j,\ell}$ observations in sample j (resp. ℓ).] Other options include estimating $R_{j,\ell}$ by fitting extreme value copulae and using a relationship between tail copulae and Pickands dependence functions, see Section 3.1 of [14]. We shall show in Section 5.1 that our proposed approach, which has the advantage of simplicity, performs well in a wide range of models. Adapting Lemma 7 from [19] shows that the estimator $\widehat{R}_{j,\ell}$ above is a locally uniformly consistent estimator of $R_{j,\ell}$ on $(0,\infty)^2$ under our conditions. This leads to the estimators

$$\widehat{\boldsymbol{B}}_{\boldsymbol{c}} = \sqrt{k}\left(\frac{\widehat{\lambda}_1/\sqrt{k_1}}{1-\widehat{\rho}_1}, \frac{\widehat{\lambda}_2/\sqrt{k_2}}{1-\widehat{\rho}_2},\ldots,\frac{\widehat{\lambda}_m/\sqrt{k_m}}{1-\widehat{\rho}_m}\right)^\top$$

$$\text{and } [\widehat{\mathbf{V}}_{\boldsymbol{c}}]_{j,\ell} = k\,\widehat{\gamma}_n^2(\boldsymbol{\omega})\begin{cases}\dfrac{1}{k_j} & \text{if } j=\ell,\\[2ex] \dfrac{1}{k_j}\times\dfrac{\widehat{R}_{j,\ell}(k_j/k_\ell, n_j/n_\ell)}{\max(1,n_j/n_\ell)} & \text{if } j\neq\ell.\end{cases}$$

Plugging in these estimators in place of $\boldsymbol{B}_{\boldsymbol{c}}$ and $\mathbf{V}_{\boldsymbol{c}}$ in the expressions of $\boldsymbol{\omega}^{(\mathrm{Var})}$ and $\boldsymbol{\omega}^{(\mathrm{AMSE})}$ results in estimators $\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})}$ and $\widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})}$ of these optimal weights. Next, we provide the asymptotic properties of the composite pooled estimators based on these estimated weights.

Corollary 1. *Work under the conditions of Theorem 1 with $\boldsymbol{\gamma} = \gamma\mathbf{1}$, $\rho_j < 0$, and $A_j(t) = \gamma\beta_j t^{\rho_j}$ for all j. Assume that the matrix $\mathbf{V}$ is positive definite (hence $\mathbf{V}_{\boldsymbol{c}}$ is). Assume further that, for all $j \in \{1,\ldots,m\}$, $\widehat{\beta}_j$ is a consistent estimator of β_j and $(\widehat{\rho}_j - \rho_j)\log n_j = \mathrm{o}_{\mathbb{P}}(1)$. Then*

$$\sqrt{k}(\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})}) - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\mathbf{1}^\top\mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}}}{\mathbf{1}^\top\mathbf{V}_{\boldsymbol{c}}^{-1}\mathbf{1}}, \frac{1}{\mathbf{1}^\top\mathbf{V}_{\boldsymbol{c}}^{-1}\mathbf{1}}\right), \text{ and}$$

$$\sqrt{k}(\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})}) - \gamma) \xrightarrow{d} \mathcal{N}\left((\boldsymbol{\omega}^{(\mathrm{AMSE})})^\top\boldsymbol{B}_{\boldsymbol{c}}, (\boldsymbol{\omega}^{(\mathrm{AMSE})})^\top\mathbf{V}_{\boldsymbol{c}}\boldsymbol{\omega}^{(\mathrm{AMSE})}\right), \text{ where}$$

$$((\boldsymbol{\omega}^{(\mathrm{AMSE})})^\top\boldsymbol{B}_{\boldsymbol{c}})^2 + (\boldsymbol{\omega}^{(\mathrm{AMSE})})^\top\mathbf{V}_{\boldsymbol{c}}\boldsymbol{\omega}^{(\mathrm{AMSE})} = \frac{1+\boldsymbol{B}_{\boldsymbol{c}}^\top\mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}}}{(1+\boldsymbol{B}_{\boldsymbol{c}}^\top\mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top\mathbf{V}_{\boldsymbol{c}}^{-1}\mathbf{1}) - (\mathbf{1}^\top\mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}})^2}.$$

Remark 1 (On the variance-optimal choice). *The variance-optimal weights* $\boldsymbol{\omega}^{(\mathrm{Var})}$ *were suggested by [14], but their analog of Theorem 1 requires stronger regularity conditions about the* F_j *and an assumption akin to a second-order condition on the survival copulae* $\overline{C}_{j,\ell}$*. Besides, the use of their estimator for* $\boldsymbol{\omega}^{(\mathrm{Var})}$ *lacks a theoretical justification similar to Corollary 1. A variance-optimal convex combination is advocated in [6] under the same conditions as in [14]. In fact, at least when* $m \le 3$ *and* $c_2 = c_3 = 1$*, the variance-optimal set of weights we propose is a convex combination, and thus provides a simple closed form to the constrained variance minimization problem in [6]. For* $m = 2$*, this is more generally true for arbitrary* $c_2 > 0$*, and follows from a direct calculation; when* $c_2 = 1$*, one finds* $\boldsymbol{\omega}^{(\mathrm{Var})} = (1/2, 1/2)^\top$ *irrespective of* $\mathbf{V_1} = \mathbf{V}$*, corresponding to the naive average. For* $m = 3$*, the discussion is more complex and involves the identification of tail correlation matrices with a convex polytope that is a proper subset of the elliptope, a Riemannian quotient manifold representing the set of standard correlation matrices (see Section C.2 of the Supplementary Material document for more details). As a consequence, pooling together* $m \le 3$ *Hill estimators with equal effective sample fractions* k_j/n_j *can never outperform, in terms of asymptotic variance, a Hill estimator built from the top* $k = \sum_{j=1}^m k_j$ *observations coming from a pooled sample of independent data of size* $n = \sum_{j=1}^m n_j$*. Perhaps surprisingly, this conclusion is not reached in general pooling problems, e.g., for positively correlated sample means, because the variance-optimal pooled estimator may then not be a convex combination. What can happen in the case* $m > 3$ *remains an open question.*

Remark 2 (On the AMSE-optimal choice). *To the best of our knowledge, the AMSE-optimal weights* $\boldsymbol{\omega}^{(\mathrm{AMSE})}$ *have not been considered before in the literature. They should be favored in practice in the case of highly different sample fractions* k_j/n_j*, as demonstrated below in Section 3.2 in the distributed inference framework. Their construction is made possible in our setting because Theorem 1 gives a precise quantification of the asymptotic bias of the pooled tail index estimator, unlike the results of [6] and [14] which assume that this bias is negligible.*

Remark 3 (On optimal choices of the k_j **in pooled estimators).** *Contrary to the AMSE-optimal choice of weights, it seems hard to find an AMSE-optimal choice of the* k_j *for the pooled tail index estimator. We give an explanation here in the simple situation when* $m = 2$ *and* $R_{1,2} = 0$ *(i.e.* X_1 *and* X_2 *are asymptotically independent). The AMSE of each individual estimator* $\widehat{\gamma}_j(k_j)$ *is* $AMSE_j(k_j) = \gamma^2(b_j^2(n_j/k_j)^{2\rho_j}/(1-\rho_j)^2 + 1/k_j)$*, minimal at* $k_j^\star = (-(1-\rho_j)^2 n_j^{2\rho_j}/(2\rho_j b_j^2))^{1/(1-2\rho_j)}$*. Meanwhile, the AMSE of the pooled estimator with weights* $(\omega, 1-\omega)$ *is*

$$AMSE(\omega, k_1, k_2) = \omega^2 AMSE_1(k_1) + (1-\omega)^2 AMSE_2(k_2) + 2\omega(1-\omega)\frac{\gamma^2 b_1 b_2}{(1-\rho_1)(1-\rho_2)}\left(\frac{n_1}{k_1}\right)^{\rho_1}\left(\frac{n_2}{k_2}\right)^{\rho_2}$$

due to the asymptotic independence assumption. Then clearly, in the nontrivial case where $\omega \notin \{0,1\}$*,*

$$\frac{\partial AMSE}{\partial k_1}(\omega, k_1^\star, k_2^\star) \neq 0 \text{ and } \frac{\partial AMSE}{\partial k_2}(\omega, k_1^\star, k_2^\star) \neq 0.$$

This means that the optimal choices of k_1 *and* k_2 *in each individual sample will not constitute a pair of optimal choices for the pooled estimators. A simple closed form for the latter seems very difficult (if not impossible) to obtain in general, because canceling the partial derivatives of* $AMSE(\omega, k_1, k_2)$ *involves finding roots of pairs of polynomials in* (k_1, k_2) *which are not merely linear or quadratic functions.*

We conclude this section by discussing bias-reduced versions of the variance-optimal and AMSE-optimal pooled estimators, defined as $\overline{\gamma}_n(\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})}) = \widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})}) - (\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})})^\top \widehat{\boldsymbol{B}}_{\boldsymbol{c}}/\sqrt{k}$ and $\overline{\gamma}_n(\widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})}) = \widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})}) - (\widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})})^\top \widehat{\boldsymbol{B}}_{\boldsymbol{c}}/\sqrt{k}$.

Corollary 2. *Under the conditions of Corollary 1,* $\sqrt{k}(\overline{\gamma}_n(\widehat{\omega}_n^{(\mathrm{Var})}) - \gamma) \xrightarrow{d} \mathcal{N}(0, 1/(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1}\mathbf{1}))$ *and*

$$\sqrt{k}(\overline{\gamma}_n(\widehat{\omega}_n^{(\mathrm{AMSE})}) - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \frac{(1+\boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}})^2(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1}\mathbf{1}) - (2+\boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}})^2}{[(1+\boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1}\mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1}\boldsymbol{B}_{\boldsymbol{c}})^2]^2}\right).$$

If the marginal distributions are equal across samples, then one can improve the estimation of the weights by also pooling the second-order parameter estimators, see Section 3.2 below.

2.3. Weighted geometric pooling of extreme quantile estimators

Consider a very small exceedance probability $p = p(n) \to 0$ as $n \to \infty$. In each sample, the quantile $q_j(1-p)$ at level $1-p$ can be estimated by the extrapolated Weissman estimator of [22]:

$$\widehat{q}_j^\star(1-p|k_j) = \left(\frac{k_j}{n_j p}\right)^{\widehat{\gamma}_j(k_j)} X_{n_j-k_j:n_j,j},$$

where k_j satisfies $k_j/(n_j p) \to \infty$. The typical case of interest is when np is bounded, in which case only a few or no observations exceeding $q_j(1-p)$ are available. When the samples are believed or known to have the same tail index γ, it is natural to substitute the weighted estimator $\widehat{\gamma}_n(\boldsymbol{\omega})$ in place of the individual estimator $\widehat{\gamma}_j(k_j)$, to get $\widehat{q}_j^\star(1-p|k_j,\boldsymbol{\omega}) = (k_j/(n_j p))^{\widehat{\gamma}_n(\boldsymbol{\omega})} X_{n_j-k_j:n_j,j}$. Going one step further, the marginal quantile estimators themselves can be pooled when the datasets have the same extreme quantiles, or equivalently, if one assumes that

($\mathcal{H}$) For any $j, \ell \in \{1,\ldots,m\}$ with $j \neq \ell$, we have $U_j(t)/U_\ell(t) \to 1$ as $t \to \infty$.

Under assumption ($\mathcal{H}$), it is tempting to take again a weighted sum of the $\widehat{q}_j^\star(1-p|k_j)$, as $\overline{q}_n^\star(1-p|\boldsymbol{\omega}) = \sum_{j=1}^m \omega_j \widehat{q}_j^\star(1-p|k_j)$. This would be sensible if one were estimating central quantiles, see [15], but it is no longer the best solution when it comes to pooling the Weissman estimators, because the use of geometric weighted sums better suits their multiplicative and power structure (see Section 5.1.1, and Section D.1 of the Supplementary Material document, for numerical evidence). The crucial point to note here is that the log-Weissman quantile estimator can be rewritten as

$$\log \widehat{q}_j^\star(1-p|k_j) = \log\left(\frac{k}{np}\right)\widehat{\gamma}_j(k_j) + \left[\log\left(\frac{k_j}{k}\right) - \log\left(\frac{n_j}{n}\right)\right]\widehat{\gamma}_j(k_j) + \log X_{n_j-k_j:n_j,j}.$$

In the first term, the γ estimator appears on the standard scale. This suggests the use of the estimator defined as $\log \widehat{q}_n^\star(1-p|\boldsymbol{\omega}) = \sum_{j=1}^m \omega_j \log \widehat{q}_j^\star(1-p|k_j)$. This estimator is a weighted geometric mean of the $\widehat{q}_j^\star(1-p|k_j)$. We now derive the asymptotic normality of $\widehat{q}_j^\star(1-p|k_j,\boldsymbol{\omega})$ and $\widehat{q}_n^\star(1-p|\boldsymbol{\omega})$.

Theorem 2. *Work under the conditions and with the notation of Theorem 1 with* $\boldsymbol{\gamma} = \gamma\mathbf{1}$ *and* $\rho_j < 0$ *for all* $j \in \{1,\ldots,m\}$. *Pick* $p = p(n) \to 0$ *such that* $k/(np) \to \infty$ *and* $\sqrt{k}/\log(k/(np)) \to \infty$. *Let* $\boldsymbol{\omega}, \widehat{\boldsymbol{\omega}}_n$ *be such that* $\widehat{\boldsymbol{\omega}}_n^\top \mathbf{1} = 1$ *and* $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$. *Then, for any* j,

$$\frac{\sqrt{k}}{\log(k/(np))}\left(\frac{\widehat{q}_j^\star(1-p|k_j,\widehat{\boldsymbol{\omega}}_n)}{q_j(1-p)} - 1\right) = \sqrt{k}(\widehat{\gamma}_n(\boldsymbol{\omega}) - \gamma) + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\omega}^\top \boldsymbol{B}_{\boldsymbol{c}}, \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}}\boldsymbol{\omega}\right).$$

If moreover assumption ($\mathcal{H}$) *holds then, for any* j,

$$\frac{\sqrt{k}}{\log(k/(np))}\left(\frac{\widehat{q}_n^\star(1-p|\widehat{\boldsymbol{\omega}}_n)}{q_j(1-p)} - 1\right) = \sqrt{k}(\widehat{\gamma}_n(\boldsymbol{\omega}) - \gamma) + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\omega}^\top \boldsymbol{B}_{\boldsymbol{c}}, \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}}\boldsymbol{\omega}\right).$$

An analog of Corollary 1 is feasible for optimally-pooled extreme quantile estimation where $\widehat{\boldsymbol{\omega}}_n \in \{\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})}, \widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})}\}$, since the asymptotic distribution of $\widehat{q}_n^{\star}(1-p|\widehat{\boldsymbol{\omega}}_n)$ is governed by that of $\widehat{\gamma}_n(\boldsymbol{\omega})$. Similar results based on the bias-reduced versions $\overline{\gamma}_n(\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})})$ and $\overline{\gamma}_n(\widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})})$ are omitted.

2.4. Inference using pooled extreme value estimators

Equality of tail indices or extreme quantiles is of course a modelling assumption. We briefly present here an approach to testing this assumption, motivated by testing for nested models. Suppose that $\boldsymbol{Z}$ is an $m-$dimensional Gaussian random vector with mean $\boldsymbol{\mu}$ and known positive definite covariance matrix $\boldsymbol{V}$, and consider the testing problem of $M_0 : \mu_1 = \cdots = \mu_m = \mu$ versus $M_1 : \exists(j,\ell)$ with $j \neq \ell$ such that $\mu_j \neq \mu_\ell$. The log-likelihood ratio deviance statistic for testing the validity of model M_0 based on $\mathbf{Z}$ is $\Lambda = (\mathbf{Z} - \widehat{\mu}\mathbf{1})^\top \mathbf{V}^{-1}(\mathbf{Z} - \widehat{\mu}\mathbf{1})$, with $\widehat{\mu} = (\mathbf{1}^\top\mathbf{V}^{-1}\mathbf{Z})/(\mathbf{1}^\top\mathbf{V}^{-1}\mathbf{1})$. In model M_0, the statistic Λ has a chi-square distribution with $m-1$ degrees of freedom. In our context, under the assumptions of Theorem 1 and if all the λ_j are 0 (see Remark 6 below for a discussion of this assumption), one has $\sqrt{k}(\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}) = (\sqrt{k}(\widehat{\gamma}_1(k_1)-\gamma_1), \ldots, \sqrt{k}(\widehat{\gamma}_m(k_m)-\gamma_m))^\top \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_{\boldsymbol{c}})$ with

$$[\mathbf{V}_{\boldsymbol{c}}]_{j,\ell} = \left(\sum_{i=1}^m c_i^{-1}\right)\begin{cases}\gamma_j^2 c_j & \text{if } j=\ell,\\ \gamma_j\gamma_\ell b_j c_j R_{j,\ell}(c_\ell/c_j, b_\ell/b_j)/\max(b_j,b_\ell) & \text{if } j<\ell.\end{cases}$$

In other words, the distribution of $\widehat{\boldsymbol{\gamma}}_n$ is approximately $\mathcal{N}(\boldsymbol{\gamma}, k^{-1}\mathbf{V}_{\boldsymbol{c}})$. Plugging in $\widehat{\gamma}_j(k_j)$ and $\widehat{R}_{j,l}(k_j/k_\ell, n_j/n_\ell)$ in place of γ_j and $R_{j,\ell}(c_\ell/c_j, b_\ell/b_j)$, respectively, one obtains an estimator $\overline{\mathbf{V}}_{\boldsymbol{c}}$ of $\mathbf{V}_{\boldsymbol{c}}$ and hence a deviance statistic for testing $H_0 : \boldsymbol{\gamma} = \gamma\mathbf{1}$ versus $H_1 : \boldsymbol{\gamma} \neq \gamma\mathbf{1}$ as

$$\Lambda_n = k(\widehat{\boldsymbol{\gamma}}_n - \overline{\mu}_n\mathbf{1})^\top \overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}(\widehat{\boldsymbol{\gamma}}_n - \overline{\mu}_n\mathbf{1}), \text{ with } \overline{\mu}_n = \frac{\mathbf{1}^\top\overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\widehat{\boldsymbol{\gamma}}_n}{\mathbf{1}^\top\overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\mathbf{1}} = (\overline{\boldsymbol{\omega}}_n^{(\mathrm{Var})})^\top\widehat{\boldsymbol{\gamma}}_n = \widehat{\gamma}_n(\overline{\boldsymbol{\omega}}_n^{(\mathrm{Var})}).$$

The (squared Mahalanobis distance) test statistic Λ_n compares the vector $\widehat{\boldsymbol{\gamma}}_n$ with an estimate of the variance-optimal pooled estimator on a scale adapted to the extremal pairwise dependence in the components of $\boldsymbol{X}$. A somewhat different proposal is outlined in [14]. We explain the main differences in Section C.3 of the Supplementary Material document. Our proposed testing procedure, of asymptotic significance level α, is to reject H_0 if $\Lambda_n > \chi^2_{m-1,1-\alpha}$, where $\chi^2_{m-1,1-\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution with $m-1$ degrees of freedom. Next, we establish the consistency of this test and give a symmetric asymptotic confidence interval for the common tail index γ under H_0.

Corollary 3. *Under the conditions of Theorem 1 and if $\lambda_j = 0$ for all j, we have $\mathbb{P}(\Lambda_n > \chi^2_{m-1,1-\alpha}) \to \alpha$ under H_0, and $\Lambda_n \xrightarrow{\mathbb{P}} +\infty$ under H_1. Moreover, under H_0, if $\widehat{\boldsymbol{\omega}}_n^\top\mathbf{1} = 1$ with $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$, then*

$$\forall\alpha\in(0,1),\ \lim_{n\to\infty}\mathbb{P}\left(\gamma \in \left[\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n) \pm z_{1-\alpha/2}\sqrt{(\widehat{\boldsymbol{\omega}}_n^\top\widehat{\mathbf{V}}_{\boldsymbol{c}}\widehat{\boldsymbol{\omega}}_n)/k}\right]\right) = 1-\alpha$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the standard Gaussian distribution. [In this asymptotic confidence interval, $\widehat{\mathbf{V}}_{\boldsymbol{c}}$ is calculated as described in Section 2.2.]

Remark 4 (Identification of tail homogeneous subgroups). *When H_0 is rejected, it is of interest to identify tail homogeneous subgroups. A simple yet effective method, employed in the real data application in Section 5.2.2, is to list the estimated tail indices and manually (or automatically, using an* *unsupervised clustering algorithm such as k−means) construct subgroups in which these estimates are similar. An alternative procedure, only tractable when m is low but having the advantage of being statistically principled and exhaustive over all possible choices of subgroups, is to calculate the set of log-likelihoods of $\widehat{\boldsymbol{\gamma}}_n = (\widehat{\gamma}_1(k_1), \ldots, \widehat{\gamma}_m(k_m))$ under the multivariate Gaussian model and all possible equality constraints on $\boldsymbol{\gamma}$: for example, when $m = 3$, one can have pairwise different (i.e. unconstrained) γ_j, or exactly two equal γ_j (three possibilities), or equal γ_j. Associated to each of these submodels will be Akaike's Information Criterion, with the minimal AIC value suggesting the most likely subgroups.*

Remark 5 (With asymptotic independence across subsamples). *An important subcase is when pairs of data points taken from two different subsamples are asymptotically independent. In this case, all tail copulae $R_{j,\ell}$ are identically zero, so one can estimate $\mathbf{V}_{\boldsymbol{c}}$ with $k\,\mathrm{diag}(\widehat{\gamma}_1^2(k_1)/k_1, \ldots, \widehat{\gamma}_m^2(k_m)/k_m)$. The test statistic Λ_n becomes $\Lambda_n = \sum_{j=1}^m k_j(\widehat{\gamma}_j(k_j) - \widehat{\gamma}_n(\overline{\boldsymbol{\omega}}_n^{(\mathrm{Var})}))^2/\widehat{\gamma}_j^2(k_j)$. This has the familiar look of a Pearson goodness-of-fit statistic, with the weight k_j adjusting for the different rates of convergence of the $\widehat{\gamma}_j(k_j)$. If all the k_j are equal, then $\Lambda_n = \frac{k}{m}\sum_{j=1}^m(\widehat{\gamma}_n(\overline{\boldsymbol{\omega}}_n^{(\mathrm{Var})})/\widehat{\gamma}_j(k_j) - 1)^2$. Our proposed statistic Λ_n then bears some similarity with a test statistic of [10] for the validity of a multivariate regular variation model that assumes equality of tail indices across marginal distributions.*

Remark 6 (Inference and bias correction). *Typically, assuming $\lambda_j = 0$ to omit the asymptotic bias terms is sensible as long as the second-order parameters ρ_j remain reasonably far away from 0. Based on finite-sample experiments with a total sample size $n = 1{,}000$, marginal Burr distributions and $2 \le m \le 5$ with both balanced and unbalanced samples, the confidence interval provided in Corollary 3 seems to perform very well at least when $|\rho_j| > 3/4$. Estimating the bias terms in such a situation is in fact detrimental, because of increased variability of the resulting interval estimator that is not accounted for in the estimated variance of the Gaussian limiting distribution.*

Remark 7 (Tail homogeneity and tail homoskedasticity). *When all the parameters ρ_j are negative, as in Corollary 3, one has $t^{-\gamma_j}U_j(t) \to C_j \in (0,\infty)$ as $t \to \infty$, see the equation below Equation (2.3.23) in [5]. Testing $H_0 : \boldsymbol{\gamma} = \gamma\mathbf{1}$ versus $H_1 : \boldsymbol{\gamma} \neq \gamma\mathbf{1}$ is then equivalent to testing*

$$
\begin{aligned}
& H_0' : \forall j, \ell \in \{1, \ldots, m\},\ \lim_{\tau\uparrow 1} q_j(\tau)/q_\ell(\tau) \in (0,\infty), \\
\textit{versus}\ \ & H_1' : \exists j, \ell \in \{1, \ldots, m\} \textit{ with } j \neq \ell \textit{ and } \lim_{\tau\uparrow 1} q_j(\tau)/q_\ell(\tau) \in \{0,\infty\}.
\end{aligned}
$$

The testing procedure based on the statistic Λ_n is therefore, under very mild conditions, exactly a test for asymptotic proportionality of marginal extreme quantiles. It can thus be used to detect tail homogeneity (equal tail indices and therefore asymptotically proportional tail quantiles) as opposed to tail heterogeneity (one marginal distribution having a heavier tail than the others). We discuss below the testing of the stronger property when all limits are equal to 1 in H_0', corresponding to the asymptotic equivalence of extreme quantiles ($\mathcal{H}$), and referred to as tail homoskedasticity.

Testing for tail homoskedasticity can be done directly using the individual Weissman estimators $\widehat{q}_j^{\star}(1-p|k_j)$. Set $\mathbf{Z}_n(p) = \log \widehat{\boldsymbol{q}}_n^{\star}(1-p) = (\log \widehat{q}_1^{\star}(1-p|k_1), \ldots, \log \widehat{q}_m^{\star}(1-p|k_m))$ and

$$
L_n(p) = \frac{k}{\log^2(k/(np))}\left(\mathbf{Z}_n(p) - \frac{\mathbf{1}^\top \overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\mathbf{Z}_n(p)}{\mathbf{1}^\top \overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\mathbf{1}}\mathbf{1}\right)^\top \overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\left(\mathbf{Z}_n(p) - \frac{\mathbf{1}^\top \overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\mathbf{Z}_n(p)}{\mathbf{1}^\top \overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\mathbf{1}}\mathbf{1}\right).
$$

A testing procedure of asymptotic significance level α of ($\mathcal{H}$) versus ($\mathcal{H}'$) : $\exists j, \ell \in \{1, \ldots, m\}$ with $j \neq \ell$ and $\lim_{\tau\uparrow 1} q_j(\tau)/q_\ell(\tau) \neq 1$ is to reject ($\mathcal{H}$) if $L_n(p) > \chi^2_{m-1,1-\alpha}$. This is established below.

Corollary 4. *Under the conditions of Corollary 3 and $\rho_j < 0$ for all j, if $p = p(n) \to 0$ is such that $k/(np) \to \infty$ and $\sqrt{k}/\log(k/(np)) \to \infty$, then we have $\mathbb{P}(L_n(p) > \chi^2_{m-1,1-\alpha}) \to \alpha$ under $(\mathcal{H})$, and $L_n(p) \xrightarrow{\mathbb{P}} +\infty$ under $(\mathcal{H}')$. Moreover, under $(\mathcal{H})$, if $\widehat{\boldsymbol{\omega}}_n^\top \mathbf{1} = 1$ with $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$ then, for all j,*

$$\forall \alpha \in (0,1), \ \lim_{n\to\infty} \mathbb{P}\left(q_j(1-p) \in \left[\widehat{q}_n^{\star}(1-p|\widehat{\boldsymbol{\omega}}_n) \exp\left(\pm z_{1-\alpha/2} \log\left[\frac{k}{np}\right] \sqrt{\frac{\widehat{\boldsymbol{\omega}}_n^\top \widehat{\mathbf{V}}_{\boldsymbol{c}} \widehat{\boldsymbol{\omega}}_n}{k}} \right) \right] \right) = 1-\alpha.$$

Unlike the tail homoskedasticity test suggested in Section 3.3 of [16], the present test does not require integrability assumptions on the X_j. The use of the log-scale is equivalent in theory to the relative scale employed in Theorem 2, but it tends to provide more accurate asymptotic confidence intervals for extreme quantiles, as indicated for instance by [8]. Of course, one may also consider inference procedures based on AMSE-optimal weights, but as we shall show in our finite-sample experiments in Section 5.1, these would in general not perform better than those based on variance-optimal weights.

3. The framework of distributed inference

In the framework of distributed inference, very restricted or no communication is allowed between the m machines in which individual subsamples are stored. In particular, the end user operating from a central machine only has access to limited information, such as the subsample estimates and associated n_j and k_j, which is not sufficient for estimating pairwise tail dependence structures. We thus assume that the data points within and across machines are independent, that is, the $X_{i,j}$ are i.i.d. for $1 \le i \le n_j$ and $1 \le j \le m$, with a distribution satisfying the second-order condition $C_2(\gamma, \rho, A)$.

3.1. Distributed estimation of the tail index

With i.i.d. data, a benchmark for the distributed estimator $\widehat{\gamma}_n(\boldsymbol{\omega})$ is the Hill estimator based on the unfeasible combination of subsamples $\{X_i, 1 \le i \le n\} = \{X_{i,j}, 1 \le j \le m, 1 \le i \le n_j\}$ with effective sample size $k = \sum_{j=1}^m k_j$, that is, $\widehat{\gamma}_n^{(\mathrm{Hill})}(k) = k^{-1}\sum_{i=1}^k \log(X_{n-i+1:n}/X_{n-k:n})$ where $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ are the order statistics of $X_1, \ldots, X_n$. Assume that the n_j and the k_j are asymptotically proportional but possibly unbalanced, *i.e.* $n_1/n_j \to b_j \in (0,\infty)$ and $k_1/k_j \to c_j \in (0,\infty)$. Since A is regularly varying with index ρ, the existence of $\lambda_j = \lim_{n\to\infty} \sqrt{k_j} A(n_j/k_j)$ is equivalent to that of $\lambda = \lim_{n\to\infty} \sqrt{k} A(n/k)$, and then $\lambda_j = c_j^{\rho-1/2} b_j^{-\rho} \lambda_1 = c_j^{\rho-1/2} b_j^{-\rho} (\sum_{j=1}^m c_j^{-1})^{\rho-1/2} (\sum_{j=1}^m b_j^{-1})^{-\rho} \lambda$. Hence, we have the following corollary of Theorem 1.

Corollary 5. *Assume that condition $C_2(\gamma,\rho,A)$ holds. Suppose that $n_1/n_j \to b_j \in (0,\infty)$ and $k_1/k_j \to c_j \in (0,\infty)$ (with then $b_1 = c_1 = 1$) for any $j \in \{1,\ldots,m\}$, and then that $k \to \infty$ with $k/n \to 0$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^\top$ be such that $\boldsymbol{\omega}^\top \mathbf{1} = 1$. Then*

$$\sqrt{k}(\widehat{\gamma}_n(\boldsymbol{\omega}) - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho} \sum_{j=1}^m d_j^{\rho} \omega_j, \gamma^2 \sum_{j=1}^m \frac{1}{c_j} \sum_{j=1}^m c_j \omega_j^2 \right)$$

where $d_j = (c_j \sum_{i=1}^m 1/c_i)/(b_j \sum_{i=1}^m 1/b_i)$. If $\widehat{\boldsymbol{\omega}}_n^\top \mathbf{1} = 1$ and $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$ then $\sqrt{k}(\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n) - \widehat{\gamma}_n(\boldsymbol{\omega})) = \mathrm{o}_{\mathbb{P}}(1)$ and so the above convergence remains valid for $\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n)$.

We compare next the variance- and AMSE-optimal distributed estimators to the benchmark Hill estimator, which satisfies $\sqrt{k}(\widehat{\gamma}_n^{(\mathrm{Hill})}(k)-\gamma) \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho),\gamma^2)$ (see Theorem 3.2.5, p.74 in [5]).

3.2. Variance-optimal and AMSE-optimal combinations

The variance-optimal weights are $\omega_j^{(\mathrm{Var})} = (\sum_{i=1}^m c_i^{-1})^{-1} c_j^{-1}$ for all j. Estimating $1/c_i$ by k_i/k_1 leads to the estimated version $\widetilde{\boldsymbol{\omega}}_n^{(\mathrm{Var})} = (k_1/k,\ldots,k_m/k)$. The variance-optimal distributed estimator therefore has a much simpler expression than in the general setup of Section 2, and its computation only requires reporting the k_j and $\widehat{\gamma}_j(k_j)$ to the central machine. Being a nonrandom convex combination, this estimator is immune to the instability issues possibly caused by estimation of weights (see Remark A.3 in the Supplementary Material document). The following result gives its asymptotic distribution.

Corollary 6. *Under the conditions and with the notation of Corollary 5,*

$$\sqrt{k}(\widehat{\gamma}_n(\widetilde{\boldsymbol{\omega}}_n^{(\mathrm{Var})})-\gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}\left(\sum_{j=1}^m \frac{1}{c_j}\right)^{-1}\left(\sum_{j=1}^m \frac{d_j^\rho}{c_j}\right),\gamma^2\right).$$

It follows immediately that the proportion of variance gained by using the variance-optimal estimator instead of the naive estimator with constant weights $\omega_j = 1/m$ (for $1\le j\le m$) can be quantified as

$$\frac{\text{variance(naive)} - \text{optimal variance}}{\text{variance(naive)}} = \frac{\frac{1}{m}\sum_{j=1}^m \frac{1}{c_j}\times\frac{1}{m}\sum_{j=1}^m c_j - 1}{\frac{1}{m}\sum_{j=1}^m \frac{1}{c_j}\times\frac{1}{m}\sum_{j=1}^m c_j} = 1-\left(\frac{1}{m}\sum_{j=1}^m\frac{1}{c_j}\right)^{-1}\left(\frac{1}{m}\sum_{j=1}^m c_j\right)^{-1}.$$

This proportion is close to 1 when the vector $(c_1,c_2,\ldots,c_m)=(1,c_2,\ldots,c_m)$ is far from $(1,\ldots,1)$, *i.e.* when there is a strong degree of unbalance between effective sample sizes.

Remark 8 (Asymptotic bias comparison). *The unfeasible Hill estimator has asymptotic bias $\mu^{(\mathrm{Hill})} = \lambda/(1-\rho)$. If $\lambda\neq 0$ and $\mu^{(\mathrm{Var})}$ denotes the asymptotic bias of $\widehat{\gamma}_n(\widetilde{\boldsymbol{\omega}}_n^{(\mathrm{Var})})$, then Corollary 5 leads to $\mu^{(\mathrm{Hill})}=\mu^{(\mathrm{Var})}$ when $\rho=0$: the intuition is that, since $|A|$ is slowly varying when $\rho=0$, the asymptotic bias of $\widehat{\gamma}_n(\boldsymbol{\omega})$ is $\sum_{j=1}^m \omega_j A(n_j/k_j)\approx(\sum_{j=1}^m\omega_j)A(n/k)=A(n/k)$, which is nothing but the asymptotic bias of the Hill estimator, regardless of the choice of weights ω_j. Otherwise, the Hölder inequality for the conjugate exponents $p=-(1-\rho)/\rho$ and $q=1-\rho$ provides $\mu^{(\mathrm{Hill})}/\mu^{(\mathrm{Var})}\le 1$. Equality holds if and only if $b_j/c_j=K$, a constant independent of j (see Section C.4 of the Supplementary Material document for technical details), i.e. $b_j/c_j=b_1/c_1=1$. In other words, $|\mu^{(\mathrm{Hill})}|\le|\mu^{(\mathrm{Var})}|$ with equality $\mu^{(\mathrm{Hill})}=\mu^{(\mathrm{Var})}$ if and only if either $\rho=0$ or $k_1/n_1=(k_j/n_j)(1+\mathrm{o}(1))$ for any j, meaning that the sample fraction in each machine should be (asymptotically) the same for asymptotic bias equality to hold. In this situation, $d_j=1$ for any j and so, by Corollary 5, the variance- and AMSE-optimal weights are actually identical.*

Corollary 6 and Remark 8 motivate the following result in the case of equal sample fractions. The proof of the second statement therein requires specific theoretical arguments.

Theorem 3. *Under the conditions of Corollary 6 with $\lambda\neq 0$, $\sqrt{k}(\widehat{\gamma}_n(\widetilde{\boldsymbol{\omega}}_n^{(\mathrm{Var})})-\gamma)\xrightarrow{d}\mathcal{N}(\lambda/(1-\rho),\gamma^2)$ if and only if $\rho=0$ or $k_j/n_j=(k/n)(1+\mathrm{o}(1))$ for any $j\in\{1,\ldots,m\}$. If moreover $k_j/n_j=(k/n)(1+\mathrm{O}(1/\sqrt{k}))$ for any j, we have in fact $\sqrt{k}(\widehat{\gamma}_n(\widetilde{\boldsymbol{\omega}}_n^{(\mathrm{Var})})-\widehat{\gamma}_n^{(\mathrm{Hill})}(k))=\mathrm{o}_{\mathbb{P}}(1)$.*

Adjusting the effective sample sizes k_j such that k_j/n_j is constant in j thus produces a variance- and AMSE-optimal estimator that is asymptotically equivalent to the unfeasible Hill estimator built from the combined subsamples. This property is much stronger than only sharing the same asymptotic distribution. It is, however, unlikely that such an adjustment can be performed in general, since each machine will typically pick its own k_j following a selection rule based only on its subsample [6,14].

This motivates the focus on AMSE-optimal pooling in the case of unequal k_j/n_j. In this setting, the estimator $\widehat{\gamma}_j(k_j)$ with the lowest sample fraction will be the one carrying the smallest amount of bias, which may be substantially lower than the bias of the Hill estimator calculated on the full sample as the latter will typically use a larger sample fraction. As a consequence, giving the former estimator a large weight can improve the AMSE overall. The following result makes this intuition rigorous by comparing the AMSE-optimal distributed estimator with the benchmark Hill estimator.

Theorem 4. *Under the conditions and notation of Corollary 5, set*

$$AMSE^{(\mathrm{Hill})} = \frac{1}{k}\left(\frac{\lambda^2}{(1-\rho)^2} + \gamma^2\right) \text{ and } AMSE(\boldsymbol{\omega}) = \frac{1}{k}((\boldsymbol{\omega}^\top \boldsymbol{B_c})^2 + \boldsymbol{\omega}^\top \mathbf{V_c}\boldsymbol{\omega})$$

$$\text{with } \boldsymbol{B_c} = \frac{\lambda}{1-\rho}\left(d_1^\rho, \dots, d_m^\rho\right)^\top \text{ and } \mathbf{V_c} = \gamma^2\left(\sum_{i=1}^m c_i^{-1}\right)\mathrm{diag}(c_1, \dots, c_m).$$

Assume that the b_j/c_j are not all equal to 1. Then $AMSE(\boldsymbol{\omega}^{(\mathrm{AMSE})}) \geq AMSE^{(\mathrm{Hill})}$ if and only if $|\lambda| \leq \lambda_0$, with

$$\lambda_0 = \gamma(1-\rho)\sqrt{\frac{S_\rho^2 - S_0^2}{S_0 S_{2\rho} - S_\rho^2}}, \text{ and } S_\alpha = \sum_{j=1}^m \frac{d_j^\alpha}{c_j}.$$

It should be noted that Theorem 4 does not violate the minimax optimality property of the (benchmark) Hill estimator proved in [7] since it only states that the AMSE-optimal pooled estimator has a lower AMSE than the Hill estimator within a certain class of heavy-tailed distributions.

To estimate the AMSE-optimal weights in the current context, assume as in Section 2.2 that $A(t) = \gamma\beta t^\rho$ and note that $\lambda = \lim_{n\to\infty}\sqrt{k}A(n/k) = \gamma\beta\lim_{n\to\infty}\sqrt{k}(n/k)^\rho$ and $d_j = \lim_{n\to\infty}(n_j k)/(n k_j)$. The estimators $\widehat{\beta}_j$ and $\widehat{\rho}_j$, defined in Section 2.2, of $\beta \equiv \beta_j$ and $\rho \equiv \rho_j$ are restricted to each machine j. Here one may use instead the pooled versions $\widehat{\beta}_n(\boldsymbol{\omega}) = \sum_{j=1}^m \omega_j \widehat{\beta}_j$ and $\widehat{\rho}_n(\boldsymbol{\omega}) = \sum_{j=1}^m \omega_j \widehat{\rho}_j$, leading to $\widehat{\lambda} = \widehat{\lambda}(\widehat{\boldsymbol{\omega}}_\gamma, \widehat{\boldsymbol{\omega}}_\beta, \widehat{\boldsymbol{\omega}}_\rho) = \widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_\gamma)\widehat{\beta}_n(\widehat{\boldsymbol{\omega}}_\beta) \times \sqrt{k}(n/k)^{\widehat{\rho}_n(\widehat{\boldsymbol{\omega}}_\rho)}$, where $\widehat{\boldsymbol{\omega}}_\gamma$, $\widehat{\boldsymbol{\omega}}_\beta$ and $\widehat{\boldsymbol{\omega}}_\rho$ are three sets of weights. An obvious choice is $\omega_j = 1/m$ for all three estimators. A more refined choice is variance-optimal weights k_j/k for $\widehat{\boldsymbol{\omega}}_\gamma$ and n_j/n for both $\widehat{\boldsymbol{\omega}}_\beta$ and $\widehat{\boldsymbol{\omega}}_\rho$ (recall that $\widehat{\beta}_j$ and $\widehat{\rho}_j$ use almost all the available observations in each machine; see Section C.1 of the Supplementary Material document). Set

$$\widetilde{\boldsymbol{B}}_{\boldsymbol{c}} = \sqrt{\sum_{j=1}^m k_j}\,\frac{\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_\gamma)\widehat{\beta}_n(\widehat{\boldsymbol{\omega}}_\beta)}{1-\widehat{\rho}_n(\widehat{\boldsymbol{\omega}}_\rho)}\left(\left(\frac{n_1}{k_1}\right)^{\widehat{\rho}_n(\widehat{\boldsymbol{\omega}}_\rho)}, \dots, \left(\frac{n_m}{k_m}\right)^{\widehat{\rho}_n(\widehat{\boldsymbol{\omega}}_\rho)}\right)^\top$$

$$\text{and } \widetilde{\mathbf{V}}_{\boldsymbol{c}} = \left(\sum_{j=1}^m k_j\right)\widehat{\gamma}_n^2(\widehat{\boldsymbol{\omega}}_\gamma)\,\mathrm{diag}(1/k_1, \dots, 1/k_m),$$

and define $\widetilde{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})}$ by replacing $\boldsymbol{B_c}$ and $\mathbf{V_c}$ in $\boldsymbol{\omega}^{(\mathrm{AMSE})}$ with $\widetilde{\boldsymbol{B}}_{\boldsymbol{c}}$ and $\widetilde{\mathbf{V}}_{\boldsymbol{c}}$. These estimators featuring pooled second-order parameter estimates require, of course, communicating $k_j, n_j, \widehat{\beta}_j, \widehat{\rho}_j$ to the central machine. We obtain the following result as an immediate consequence of Corollary 5.

Corollary 7. *Work under the conditions of Corollary 5 with $\rho < 0$ and $A(t) = \gamma\beta t^{\rho}$. Assume that, for all $j \in \{1,\ldots,m\}$, $\widehat{\beta}_j$ is a consistent estimator of β and $(\widehat{\rho}_j - \rho)\log n_j = \mathrm{o}_{\mathbb{P}}(1)$. Then*

$$\sqrt{k}(\widehat{\gamma}_n(\widetilde{\omega}_n^{(\mathrm{AMSE})}) - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}\sum_{j=1}^m d_j^{\rho}\omega_j^{(\mathrm{AMSE})}, \gamma^2 \sum_{j=1}^m \frac{1}{c_j}\sum_{j=1}^m c_j(\omega_j^{(\mathrm{AMSE})})^2\right).$$

Confidence intervals can then be constructed as in Section 2.4, and bias reduction and appropriate theory can also be developed as in Section 2.2. We omit the details for the sake of brevity.

3.3. Extreme quantile estimation

We are now ready to compare the weighted geometric distributed estimator of an extreme quantile $q(1-p)$, as well as its variance- and AMSE-optimal versions, to the classical unfeasible Weissman estimator obtained directly from the combined subsamples, each defined as

$$\widehat{q}_n^{\star}(1-p|\omega) = \prod_{j=1}^m \left[\left(\frac{k_j}{n_j p}\right)^{\widehat{\gamma}_j(k_j)} X_{n_j-k_j:n_j,j}\right]^{\omega_j}, \quad \widehat{q}_n^{\star,(\mathrm{Hill})}(1-p|k) = \left(\frac{k}{np}\right)^{\widehat{\gamma}_n^{(\mathrm{Hill})}(k)} X_{n-k:n}.$$

Corollary 8. *Work under the conditions of Corollary 5 with $\rho < 0$. Pick $p = p(n) \to 0$ such that $k/(np) \to \infty$ and $\sqrt{k}/\log(k/(np)) \to \infty$. Let $\boldsymbol{\omega}, \widehat{\boldsymbol{\omega}}_n$ be such that $\widehat{\boldsymbol{\omega}}_n^{\top}\mathbf{1} = 1$ and $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$. Then*

$$\frac{\sqrt{k}}{\log(k/(np))}\left(\frac{\widehat{q}_n^{\star}(1-p|\widehat{\boldsymbol{\omega}}_n)}{q(1-p)} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}\sum_{j=1}^m d_j^{\rho}\omega_j, \gamma^2\sum_{j=1}^m \frac{1}{c_j}\sum_{j=1}^m c_j\omega_j^2\right).$$

If also $k_j/n_j = (k/n)(1+\mathrm{O}(1/\sqrt{k}))$ for any $j \in \{1,\ldots,m\}$, then $\widehat{q}_n^{\star}(1-p|\widetilde{\boldsymbol{\omega}}_n^{(\mathrm{Var})})$ is $\sqrt{k}/\log(k/(np))$–asymptotically equivalent to $\widehat{q}_n^{\star,(\mathrm{Hill})}(1-p|k)$. Finally, if the b_j/c_j are not all equal to 1, then under the conditions of Corollary 7, $\widehat{q}_n^{\star}(1-p|\widetilde{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})})$ has a smaller AMSE than $\widehat{q}_n^{\star,(\mathrm{Hill})}(1-p|k)$ if and only if $|\lambda| > \lambda_0$, with the notation of Theorem 4.

3.4. Extension to the case of at least one, but not all, very low k_j

It may happen that the ratio $\max_{1\le j\le m} k_j/\min_{1\le j\le m} k_j$ is quite large, owing to uncertainty in data-driven selection rules. One may then want to discard the marginal estimates with a very low k_j from the pooling procedure, but these estimates will have very low bias if all machines have comparable sample sizes, and hence it is more sensible to incorporate them into the distributed estimators of the tail index and extreme quantiles. From an asymptotic point of view, we obtain the following result for the variance-optimal distributed estimators in this case of extremely unbalanced effective sample sizes.

Theorem 5. *Assume that condition $C_2(\gamma,\rho,A)$ holds. Suppose that there is $\ell \in \{1,\ldots,m-1\}$ such that on the one hand, for any $j \in \{1,\ldots,\ell\}$, $k_j \to \infty$ with $k_j/n_j \to 0$, $n_1/n_j \to b_j \in (0,\infty)$ and $k_1/k_j \to c_j \in (0,\infty)$, and $\sqrt{k_j}A(n_j/k_j) \to \lambda_j \in \mathbb{R}$; and on the other hand, for any $j \in \{\ell+1,\ldots,m\}$, $k_j = k_j(n)$ is a nondecreasing sequence with $k_j/n_j \to 0$, $k_1/k_j \to \infty$ and $\sqrt{k_j}A(n_j/k_j) = \mathrm{O}(1)$. Then,* *if* $\lambda = \lim_{n\to\infty} \sqrt{k} A(\sum_{i=1}^{\ell} n_i / \sum_{i=1}^{\ell} k_i) \in \mathbb{R}$ *and* $d_j = (c_j/b_j) \times (\sum_{i=1}^{\ell} c_i^{-1})/(\sum_{i=1}^{\ell} b_i^{-1})$,

$$\sqrt{k}(\widehat{\gamma}_n(\widetilde{\omega}_n^{(\mathrm{Var})}) - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}\left(\sum_{j=1}^{\ell}\frac{1}{c_j}\right)^{-1}\left(\sum_{j=1}^{\ell}\frac{d_j^{\rho}}{c_j}\right), \gamma^2\right).$$

If moreover $\rho < 0$ *and* $p = p(n) \to 0$ *is such that* $k/(np) \to \infty$ *and* $\sqrt{k}/\log(k/(np)) \to \infty$*, as well as*

$$\max_{1\le j\le m} \frac{\sqrt{k_j}}{\sqrt{k}}\left|\frac{\log(k_j/(n_j p))}{\log(k/(np))} - 1\right| \to 0, \tag{1}$$

then

$$\frac{\sqrt{k}}{\log(k/(np))}\left(\frac{\widehat{q}_n^{\star}(1-p|\widetilde{\omega}_n^{(\mathrm{Var})})}{q(1-p)} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}\left(\sum_{j=1}^{\ell}\frac{1}{c_j}\right)^{-1}\left(\sum_{j=1}^{\ell}\frac{d_j^{\rho}}{c_j}\right), \gamma^2\right).$$

The variance-optimal distributed estimators thus behave as if they were calculated without including the machines using a very low k_j. This provides theoretical justification for their use regardless of the lack of balance in effective sample sizes. Note that condition (1) always holds if for any $j \in \{\ell+1,\ldots,m\}$, $n_1/n_j \to b_j \in (0,\infty)$. Remark also that, if at least one of the k_j is bounded then, unlike the variance-optimal version, the naive distributed estimator is not even consistent.

3.5. The case of a large number of machines

Our results above do not consider the case when all the k_j are low, possibly even bounded. When all the $k_j = k_j(n)$ are bounded in n, consistency of the distributed estimators necessarily requires a growing number of machines, *i.e.* $m = m(n) \to \infty$, since otherwise the pooled estimators may contain only a bounded number of summands. In this context, we require the following fundamental assumption.

$(\mathcal{A})$ $m = m(n) \to \infty$ and the $n_j = n_j(n)$ satisfy $\inf_{1\le j\le m} n_j/\log m \to \infty$ as $n \to \infty$.

This condition means that the amount of data stored in each machine grows with n, although the number of machines may grow much faster than the n_j; in the balanced data setting, it reduces to $m = m(n) \to \infty$ and $m\log(m)/n \to 0$, which is for instance satisfied when $m = n^{\theta}$ for any $\theta \in (0,1)$. We require condition $(\mathcal{A})$ to establish a precise control of the statistical errors arising in each machine. The proof is thus fundamentally different from the proofs of our theorems in the case of a fixed m. Another important difference is that the weight vector $\boldsymbol{\omega} \in \mathbb{R}^m$ now implicitly varies with n, so restrictions are needed to define admissible weights. For example, the pooled estimator corresponding to the weight $(1,0,0,\ldots)$ is simply $\widehat{\gamma}_1(k_1)$, which is not consistent when k_1 is bounded. We thus introduce a *balanced allocation condition* on $\boldsymbol{\omega} = \boldsymbol{\omega}(n)$.

$(\mathcal{W})$ The weight vector $\boldsymbol{\omega} = \boldsymbol{\omega}(n) \in \mathbb{R}^m$ satisfies $\sum_{j=1}^m \omega_j = 1$ as well as

$$\lim_{n\to\infty}\sum_{j=1}^{m}\frac{\omega_j^2}{k_j} = 0 \ \text{ and } \ \exists\delta > 0,\ \lim_{n\to\infty}\frac{\sum_{j=1}^m k_j^{-\delta/2}(\omega_j^2/k_j)^{1+\delta/2}}{(\sum_{j=1}^m \omega_j^2/k_j)^{1+\delta/2}} = 0.$$

When the k_j are all bounded, or equal and tend to infinity, the condition $m(n)\sup_{1\le j\le m(n)}|\omega_j(n)| \le \omega_0 < \infty$, for all n, is sufficient for $(\mathcal{W})$ to hold. This condition is satisfied by naive pooling, for which $\omega_j = 1/m$ for $1 \le j \le m$. The distributed estimator $\widehat{\gamma}_n(\omega)$ has the following asymptotic properties.

Theorem 6. *Assume that conditions* $(\mathcal{A})$, $(\mathcal{W})$ *and* $C_2(\gamma,\rho,A)$ *hold. If moreover* $\sup_{1\le j\le m} k_j/n_j \to 0$ *and* $\sum_{j=1}^m \{\sqrt{k_j}|A(n_j/k_j)|\}^2 = \mathrm{O}(1)$, *then*

$$\left(\sum_{j=1}^m \frac{\omega_j^2}{k_j}\right)^{-1/2}\left(\widehat{\gamma}_n(\omega) - \gamma - \frac{1}{1-\rho}\sum_{j=1}^m \omega_j k_j^{\rho}\frac{\Gamma(k_j-\rho+1)}{k_j!}A(n_j/k_j)\right) \xrightarrow{d} \mathcal{N}(0,\gamma^2).$$

Here Γ *denotes Euler's Gamma function. In particular, if* $k\sum_{j=1}^m \omega_j^2/k_j \to v \in [1,\infty)$ *(note that, in this situation, necessarily* $v\ge 1$ *from the Cauchy-Schwarz inequality), one has*

$$\sqrt{k}\left(\widehat{\gamma}_n(\omega) - \gamma - \frac{1}{1-\rho}\sum_{j=1}^m \omega_j k_j^{\rho}\frac{\Gamma(k_j-\rho+1)}{k_j!}A(n_j/k_j)\right) \xrightarrow{d} \mathcal{N}(0,v\gamma^2).$$

Remark 9 (Comparison with earlier results for fixed m**).** *In Corollary 5, since* $c_j = \lim_{n\to\infty} k_1/k_j$,

$$\sum_{j=1}^m \frac{1}{c_j}\sum_{j=1}^m c_j\omega_j^2 = \lim_{n\to\infty}\sum_{j=1}^m \frac{k_j}{k_1}\sum_{j=1}^m \frac{k_1}{k_j}\omega_j^2 = \lim_{n\to\infty} k\sum_{j=1}^m \frac{\omega_j^2}{k_j}.$$

In the case $m\to\infty$, *the right-hand side is nothing but the quantity* v *of Theorem 6: as a consequence, the asymptotic variance in Theorem 6 matches the asymptotic variance obtained in Corollary 5 for fixed* m. *By contrast, the asymptotic bias term is substantially different, with condition* $\sum_{j=1}^m \{\sqrt{k_j}|A(n_j/k_j)|\}^2 = \mathrm{O}(1)$ *expressing that the sum of statistical errors arising in each machine through the use of the individual Hill estimators* $\widehat{\gamma}_j(k_j)$ *should stay bounded in a certain sense.*

Remark 10 (Comparison with earlier results about tail index estimation). *Tail index estimation when* $m\to\infty$ *has been tackled in [3] using naive pooling. Under the weaker version* $(\mathcal{A})$ *of their Condition A, Theorem 6 generalizes their Theorems 1, 2 and 3 by dealing with weighted distributed estimation, by unifying the cases of unbounded and bounded* k_j *and by allowing for unbalanced sample sizes* n_j. *We also remark that the asymptotic variance* γ^2 *obtained in Theorem 2 of [3] for naive pooling is unfortunately not correct. We prove that the asymptotic variance is in fact* $v\gamma^2$, *where in general* $v>1$. *See Section C.5 of the Supplementary Material document for further theoretical and numerical details. As such, the distributed Hill estimator with* $\omega_j = 1/m$ *typically does not achieve the so-called "oracle property" claimed in [3] even if* $\sqrt{k}\sup_{1\le j\le m}|A(n_j/k_j)| \to 0$ *or* $\rho = 0$. *Finally, even though [17] is not written with the distributed inference problem in mind, the estimator* $\widehat{\gamma}_N$ *considered therein is in fact our distributed estimator with equal sample sizes and equal fixed effective sample sizes; using the relationship* $r^{-1}\sum_{j=1}^r \Gamma(j-\rho)/(j-1)! = \Gamma(r+1-\rho)/((1-\rho)r!)$, *following from a straightforward proof by induction on* r, *shows that our Theorem 6 generalizes Theorems 1 and 2 of [17] also.*

Remark 10 motivates the following corollary for the variance-optimal weights $\omega_j = \widehat{\omega}_j^{(\mathrm{Var})} = k_j/k$ (for $1\le j\le m$), which satisfy condition $(\mathcal{W})$ regardless of the behavior of the sequences $k_j = k_j(n)$.

Corollary 9. *Assume that conditions* $(\mathcal{A})$ *and* $C_2(\gamma,\rho,A)$ *hold. If moreover* $\sup_{1\le j\le m} k_j/n_j \to 0$ *and* $\sum_{j=1}^m \{\sqrt{k_j}|A(n_j/k_j)|\}^2 = \mathrm{O}(1)$, *then*

$$\sqrt{k}\left(\widehat{\gamma}_n(\widehat{\omega}_n^{(\mathrm{Var})}) - \gamma - \frac{1}{1-\rho}\sum_{j=1}^m \frac{k_j}{k}\times k_j^{\rho}\frac{\Gamma(k_j-\rho+1)}{k_j!}A(n_j/k_j)\right) \xrightarrow{d} \mathcal{N}(0,\gamma^2).$$

Therefore, as in the case of bounded m, the distributed estimator with $\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})}$ is asymptotically variance-optimal. It is this weighted estimator which possesses the "oracle property" introduced in [3].

We now turn to the asymptotic behavior of the geometrically weighted extreme quantile estimator $\widehat{q}_n^{\star}(1-p|\boldsymbol{\omega})$. Perhaps surprisingly, this estimator is generally not consistent when the k_j are bounded, *i.e.* $\limsup_{n\to\infty} \sup_{1\le j\le m(n)} k_j(n) < \infty$. The rationale is that, while the bias of the shape parameter estimators $\widehat{\gamma}_j(k_j)$ is small, and so averaging them out as $\widehat{\gamma}_n(\boldsymbol{\omega})$ creates a consistent estimator as the number of machines increases, the scale parameter estimators $X_{n_j-k_j:n_j,j}$ are fundamentally biased estimators of $q(1-k_j/n_j)$ when k_j is fixed, so the classical Weissman extrapolation of $X_{n_j-k_j:n_j,j}$ can no longer be correctly applied. However, when $k_j\to\infty$ with $k_j/n_j\to 0$, the distributed extreme quantile estimator may be consistent and asymptotically normal with essentially the same asymptotic distribution as $\widehat{\gamma}_n(\boldsymbol{\omega})$ investigated in Theorem 6. These insights are summarized in the following result.

Theorem 7. *Work under the conditions of Theorem 6, including the assumption* $k\sum_{j=1}^m \omega_j^2/k_j \to v \in [1,\infty)$*, and suppose that* $\rho<0$*. Pick* $p=p(n)\to 0$ *such that* $k/(np)\to\infty$ *and* $\sqrt{k}/\log(k/(np))\to\infty$*, and assume that*

$$\sup_{1\le j\le m}\left|\frac{\log(k_j/(n_j p))}{\log(k/(np))}-1\right|\to 0.$$

(i) *If the* $k_j=k_j(n)$ *are bounded, i.e.* $\limsup_{n\to\infty}\sup_{1\le j\le m(n)} k_j(n)<\infty$*, and* $\omega_j\ge 0$ *for any* j*, then* $\widehat{q}_n^{\star}(1-p|\boldsymbol{\omega})/q(1-p)$ *does not converge to 1 in probability.*

(ii) *If the* $k_j=k_j(n)$ *are such that* $\inf_{1\le j\le m(n)} k_j(n)\to\infty$*, then* $\widehat{q}_n^{\star}(1-p|\boldsymbol{\omega})/q(1-p)\xrightarrow{\mathbb{P}}1$*. If moreover* $\sum_{j=1}^{m(n)} 1/k_j(n)\to 0$*, then*

$$\frac{\sqrt{k}}{\log(k/(np))}\left(\frac{\widehat{q}_n^{\star}(1-p|\boldsymbol{\omega})}{q(1-p)}-1-\frac{1}{1-\rho}\sum_{j=1}^m \omega_j A(n_j/k_j)\right)\xrightarrow{d}\mathcal{N}(0,v\gamma^2).$$

To the best of our knowledge, only [2] have studied extreme quantile estimation when $m\to\infty$. They do not use geometric weighted pooling and only allow for equal k_j and n_j. In that setting, a consequence of their condition (C2) is that $m\le k_1^{1-\delta}$ for some $\delta>0$, while our condition $\sum_{j=1}^m 1/k_j\to 0$ is just the weaker condition $m/k_1\to 0$. In general, our condition is satisfied as long as $m/\inf_{1\le j\le m} k_j\to 0$. Our result, unlike theirs, also investigates the case when the k_j are bounded. The extra assumption linking the k_j, n_j, k, n and p ensures that the sample fractions k_j/n_j are not too dissimilar across machines. It is satisfied if $0<\liminf_{n\to\infty}(n/k)\inf_{1\le j\le m} k_j/n_j \le \limsup_{n\to\infty}(n/k)\sup_{1\le j\le m} k_j/n_j<\infty$. A weaker version of this condition already appears in Theorem 5 for finite m.

4. Filtering to handle serial dependence and covariates

In many applications, the data are recorded with relevant covariates, or are stationary but weakly dependent in a way that can be modeled by a standard time series. Besides, when the X_j share the same tail index, they typically have also asymptotically proportional extreme quantiles (see Remark 7). This suggests that $(X_1,\ldots,X_m)$ can be modeled in many situations by a general location-scale model

$$X_j = g_j(\mathbf{Z}_j)+\sigma_j(\mathbf{Z}_j)\varepsilon_j,\ 1\le j\le m, \tag{2}$$

where the unobserved noise vector $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_m)$ has marginal tail quantile functions U_j satisfying $C_2(\gamma\mathbf{1},\boldsymbol{\rho},\boldsymbol{A})$ and bivariate survival copulae $\overline{C}_{j,\ell}$ satisfying $\mathcal{J}(\boldsymbol{R})$. The g_j and $\sigma_j>0$ are unknown measurable functions of $\boldsymbol{Z}_j \in \mathbb{R}^{l_j}$, for some $l_j \geq 1$. The covariates $\boldsymbol{Z}_j$ can be fully observed (in traditional regression settings) or partially or not at all observed (in a time series model with past unobserved innovations or volatility terms). The noise variable ε_j is assumed to be independent of $\boldsymbol{Z}_j$. In this model, the jth tail index γ_j is a constant function of $\boldsymbol{Z}_j$, which is a typical assumption in extremal regression based on heteroskedastic location-scale or quantile regression models [20,21], and more generally a common assumption in the analysis of heteroskedastic extremes, see for instance [9].

In this setting, there are two main approaches to the extreme value analysis of the X_j. The first one is the conditional extreme value analysis of high $(1-p)$th quantiles of X_j given $\boldsymbol{Z}_j = z_j$, expressed as $q_{X_j|\boldsymbol{Z}_j=z_j}(1-p) = g_j(z_j) + \sigma_j(z_j) q_j(1-p)$, where q_j is the quantile function of ε_j. When the $\boldsymbol{Z}_j$ consist in lagged values of the X_j, estimating these so-called dynamic quantiles is typically appropriate for short-term risk management. In finance, for example, it is of interest to capture the key dynamic properties of financial asset returns, such as volatility clustering, so as to give a better understanding of the current riskiness of a portfolio. In environmental science, one may be interested in forecasting extreme rainfall levels at a short time horizon given values of weather parameters today, or given outputs from a climate model [20,21]. The second approach is the estimation of extreme parameters of the stationary distribution of each X_j. In the context where (2) is a time series model, this will be appropriate for long-term risk management, for instance when trying to estimate high quantiles or return levels over hundreds or thousands of years. In this section we focus on the first approach.

Let then the pairs $(X_{i,j}, \boldsymbol{Z}_{i,j})$, $1 \leq i \leq n_j$, be part of a strictly stationary sequence such that $X_{i,j} = g_j(\boldsymbol{Z}_{i,j}) + \sigma_j(\boldsymbol{Z}_{i,j})\varepsilon_{i,j}$ for $1 \leq j \leq m$ and $1 \leq i \leq n_j$. The $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,m})$ are assumed to be independent copies of $\boldsymbol{\varepsilon}$ as above. To estimate extreme conditional quantiles, we first estimate the location and scale components g_j and σ_j of the model (under suitable identifiability and regularity conditions), and then filter the $X_{i,j}$ to obtain residuals $\widehat{\varepsilon}_{i,j}^{(n_j)}$ close to the unobserved errors $\varepsilon_{i,j}$. This results in j residual-based Hill estimators $\widehat{\gamma}_j(k_j) = k_j^{-1} \sum_{i=1}^{k_j} \log(\widehat{\varepsilon}^{(n_j)}_{n_j-i+1:n_j,j} / \widehat{\varepsilon}^{(n_j)}_{n_j-k_j:n_j,j})$. These can be combined in a pooled version $\widehat{\gamma}_n(\boldsymbol{\omega}) = \sum_{j=1}^m \omega_j \widehat{\gamma}_j(k_j)$ whose asymptotic normality can be proved under a high-level condition on the discrepancy between the errors and the corresponding residuals.

Theorem 8. *Assume that $\boldsymbol{\varepsilon}$ satisfies $\mathcal{C}_2(\gamma\mathbf{1}, \boldsymbol{\rho}, \boldsymbol{A})$ and $\mathcal{J}(\boldsymbol{R})$. Under the conditions of Theorem 1 on k_j, n_j and $\boldsymbol{\omega}$, if $\max_{1\leq j\leq m} \sqrt{k_j} \max_{1\leq i\leq n_j} |\widehat{\varepsilon}_{i,j}^{(n_j)} - \varepsilon_{i,j}|/(1+|\varepsilon_{i,j}|) \xrightarrow{\mathbb{P}} 0$, then $\sqrt{k}(\widehat{\gamma}_n(\boldsymbol{\omega}) - \gamma) \xrightarrow{d} \mathcal{N}(\boldsymbol{\omega}^\top \boldsymbol{B_c}, \boldsymbol{\omega}^\top \mathbf{V_c} \boldsymbol{\omega})$ with $\boldsymbol{B_c}$ and $\mathbf{V_c}$ defined as in Theorem 1. If $\widehat{\boldsymbol{\omega}}_n^\top \mathbf{1} = 1$ with $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$, then $\sqrt{k}(\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n) - \widehat{\gamma}_n(\boldsymbol{\omega})) = \mathrm{o}_{\mathbb{P}}(1)$. In particular, $\widehat{\gamma}_n(\widehat{\boldsymbol{\omega}}_n)$ has the same $\sqrt{k}$–asymptotic behavior as $\widehat{\gamma}_n(\boldsymbol{\omega})$.*

Remark 11 (Linking the regression and unconditional settings). *In the unrealistic setting where the g_j and σ_j were known, the residuals would be exactly the $\varepsilon_{i,j}$, which would yield the asymptotic normality of the pooled estimator based on the $\widetilde{\gamma}_j(k_j) = k_j^{-1} \sum_{i=1}^{k_j} \log(\varepsilon_{n_j-i+1:n_j,j}/\varepsilon_{n_j-k_j:n_j,j})$. This means that Theorem 1 can be seen as a (statistically unnatural) consequence of Theorem 8.*

Remark 12 (On pooling residuals). *One might argue that pooling the residuals in a single sample and using the Hill estimator is more efficient than pooling residual-based Hill estimators calculated in individual subsamples. However, if the model is misspecified, heteroskedasticity could still remain in the residuals, and those with the largest scale might swamp the other residuals in the pooled sample, resulting thus in a large loss of estimation accuracy. Pooling the residual-based Hill estimates instead provides more protection against departures from the assumed location-scale model.*

The condition on the discrepancy between residuals and innovations in Theorem 8 is typically satisfied as soon as the location and scale components g_j and σ_j are estimated at a faster rate than $\sqrt{k_j}$. This can be checked theoretically in a variety of regression models, see [11] for examples.

Given consistent estimators $\widehat{g}_j(z_j)$ of $g_j(z_j)$ and $\widehat{\sigma}_j(z_j)$ of $\sigma_j(z_j)$, and using the residual-based Weissman estimator of $q_j(1-p)$, we estimate $q_{X_j|\mathbf{Z}_j=z_j}(1-p)$ by the location-scale estimator

$$\widehat{q}^{\star}_{X_j|\mathbf{Z}_j=z_j}(1-p|k_j,\boldsymbol{\omega}) = \widehat{g}_j(z_j) + \widehat{\sigma}_j(z_j)\left(\frac{k_j}{n_j p}\right)^{\widehat{\gamma}_n(\boldsymbol{\omega})} \widehat{\varepsilon}^{(n_j)}_{n_j-k_j:n_j,j}.$$

When $\boldsymbol{\varepsilon}$ satisfies the tail homoskedasticity condition $(\mathcal{H})$ of Section 2.3, the quantiles $q_j(1-p)$ can be estimated by a geometrically pooled estimator, which leads to the estimator of $q_{X_j|\mathbf{Z}_j=z_j}(1-p)$ below:

$$\widetilde{q}^{\star}_{X_j|\mathbf{Z}_j=z_j}(1-p|\boldsymbol{\omega}) = \widehat{g}_j(z_j) + \widehat{\sigma}_j(z_j)\prod_{i=1}^{m}\left[\left(\frac{k_i}{n_i p}\right)^{\widehat{\gamma}_i(k_i)} \widehat{\varepsilon}^{(n_i)}_{n_i-k_i:n_i,i}\right]^{\omega_i}.$$

Our final weak convergence result establishes the asymptotic normality of these two estimators.

Theorem 9. *Work under the conditions of Theorem 8 with $\rho_j<0$ for all $j\in\{1,\dots,m\}$. Pick $p=p(n)\to 0$ such that $k/(np)\to\infty$ and $\sqrt{k}/\log(k/(np))\to\infty$. Let $\boldsymbol{\omega},\widehat{\boldsymbol{\omega}}_n$ be such that $\widehat{\boldsymbol{\omega}}_n^\top\mathbf{1}=1$ and $\widehat{\boldsymbol{\omega}}_n\xrightarrow{\mathbb{P}}\boldsymbol{\omega}$. Finally, assume that the estimators $\widehat{g}_j(z_j)$ and $\widehat{\sigma}_j(z_j)$ satisfy $\widehat{g}_j(z_j)-g_j(z_j)=\mathrm{O}_{\mathbb{P}}(1)$ and $\sqrt{k_j}(\widehat{\sigma}_j(z_j)-\sigma_j(z_j))=\mathrm{O}_{\mathbb{P}}(1)$. Then, for any j,*

$$\frac{\sqrt{k}}{\log(k/(np))}\left(\frac{\widehat{q}^{\star}_{X_j|\mathbf{Z}_j=z_j}(1-p|k_j,\widehat{\boldsymbol{\omega}}_n)}{q_{X_j|\mathbf{Z}_j=z_j}(1-p)}-1\right) = \sqrt{k}(\widehat{\gamma}_n(\boldsymbol{\omega})-\gamma)+\mathrm{o}_{\mathbb{P}}(1)\xrightarrow{d}\mathcal{N}(\boldsymbol{\omega}^\top\boldsymbol{B_c},\boldsymbol{\omega}^\top\mathbf{V_c}\boldsymbol{\omega}).$$

If moreover assumption $(\mathcal{H})$ holds then, for any j,

$$\frac{\sqrt{k}}{\log(k/(np))}\left(\frac{\widetilde{q}^{\star}_{X_j|\mathbf{Z}_j=z_j}(1-p|\widehat{\boldsymbol{\omega}}_n)}{q_{X_j|\mathbf{Z}_j=z_j}(1-p)}-1\right) = \sqrt{k}(\widehat{\gamma}_n(\boldsymbol{\omega})-\gamma)+\mathrm{o}_{\mathbb{P}}(1)\xrightarrow{d}\mathcal{N}(\boldsymbol{\omega}^\top\boldsymbol{B_c},\boldsymbol{\omega}^\top\mathbf{V_c}\boldsymbol{\omega}).$$

5. Finite-sample study

5.1. Simulation experiments

We briefly describe the finite-sample performance of the pooling approach for extreme value estimation and inference, in the general pooling framework (Section 2), the distributed inference framework (Section 3) and after filtering in regression models (Section 4). A complete description of our setup and numerical results can be found in Section D.1 of the Supplementary Material document.

5.1.1. General setup: Pooling for tail index and extreme quantile inference

Dimensions $m\in\{2,3,4,5\}$ are considered, with balanced and unbalanced sample sizes. We use either unit Fréchet or absolute Student marginals (*i.e.* the absolute value of a Student) with 1 degree of freedom (and therefore $\gamma_j=1$ for any j) with a dependence structure given by a Gaussian, Student, Clayton or Gumbel copula. The Clayton and Gaussian copulae represent cases of asymptotic independence, while the Gumbel and Student copulae are cases of asymptotic dependence.

For a total sample size of $n=\sum_{j=1}^m n_j=1{,}000$ across all subsamples, we compare four pooled tail index estimators (naive *i.e.* simple average, variance-optimal and AMSE-optimal as in Corollary 1, and

AMSE-optimal with pooled second-order estimates as described below Corollary 2) with the benchmark Hill estimator. To illustrate Theorem 2, we compare the related four geometrically pooled extreme quantile estimators $\widehat{q}_n^\star(1-p|\widehat{\omega}_n)$ at level $1-p=0.999$ with the arithmetic mean of the Weissman estimators $\widehat{q}_j^\star(1-p|k_j)$, *i.e.* $\overline{q}_n^\star(1-p|1/m,\ldots,1/m)$ with the notation of Section 2.3, and the benchmark Weissman estimator for the pooled dataset. We compute Monte Carlo approximations of the Mean Squared Error (MSE) and (to illustrate Corollaries 3 and 4) of the actual coverage probability for the asymptotic confidence intervals with 95% nominal level arising from our asymptotic theory; for the Hill (resp. Weissman) estimator, the asymptotic distribution is normal with mean 0 and variance γ^2/k (resp. $\gamma^2 \times \log^2(k/(np))/k$), see Theorem 3.2.5, p.74 in [5] (resp. Theorem 4.3.8, p.138 therein). The MSE, coverage probability, and average length of each confidence interval are reported in Figures D.2–D.5 in the Supplementary Material document. We also illustrate the performance of the tail homogeneity test based on the test statistic Λ_n with 5% nominal type I error level, outlined before Corollary 3, in the same models but with the key difference that we let the tail index γ_1 of the first marginal vary between 0.2 and 5. Representing the rejection rate of the test as a function of γ_1, see Figure D.6, makes it possible to assess its power as the model under consideration gets away from tail homogeneity.

The variance-optimal and AMSE-optimal estimators outperform by far the naive pooled estimator on the basis of the MSE, when there is strong unbalance between sample sizes. Differences in performance get larger as the degree of unbalance increases. Confidence intervals based on variance-optimal and AMSE-optimal estimators with pooled second-order estimates are typically substantially narrower than with naive pooling and their coverage is close to nominal, while the finite-sample coverage of the AMSE-optimal intervals without pooling in the second-order parameter estimates is sometimes less satisfactory. This is due to the fact that second-order parameter estimators in each marginal distribution may sometimes have poor performance; pooling them together substantially reduces the high variability they suffer from. Most importantly, the 95% confidence intervals constructed using the benchmark Hill estimator have actual coverage that can be as low as 75% when substantial asymptotic dependence between subsamples is present. Conclusions about extreme quantile estimation are similar, with the added value that geometrically pooled estimators outperform by far the naive mean of Weissman estimators. The tail homogeneity test, meanwhile, has sensible power overall for a common effective sample fraction $k_j/n_j = 20\%$ across subsamples, although power curves tend to be asymmetric and with a minimum not necessarily located at $\gamma_1 = 1$ when sample sizes are highly unbalanced. The explanation is that Λ_n compares the marginal estimators $\widehat{\gamma}_j(k_j)$ to the variance-optimal pooled estimator, so that if one of the $\widehat{\gamma}_j(k_j)$ behaves poorly, for example due to low size of the jth subsample, then so will Λ_n. Figure D.7 shows that, with the same models but with equal sample sizes, power curves are indeed symmetric and more quickly reach very high values as γ_1 gets away from 1. Similar conclusions were reached with the tail homoskedasticity test based on the test statistic $L_n(p)$, outlined before Corollary 4. Results for the lower total sample size $n = 400$ are overall the same, even though in this very difficult setting certain subsamples have size as low as 20: see Figures D.8–D.11.

5.1.2. Distributed inference of extreme values

In order to illustrate the results of Section 3, we assume that the $X_{i,j}$ are i.i.d. Fréchet, absolute Student or Burr with tail index $\gamma = 1$. We consider dimensions $m \in \{2, 5, 10, 20\}$ in balanced and highly unbalanced setups, with the situation $m = 20$ representing a case when m can be considered large (see [3]). We compare again the tail index estimators described in Section 5.1.1, as well as the listed extreme quantile estimators plus the subsample Weissman estimator in the first marginal with the corresponding individual Hill estimator replaced by the variance-optimal pooled tail index estimator, that is, $\widehat{q}_1^\star(1-p|k_1,\widehat{\omega}_n)$ where $\widehat{\omega}_n$ is taken to be variance-optimal weights. The results, reported in Figures D.12–D.16 in the Supplementary Material document, indicate that, in accordance with Corollaries 5, 6, 7 and 8 (see Figures D.12–D.13, for fixed m) and Theorem 6, Corollary 9 and Theorem 7 (see

Figure D.14, for large m), our proposed variance-optimal and AMSE-optimal methods (with pooled second-order estimates) perform comparably to the unfeasible Hill and Weissman estimators applied to the pooled dataset and, as the unbalance between sample sizes increases, they outperform the naive distributed estimators, with much shorter confidence intervals having correct coverage, and lower MSE. Geometric pooling is clearly beneficial as far as extreme quantile estimation is concerned. Figures D.12–D.14 suggest that the variance-optimal estimator behaves well regardless of how unbalanced effective sample sizes are, as stated in Theorem 5. Moreover, as expected following Remark 8 and Theorem 3, the AMSE-optimal distributed estimator is very close to the variance-optimal estimator when sample fractions are equal, and in that setting, the performance of the variance-optimal estimator is virtually identical to that of the benchmark Hill estimator. By contrast, in accordance with Theorem 4, AMSE-optimal pooling is overall the best solution when sample fractions are substantially different, as Figure D.15 illustrates in the case $m = 2$; in particular, when $(n_1, n_2) = (50, 950)$ and $k_1 = k_2 = 20$, the MSE of the AMSE-optimal distributed estimator (on the log-scale) is around 0.04, with the corresponding MSE of the unfeasible Hill benchmark around 0.05, representing an improvement of about 20%. Finally, it follows from Figure D.16 that when n_1/n decreases, the performance of the subsample Weissman estimator $\widehat{q}_1^{\star}(1 - p|k_1, \widehat{\omega}_n^{(\mathrm{Var})})$ substantially deteriorates relative to that of the geometrically pooled extreme quantile estimator, with the difference in performance getting larger as ρ gets closer to 0. This illustrates the fact that when the tail homoskedasticity assumption holds, it is indeed valuable to pool not only the estimates of the tail index but also those of scale in the Pareto extrapolation relationship, particularly when sample sizes are unbalanced.

5.1.3. Pooling in location-scale models using residual-based estimators

We conclude with an illustration of the results of Section 4. We consider four autoregressive models with either 1 or 2 lags, in dimensions $m = 2$ or 3, possibly misspecified and having Student innovations with 1 degree of freedom. We first filter each univariate time series using an AR(1) filter and a Student maximum likelihood estimator of the coefficient(s), thus yielding subsamples of univariate residuals to which the tail index estimators of Section 4 are applied. Results are represented in Figures D.17-D.20. The bias and variance of the pooled estimators are reasonably low and behave as expected, with higher (resp. lower) sample fractions associated to larger bias (resp. higher variance). As suggested by Theorem 8, MSE values are in line with those observed when the data is independent through time.

5.2. Data analysis

5.2.1. Distributed inference for car insurance data

The first dataset comprises total claim amounts for car insurance companies in the five US states of Iowa (n_1 = 2,601), Kansas (n_2 = 798), Missouri (n_3 = 3,150), Nebraska (n_4 = 1,703) and Oklahoma (n_5 = 882) between January and February 2011, for a total sample size of n = 9,134 (see Figures D.21-D.22 of the Supplementary Material document for an exploratory analysis). As in [3], we assume that companies cannot share their data but each company is willing to share its statistical analysis to enhance its appraisal of tail risk. Unlike [3], however, our distributed inference method can handle the different subsample sizes n_j and hence the full dataset, and allows to estimate extreme quantiles. We therefore compare our results using the full data with those obtained by exactly reproducing the approach of [3], which consists in applying the naive pooled estimator after subsampling at random 700 observations in each state, as described in their Supplement B. As a benchmark, we use the (unfeasible) Hill and Weissman estimates, obtained from the combined n data points. Results are given in Figure 1.

We first check the equality of tail indices by testing for tail homogeneity across the 5 states on full data and the subsampled data in each state, using the theory developed in Section 2.4 under the constraint of independence between subsamples (see Remark 5). From the p-values corresponding to our test statistic Λ_n in Figure 1(A), we can comfortably conclude the equality of individual tail indices at the three significance levels 0.10, 0.05 and 0.01. It is remarkable that the p-values plot remains quite stable when moving from the full 5 samples of total size 9,134 to the subsamples of total size $5 \times 700 = 3{,}500$. This indicates that the asymptotic chi-square regime is attained reasonably quickly.

Figure 1(B) compares our variance-optimal distributed estimator $\widehat{\gamma}_n(\widehat{\omega}_n^{(\mathrm{Var})})$, based on the full data, with the naive distributed estimator $\widehat{\gamma}_n(1/m, \ldots, 1/m)$ of [3], which relies on the subsampled data, and with the benchmark Hill estimator $\widehat{\gamma}_n^{(\mathrm{Hill})}$ along with their respective asymptotic 95% confidence intervals. In contrast to the naive estimates and their associated confidence intervals, our optimal weighted estimates and their confidence intervals are, respectively, almost identical to the Hill estimates and their corresponding confidence intervals, as is to be expected from Theorem 3. Our variance-optimal confidence intervals are found to be around 40% shorter than those of [3]. We arrive at a similar conclusion, in Figure 1(C), when restricting the analysis to the branches in Kansas and Missouri, whose subsample sizes 798 and 3,150 are strongly unbalanced, and using the full data from these states for both the variance-optimal and naive distributed estimates. Here, the variance-optimal confidence intervals are found to be roughly 20% shorter than those relative to the naive estimator.

The test of extreme quantile equivalence developed in Corollary 4 is implemented for the two extreme quantile levels $1 - p = 0.999 \approx 1 - 1/\max_j n_j$ and $1 - p = 0.9999 \approx 1 - 1/n$, resulting in the p-values from the test statistic $L_n(p)$ displayed in Figure 1(D) for both full and subsampled data. The test overall allows to accept the assumption of tail homoskedasticity across states, with p-values getting higher as p decreases. The rationale behind this behavior in this distributed setting is that, as extreme quantile levels increase, the shape of the approximating Pareto distribution gets more important relative to its scale. As such, because mere differences in scale can no longer be detected in the far tail as $p \downarrow 0$, the test actually becomes less powerful against the sub-alternative of proportional quantiles. Finally, the resulting variance-optimal distributed estimates and confidence intervals for extreme quantiles are found to be virtually indistinguishable from the ideal Weissman analogs, whereas they appreciably outperform the naive distributed competitors, as can be seen in Figure 1(E) and 1(F) for $p = 0.0001$.

5.2.2. *Pooling for regional inference on extreme rainfall*

In the second dataset, rainfall measurements are collected daily by the Florida Automated Weather Network at 49 gauge stations, over different periods between December 1997 and May 2021 (see https://fawn.ifas.ufl.edu/data/fawnpub/). We focus on the eight stations indicated with pin markers in the map in Figure 2, whose aggregated monthly rainfall distributions exhibit heavy-tailed behavior; the upper tail heaviness was ascertained in an exploratory analysis using moment and generalized Hill estimators (see *e.g.* [1]). Individual sample sizes n_j are rather short, ranging from 221 to 281. Extreme value inference at each site will thus be subject to large uncertainty. By contrast, the pooling approach reduces uncertainty by borrowing tail information from stations having the same tail characteristics.

Our exploratory analysis shows first that the eight monthly time series are all stationary according to the classical KPSS and ADF tests. A first distinctive property is that the data for each station in the cluster of red pinned stations near the northern border of Florida are, in contrast to those in the green cluster stretching along the east coast, not autocorrelated according to the Ljung-Box test. The time series (X_t) in the green cluster of stations can be fitted by simple seasonal ARMA models of the form $\Phi_{12}(B^{12})\Phi(B)X_t = c + \Theta_{12}(B^{12})\Theta(B)\varepsilon_t$, where ε_t is a white noise process and c is a constant, with $\Phi(B)$, $\Theta(B)$, $\Phi_{12}(B^{12})$ and $\Theta_{12}(B^{12})$ being respectively polynomials of degree $p, q \in \{0, 1, 2\}$ in the lag operator B and $p_{12}, q_{12} \in \{1, 2\}$ in B^{12}. Following Section 4, the residuals $(\widehat{\varepsilon}_{t,j}^{(n_j)}, 1 \le t \le n_j)$, are the basic tool for estimating the tail index in station j. Figures D.23 and D.24 in the Supplementary Material document show the histograms and Hill plots obtained for the eight stations. A second distinctive property is that the three stations in the red cluster have very similar Hill estimates, while the

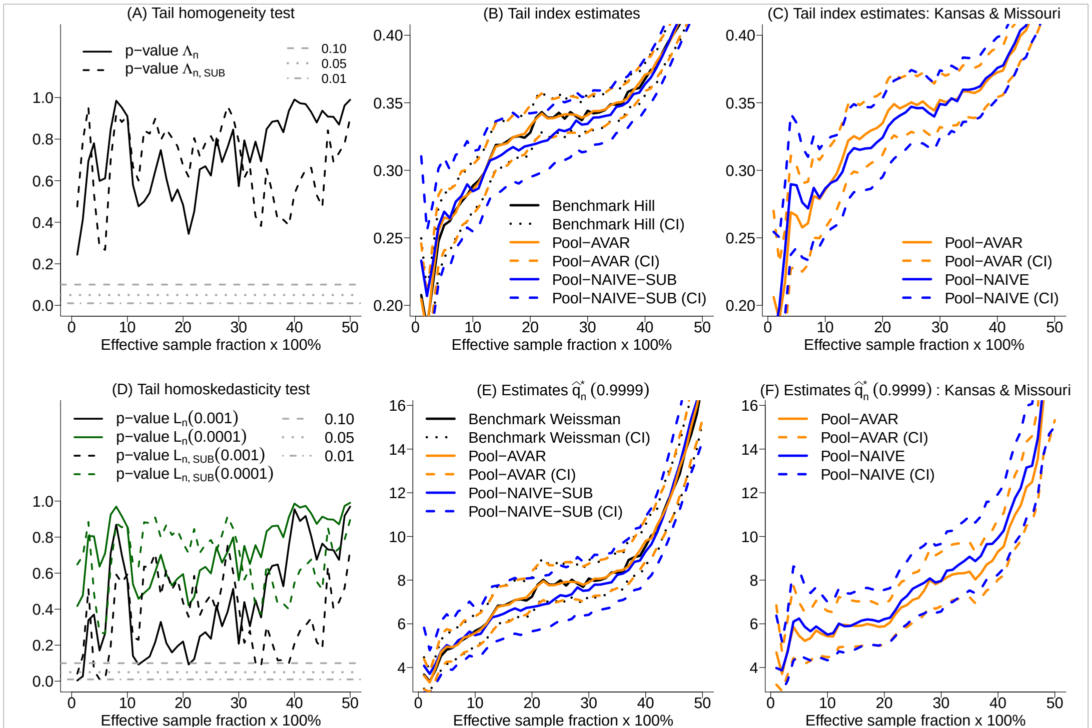

Figure 1: Car insurance data. In (A) and (D), $\Lambda_{n,\mathrm{SUB}}$ and $L_{n,\mathrm{SUB}}(p)$ denote the test statistics Λ_n and $L_n(p)$ calculated from the subsampled data of total size $5 \times 700 = 3{,}500$. In (B) and (C), Pool-AVAR and Pool-NAIVE respectively denote the variance-optimal pooled estimator and the naive pooled estimator of the tail index. In (E) and (F), Pool-AVAR and Pool-NAIVE respectively denote the variance-optimal pooled quantile estimator $\widehat{q}_n^{\star}(1-p|\widehat{\omega}_n^{(\mathrm{Var})})$ and its unweighted analog. Pool-NAIVE-SUB stands for these estimators calculated on the subsampled data. All estimates are represented as functions of the sample fraction k_j/n_j, assumed to be identical for each j.

five stations in the green cluster also have similar Hill estimates that are rather different from those in the red cluster. Table D.1 summarizes extreme value information gathered from each station.

That the eight stations do not have the same tail index is confirmed by the tail homogeneity test in Figure 2(A), where the plot of p-values from the test statistic Λ_n becomes very stable below the three significance levels 0.10, 0.05 and 0.01. By contrast, we can comfortably conclude the equality of tail indices in both red and green clusters from the tail homogeneity tests in Figure 2(B) and 2(C). The tail homoskedasticity test, implemented for the two extreme quantile levels $1-p \approx 1-1/\max_j n_j$ and $1-p \approx 1-1/n$, in fact allows to even accept the assumption of extreme quantile equivalence across stations in each cluster, see Figure 2(D) and 2(E). Therefore, the Hill and Weissman estimators of the common tail index and extreme quantiles could be directly calculated from the combined data in each cluster. However, it should be clear that the key question of inference based on these ideal estimators remains open in this particular application. Indeed, combining subsamples in each cluster of stations results in a single sample of asymptotically dependent data for which the asymptotic theory of the Hill and Weissman estimators is still unavailable in the extreme value literature. Our regional pooled estimators provide a satisfactory solution, reducing substantially the huge uncertainty inherent to local inference at each site, as can be seen from Figure 2(F)-2(I). For the red cluster, Figure 2(F) shows that both naive and variance-optimal pooled estimators of the tail index are very close to the benchmark Hill estimator, while the asymptotic 95% variance-optimal confidence intervals are quite stable and narrower relative to the Hill-based confidence intervals obtained individually from each subsample. We arrive at the same conclusion for the green cluster in Figure 2(G), where both pooling-type confidence intervals appear to be much tighter than the individual Hill-based confidence interval obtained from the largest subsample. Likewise, when estimating the extreme quantile of order $1 - p \approx 1 - 1/n$, the individual Weissman-based confidence interval obtained from the largest subsample in Figure 2(H) and 2(I), for raw data in the red cluster and for residuals in the green cluster, tends to be unstable and twice as wide as our pooling confidence intervals.

6. Discussion

We provide a wide-ranging treatment of weighted pooling strategies for inference about the tail index and extreme quantiles of m samples of heavy-tailed data, where m can be fixed or increasing with the total sample size. We allow for different sample sizes and effective sample sizes, heterogeneity in marginal distributions, dependence across samples, heteroskedasticity and/or dependence across time represented by location-scale dependence upon possibly unobserved covariates. Our weighted pooling methods minimize either the asymptotic variance or the AMSE of the pooled estimator. This results in an off-the-shelf device applicable by testing first the assumption of equal tail indices or extreme quantiles through likelihood ratio-type tests, before using our weighted pooling inferential procedure based on asymptotic Gaussian confidence intervals. Our experience indicates considerable reductions in length (from 20% up to 40% in the insurance dataset) of the Gaussian confidence intervals when using variance-optimal pooling instead of the naive version, all while staying very close to nominal coverage. This means that the variance-optimal proposal substantially reduces finite-sample uncertainty compared to naive pooling and standard extreme value inference in each individual subsample.

Another merit of the pooling approach, showcased through regional inference on extreme rainfall, is its ability to handle serial dependence within the data in a seasonal autoregressive time series model. Combining subsamples in each cluster of raingauge stations would result in a single sample of asymptotically dependent data for which the asymptotic theory of the Hill and Weissman estimators is still unavailable in the extreme value literature, thus making statistical inference based on their asymptotics unfeasible. By contrast, weighted pooling affords a remarkably simple solution that substantially reduces the huge uncertainty inherent to local inference at each site, and therefore represents significant progress. To make the proposed technique even more flexible, it would be interesting to work out the case of serially dependent errors in time series or regression models, which corresponds to the situation of model misspecification, and also to investigate the case when $m \to \infty$ in that setting, to handle applications similar to our rainfall example when data is collected across a much larger geographical region. Yet another important question, particularly with long-term risk management applications in environmental science in mind, would be to extend our theory so as to allow for pooling-based inference about stationary (rather than dynamic) extreme value parameters of the X_j when the available samples of data come from time series. Theoretical results along these lines are left for future research.

Acknowledgments

The authors acknowledge an anonymous Associate Editor and two anonymous reviewers for their helpful comments that led to an improved version of this paper. A. Daouia and G. Stupfler acknowledge financial support from the French National Research Agency under the grants ANR-19-CE40-0013 (ExtremReg project) and ANR-17-EURE-0010 (EUR CHESS), as well as from the TSE-HEC ACPR Chair and from an AXA Research Fund Award on “Mitigating risk in the wake of the COVID-19 pandemic”. G. Stupfler acknowledges further support from the Centre Henri Lebesgue (ANR-11-LABX-0020-01). S.A. Padoan is supported by the Bocconi Institute for Data Science and Analytics (BIDSA), Italy.

References

[1] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. L. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

[2] CHEN, L., LI, D. and ZHOU, C. (2021). Distributed inference for tail empirical and quantile processes. arXiv:2108.01327v1.

[3] CHEN, L., LI, D. and ZHOU, C. (2022). Distributed inference for the extreme value index. *Biometrika* **109** 257–264.

[4] COCHRAN, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Suppl. J. R. Stat. Soc.* **4** 102–118.

[5] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer.

[6] DEMATTEO, A. and CLÉMENÇON, S. (2016). On tail index estimation based on multivariate data. *J. Nonparametr. Stat.* **28** 152–176.

[7] DREES, H. (1998). Optimal rates of convergence for estimates of the extreme value index. *Ann. Statist.* **26** 434–448.

[8] DREES, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9** 617–657.

[9] EINMAHL, J. H. J., DE HAAN, L. and ZHOU, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **78** 31–51.

[10] EINMAHL, J. H. J., YANG, F. and ZHOU, C. (2021). Testing the multivariate regular variation model. *J. Bus. Econom. Statist.* **39** 907–919.

[11] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2021). Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. *Ann. Statist.* **49** 3358–3382.

[12] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2022). On automatic bias reduction for extreme expectile estimation. *Statist. Comput.* **32** 64.

[13] HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174.

[14] KINSVATER, P., FRIED, R. and LILIENTHAL, J. (2016). Regional extreme value index estimation and a test of tail homogeneity. *Environmetrics* **27** 103–115.

[15] KNIGHT, K. and BASSETT, G. W. (2003). Second order improvements of sample quantiles using subsamples. Unpublished manuscript.

[16] PADOAN, S. A. and STUPFLER, G. (2022). Joint inference on extreme expectiles for multivariate heavy-tailed distributions. *Bernoulli* **28** 1021–1048.

[17] QI, Y. (2010). On the tail index of a heavy tailed distribution. *Ann. Inst. Statist. Math.* **62** 277–298.

[18] SCHMIDT, R. and STADTMÜLLER, U. (2006). Non-parametric estimation of tail dependence. *Scand. J. Statist.* **33** 307–335.

[19] STUPFLER, G. (2019). On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. *Extremes* **22** 749–769.

[20] VELTHOEN, J., CAI, J. J., JONGBLOED, G. and SCHMEITS, M. (2019). Improving precipitation forecasts using extreme quantile regression. *Extremes* **22** 599–622.

[21] WANG, H. J., LI, D. and HE, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *J. Amer. Statist. Assoc.* **107** 1453–1464.

[22] WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73** 812–815.

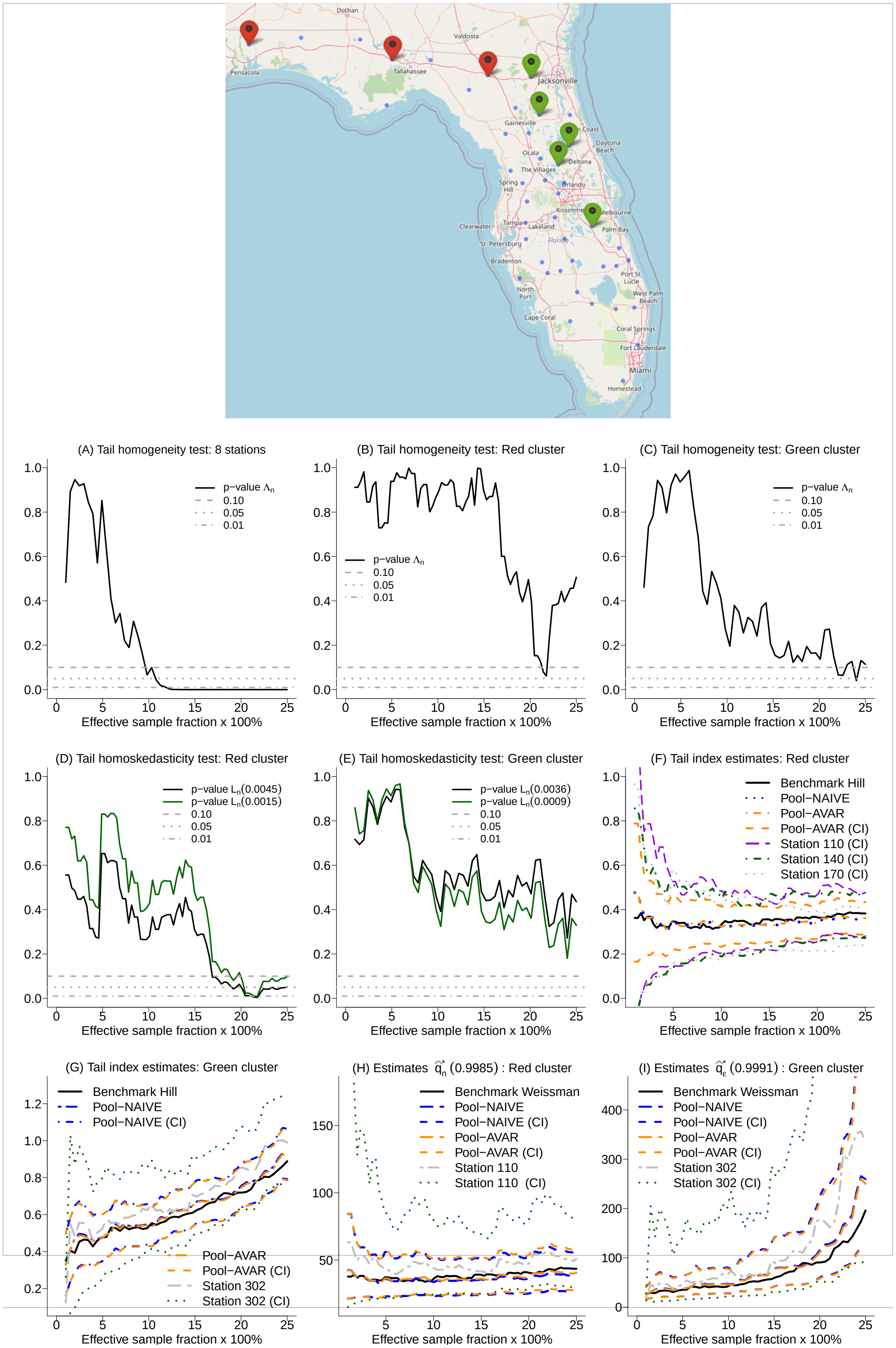

Figure 2: Florida rainfall data. Map of Florida along with its gauge stations and inferential results (notation as in Figure 1).

Appendix A: General pooling theory

Theorem A.1.

Remark A.1 (Asymptotic variance-optimal weights and pseudo-maximum likelihood).

Remark A.2 (AMSE-optimal weights and regularization of bias-optimal weights).

Remark A.3 (On the improvement in asymptotic variance and the sensitivity to uncertainty).

Appendix B: Results of the main paper and their proofs

B.1. Auxiliary results

Lemma B.1.

Proof of Lemma B.1. □

Lemma B.2.

Proof of Lemma B.2. □

Lemma B.3.

Proof of Lemma B.3. □

Lemma B.4.

Proof of Lemma B.4. □

Lemma B.5.

Proof of Lemma B.5. □

B.2. Main results

Proof of Theorem 1. □

Proof of Corollary 1. □

Proof of Corollary 2. □

Proof of Theorem 2. □

Proof of Corollary 3. □

Proof of Corollary 4. □

Proof of Corollary 5. □

Proof of Corollary 6. □

Proof of Theorem 3. □

Proof of Theorem 4. □

Proof of Corollary 7. □

Proof of Corollary 8. □

Proof of Theorem 5. □

Proof of Theorem 6. □

Proof of Theorem 7. □

Proof of Theorem 8. □

Appendix C: Further results, expanded remarks and related calculations

C.1. About the β_j and ρ_j estimators used in the estimation of bias terms

C.2. About Remark 1

Proposition C.1.

Proof of Proposition C.1. □

Proposition C.2.

Proof of Proposition C.2. □

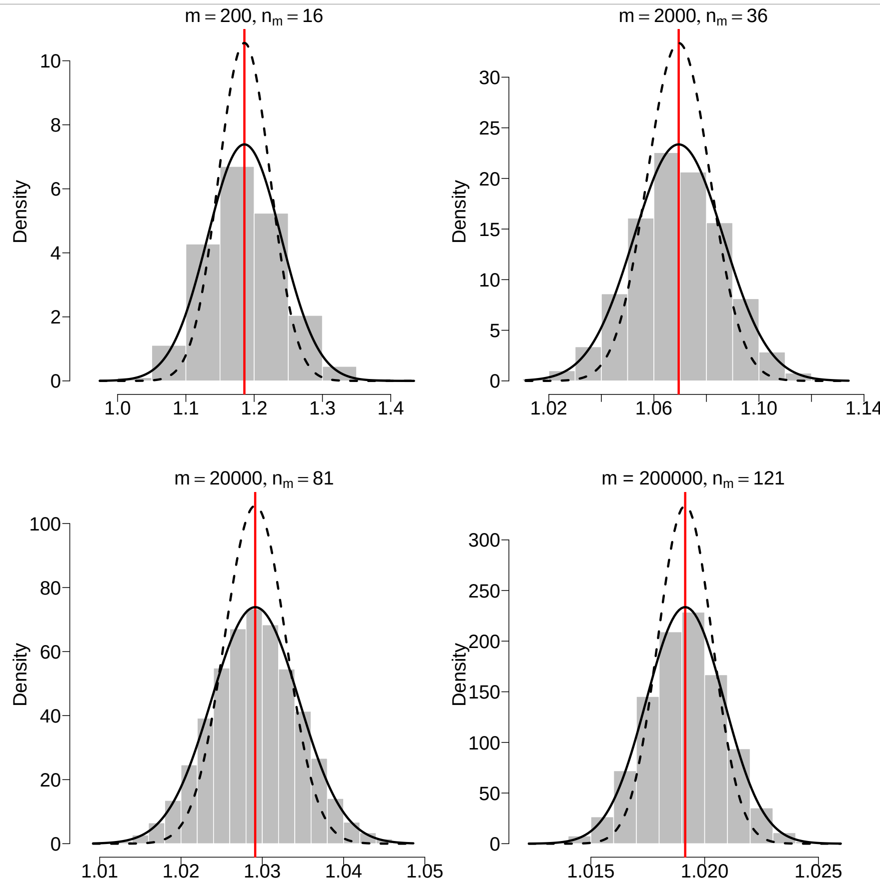

Figure C.1: Empirical distribution of the naive distributed Hill estimator in the simulation setup described in Section C.5. Top left: $m = 200$, top right: $m = 2{,}000$, bottom left: $m = 20{,}000$, bottom right: $m = 200{,}000$. The red vertical line represents the mean of the distributed Hill estimates, and the dashed (resp. solid) line represents the normal density curve with mean $\overline{\gamma}$ (*i.e.* the empirical mean of the distributed Hill estimates) and variance $1/k$ (resp. $49/(24k)$).

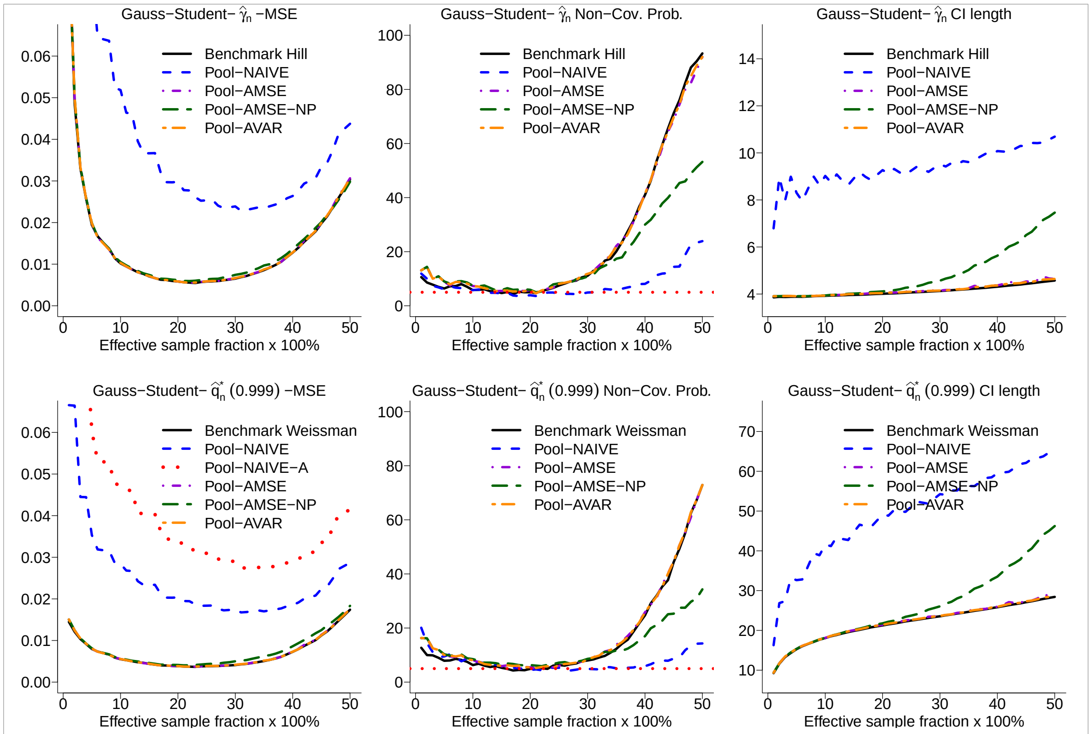

Figure D.2: Simulation results, general pooling setting, Gaussian-Student model (a). Top row: tail index estimation; bottom row: extreme quantile estimation at level $1 - p = 0.999$. Left panels: MSE of the point estimators; middle panels: non-coverage probability of the asymptotic confidence intervals, where the red horizontal dotted line represents the 5% nominal non-coverage probability; right panels: average length of the 95% confidence intervals multiplied by $\sqrt{k}$. All results are represented as functions of the sample fraction k_j/n_j (identical for each j). In the bottom left panels, the MSEs represented are the relative MSEs of the quantile estimates put beforehand on the log-scale; in the bottom right panels, the lengths reported are those of the confidence interval for $\log q(1-p)$.

C.3. About the difference between the test statistic Λ_n of equal tail indices and the test statistic of [14]

C.4. About Remark 8

C.5. About the asymptotic variance in Theorem 6

Appendix D: Finite-sample study - Further details and results

D.1. Simulation experiments

D.1.1. General setup: Pooling for tail index and extreme value inference

D.1.2. Distributed inference of extreme values

D.1.3. Pooling estimators using residuals from location-scale models

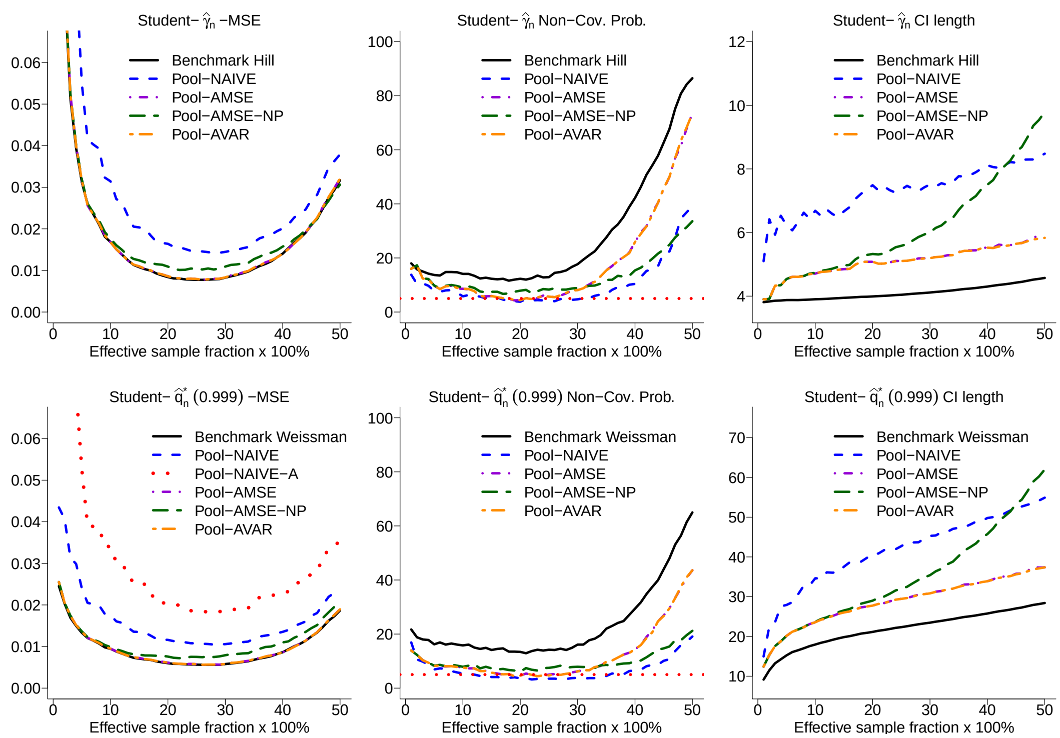

Figure D.3: As in Figure D.2, in the Multivariate Student model (b).

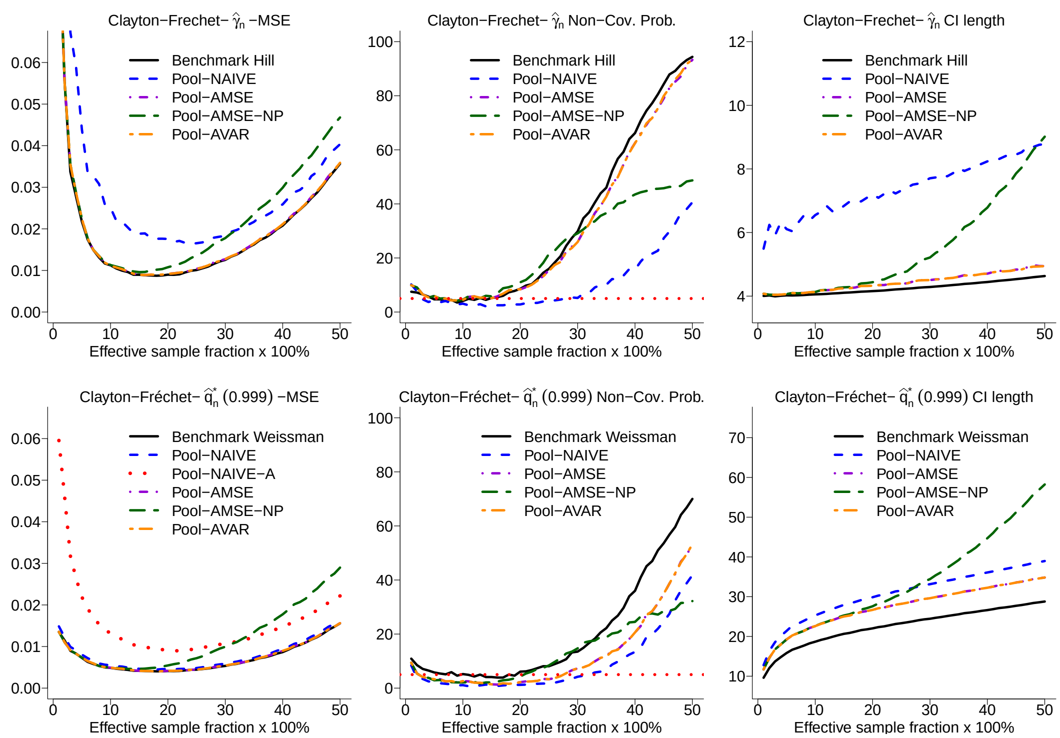

Figure D.4: As in Figure D.2, in the Clayton-Fréchet model (c).

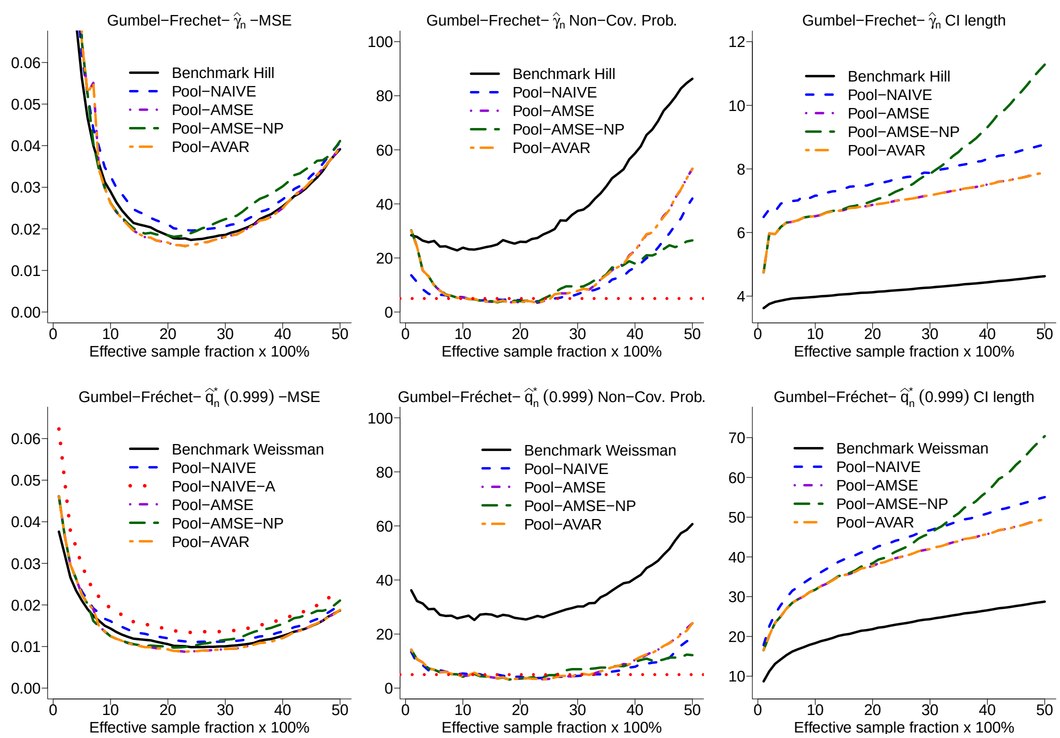

Figure D.5: As in Figure D.2, in the Gumbel-Fréchet model (d).

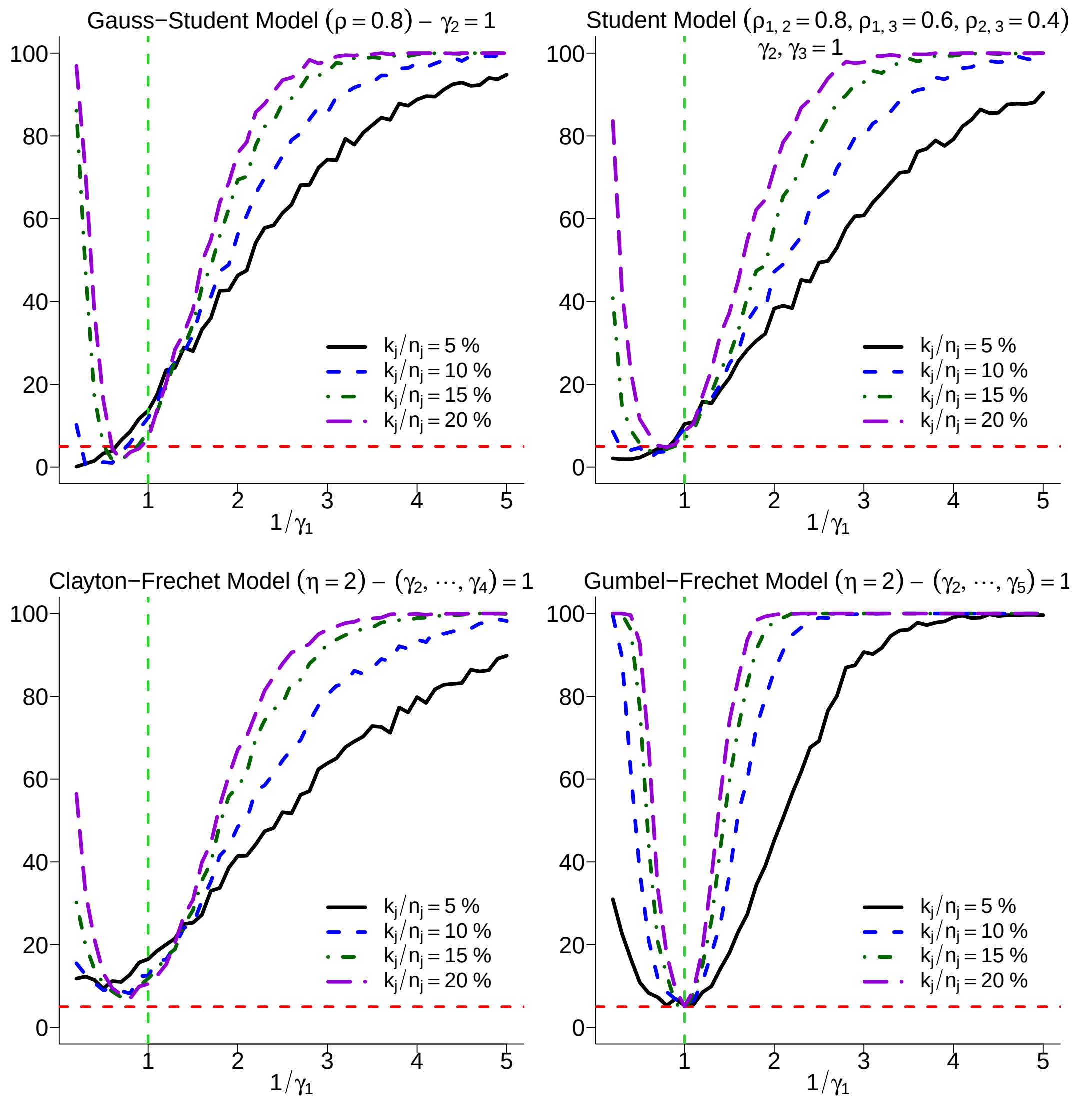

Figure D.6: Simulation results, general pooling setting, rejection rate of the test of tail homogeneity based on Λ_n with nominal type I error equal to 5%. Top left panel: Gaussian-Student model (a'), top right panel: Multivariate Student model (b'), bottom left panel: Clayton-Fréchet model (c'), bottom right panel: Gumbel-Fréchet model (d'), where the value of the tail index γ_1 in the first marginal is allowed to vary in the interval $[0.2, 5]$. The red horizontal dashed line represents the 5% nominal rejection rate under the null hypothesis, and the green vertical dashed line represents the value $\gamma_1 = 1$ under which the null hypothesis of tail homogeneity is satisfied. All results are represented as functions of $1/\gamma_1$, with the common effective sample fraction k_j/n_j used in each marginal indicated with a color code in the bottom right corner of each panel.

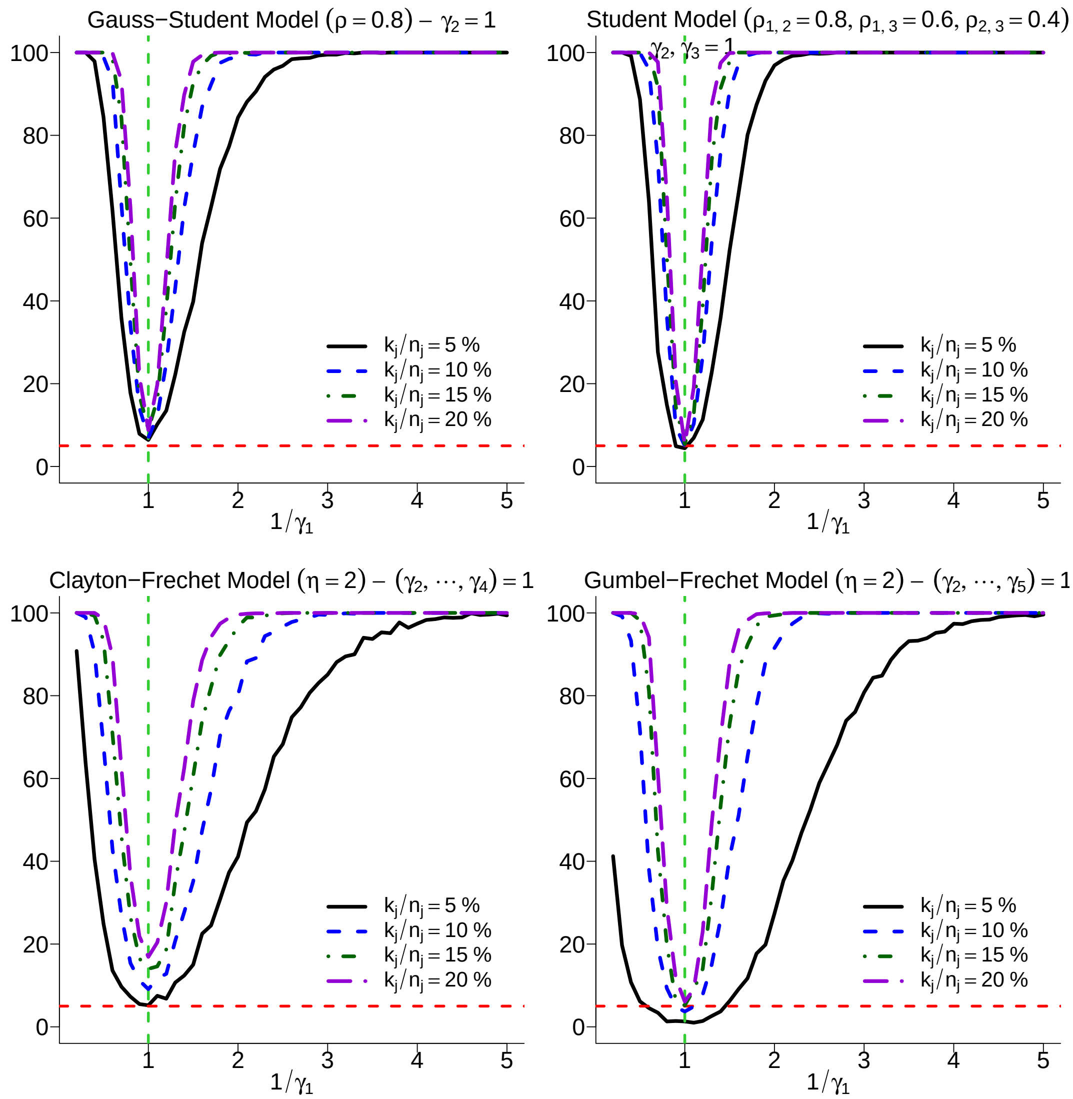

Figure D.7: As in Figure D.6, with equal subsample sizes in each model, namely $(n_1, n_2) = (500, 500)$ in model (a'), $(n_1, n_2, n_3) = (333, 333, 334)$ in model (b'), $(n_1, n_2, n_3, n_4) = (250, 250, 250, 250)$ in model (c') and $(n_1, n_2, n_3, n_4, n_5) = (200, 200, 200, 200, 200)$ in model (d').

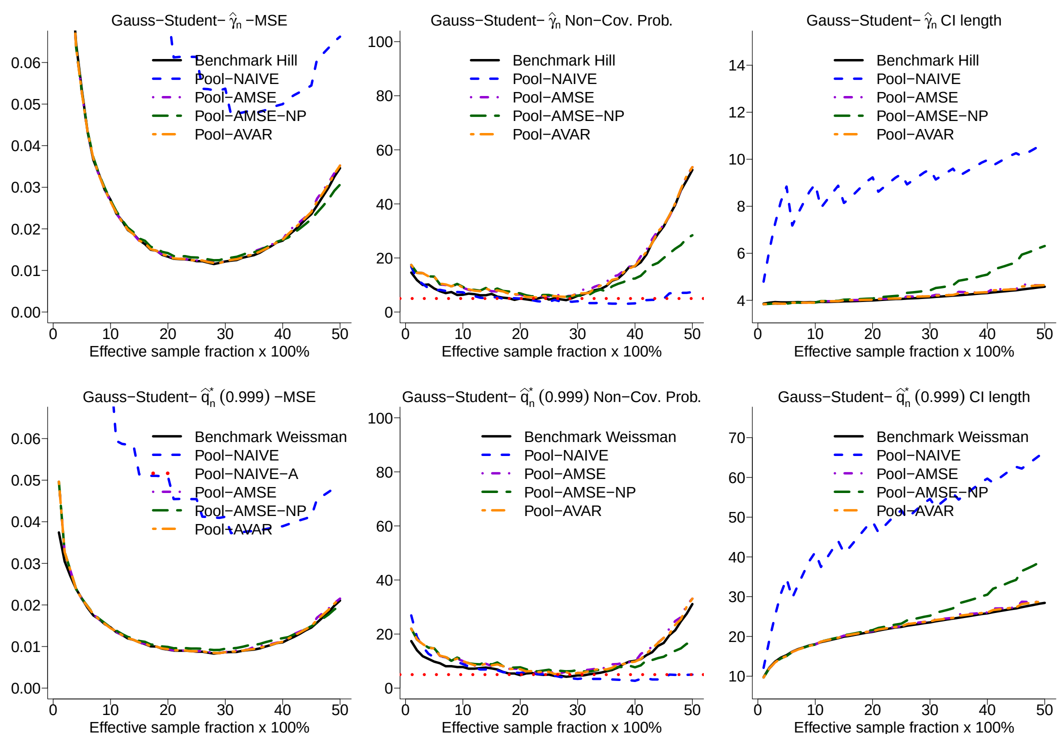

Figure D.8: As in Figure D.2, in the Gaussian-Student model (a), with $(n_1, n_2) = (20, 380)$.

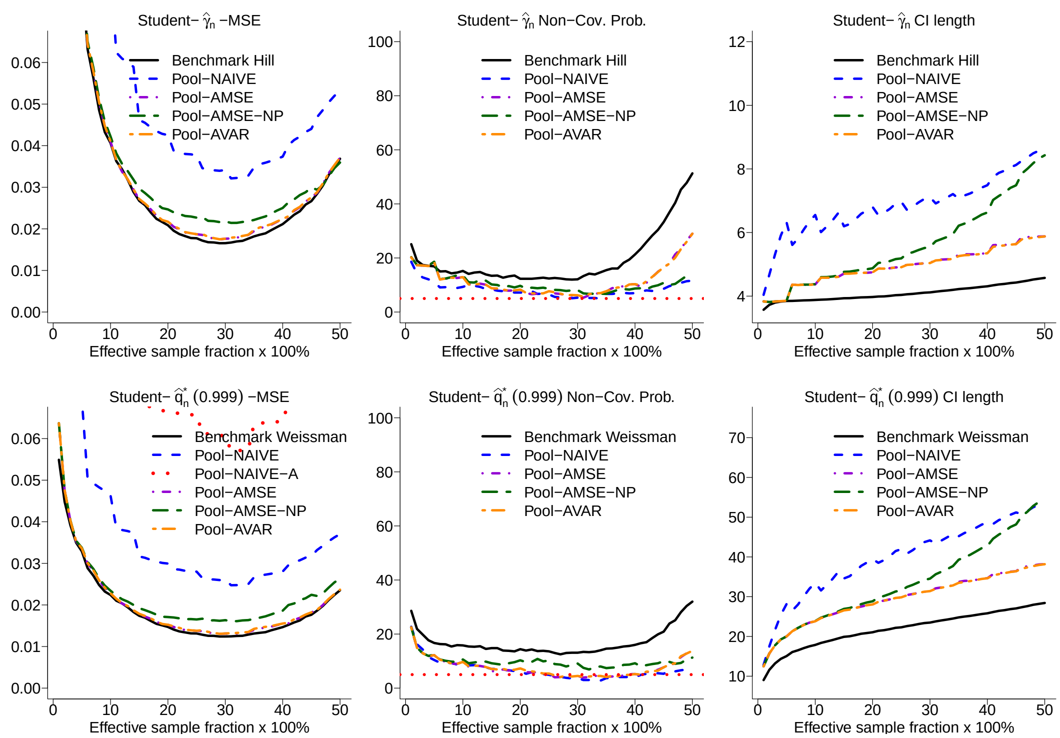

Figure D.9: As in Figure D.2, in the Multivariate Student model (b), with $(n_1, n_2, n_3) = (20, 180, 200)$.

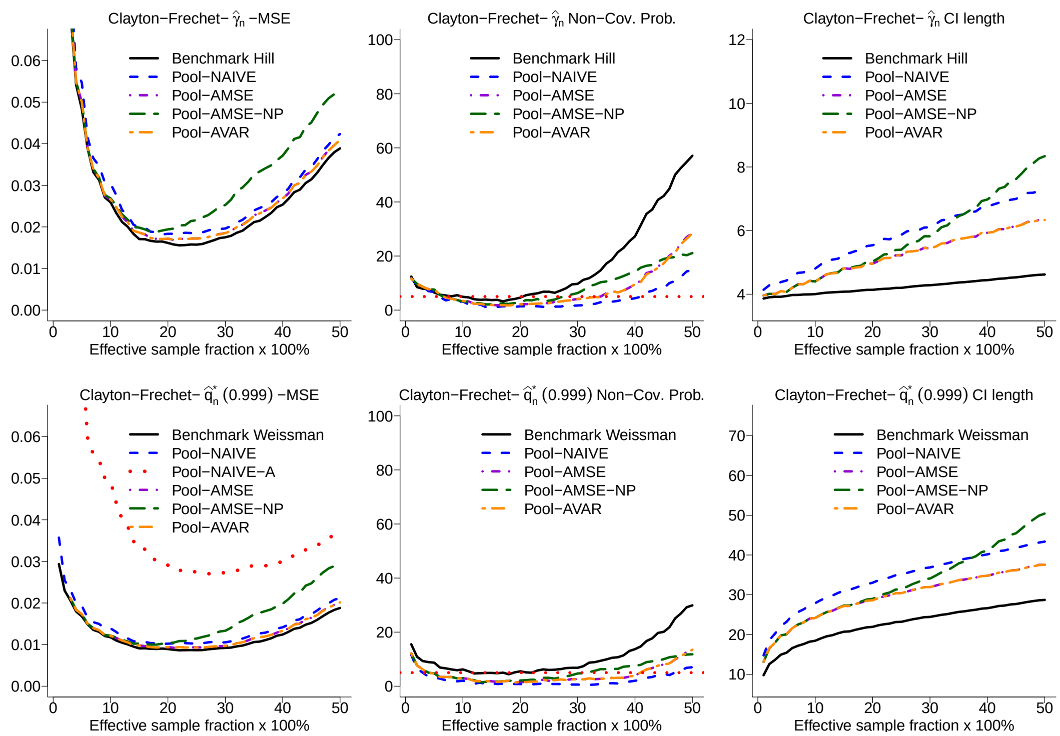

Figure D.10: As in Figure D.2, in the Clayton-Fréchet model (c), with $(n_1, n_2, n_3, n_4) = (60, 80, 100, 160)$.

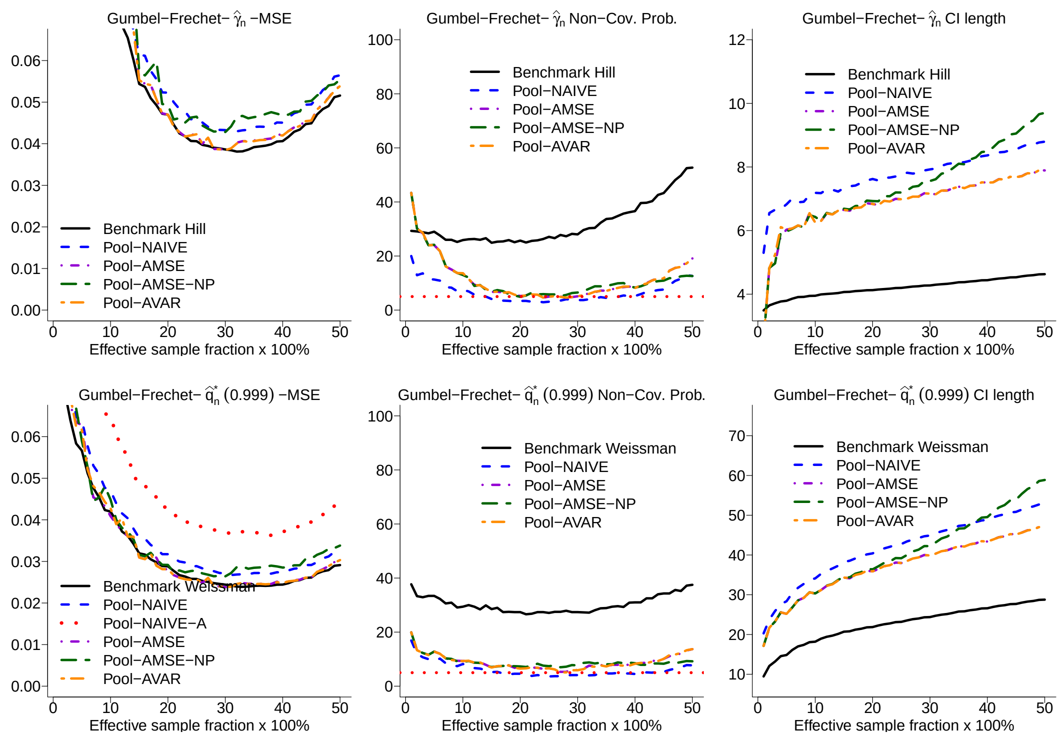

Figure D.11: As in Figure D.2, in the Gumbel-Fréchet model (d), with $(n_1, n_2, n_3, n_4, n_5) = (60, 60, 80, 80, 120)$.

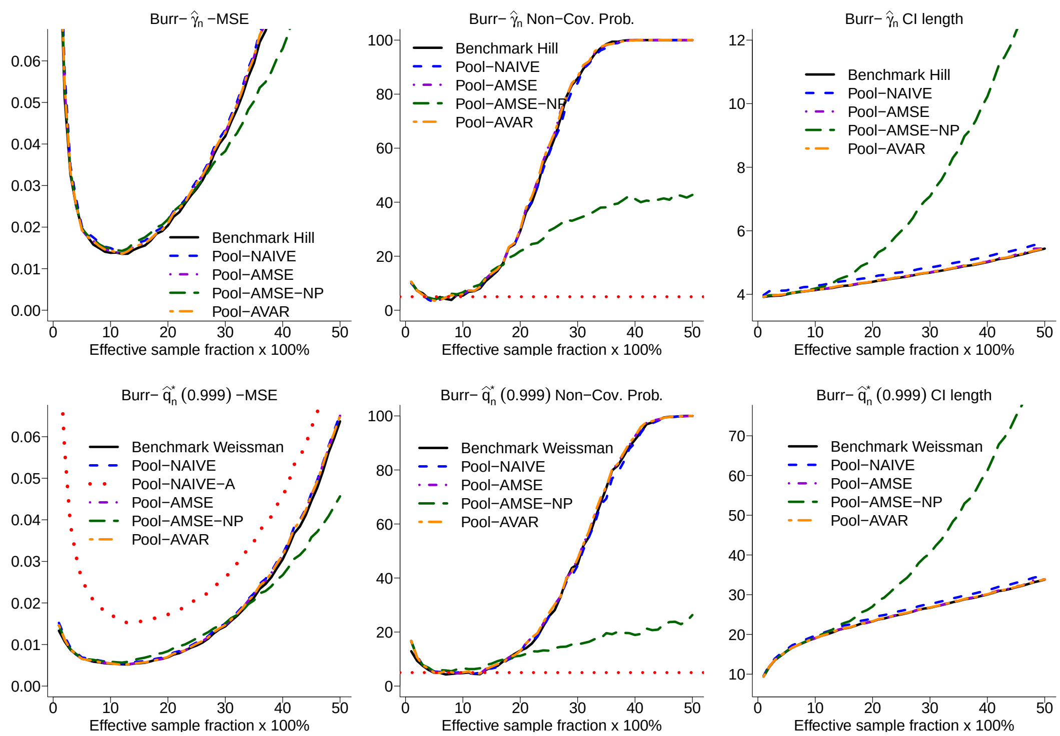

Figure D.12: Simulation results, distributed inference context, Burr model (e). Top row: tail index estimation; bottom row: extreme quantile estimation at level $1 - p = 0.999$. Left panels: MSE of the point estimators; middle panels: non-coverage probability of the asymptotic confidence intervals, where the red horizontal dotted line represents the 5% nominal non-coverage probability; right panels: average length of the 95% confidence intervals multiplied by $\sqrt{k}$. All results are represented as functions of the sample fraction k_j/n_j (identical for each j). In the bottom left panels, the MSEs represented are the relative MSEs of the quantile estimates put beforehand on the log-scale; in the bottom right panels, the lengths reported are those of the confidence interval for $\log q(1 - p)$.

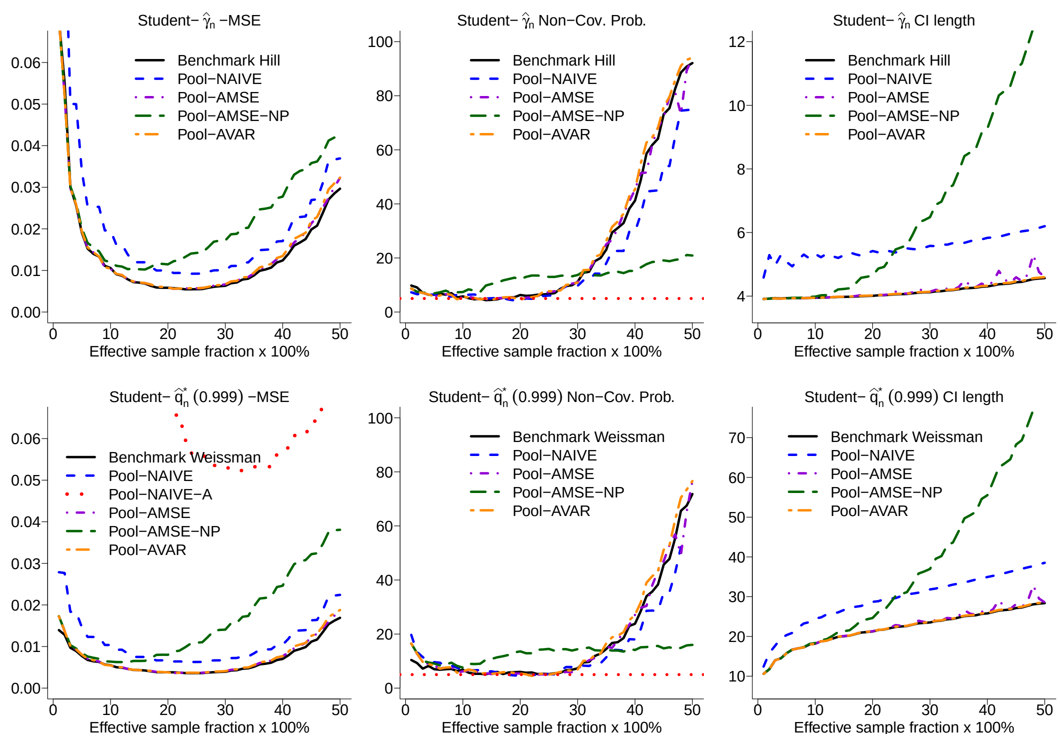

Figure D.13: As in Figure D.12, in the Student model (f).

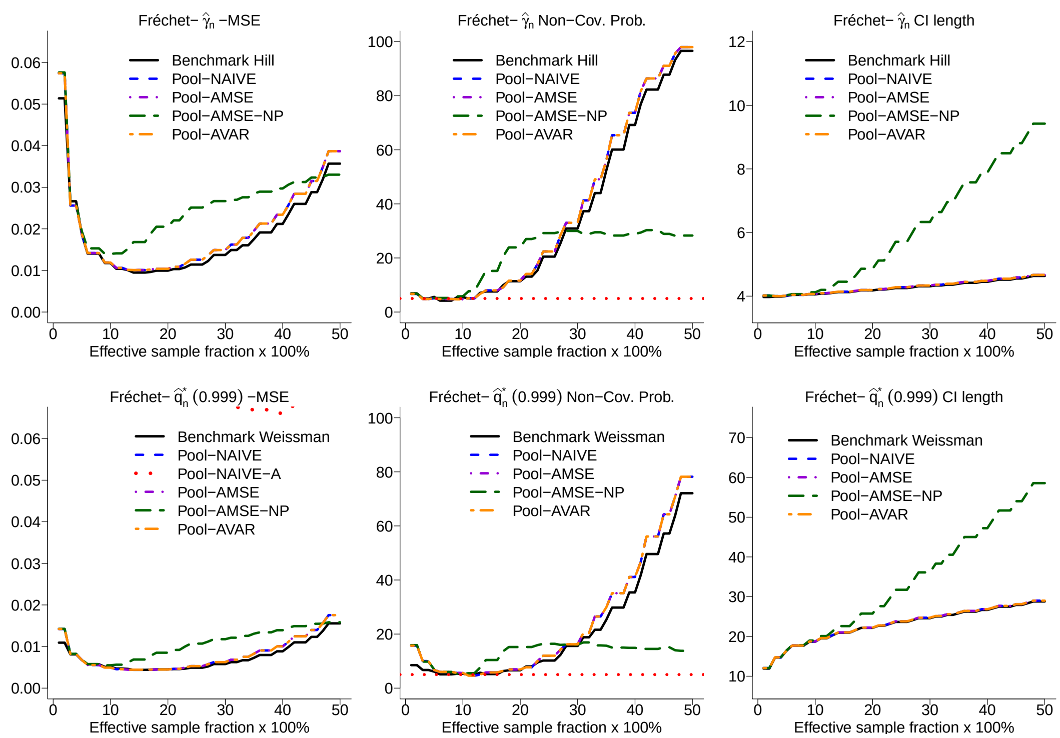

Figure D.14: As in Figure D.12, in the Fréchet model (g).

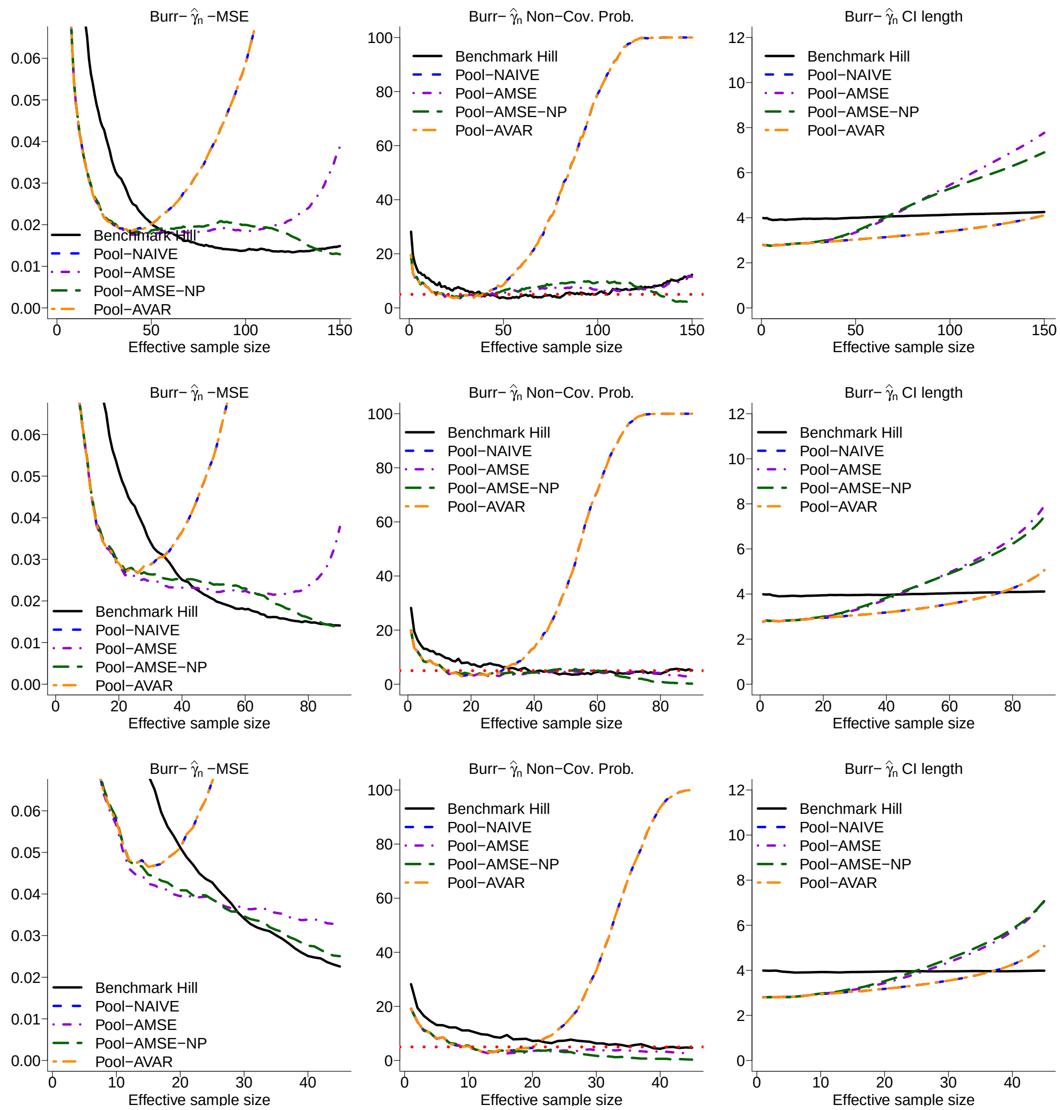

Figure D.15: Simulation results, distributed inference context, Burr model (h). Left panels: MSE of the tail index point estimators, middle panels: non-coverage probability of the asymptotic confidence intervals, right panels: average length of the 95% confidence intervals multiplied by $\sqrt{k}$. In the middle panels, the red horizontal dotted line represents the 5% nominal non-coverage probability. Top row: $(n_1, n_2) = (200, 800)$, middle row: $(n_1, n_2) = (100, 900)$, bottom row: $(n_1, n_2) = (50, 950)$. All results are represented as functions of the effective sample size $k_1 = k_2$.

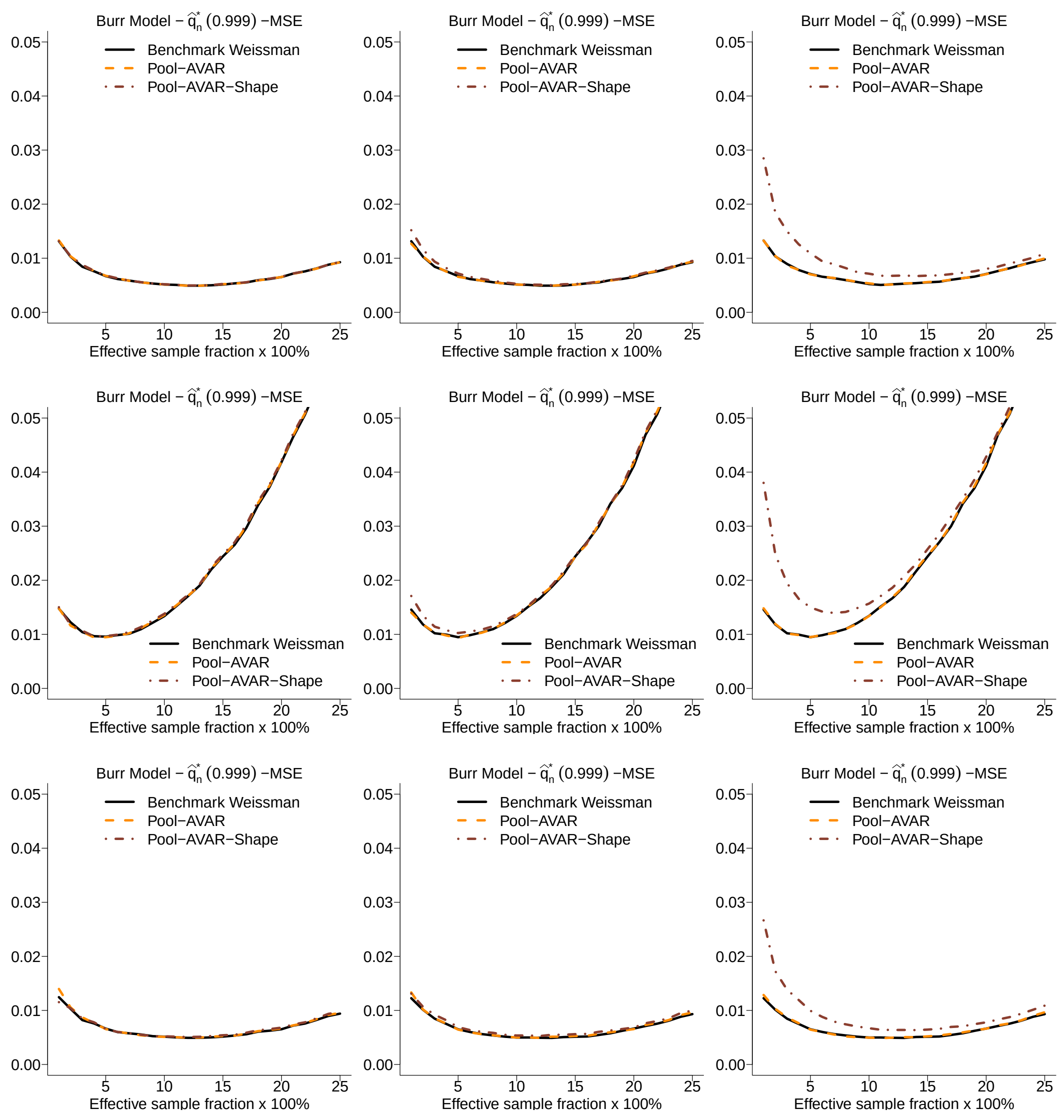

Figure D.16: Simulation results, distributed inference context, extreme quantile estimation at level $1-p=0.999$ in the Burr models (Q-a), (Q-b), (Q-c). Top row, model (Q-a), from left to right: $n_1=900$, $n_1=500$, $n_1=100$; middle row, model (Q-b), from left to right: $n_1=900$, $n_1=500$, $n_1=100$; bottom row, model (Q-c), from left to right: $n_1=800$, $n_1=500$, $n_1=100$. All results are represented as functions of the sample fraction $k/n=k_j/n_j$ (identical for each j). In each panel, the MSEs represented are the relative MSEs of the quantile estimates put beforehand on the log-scale.

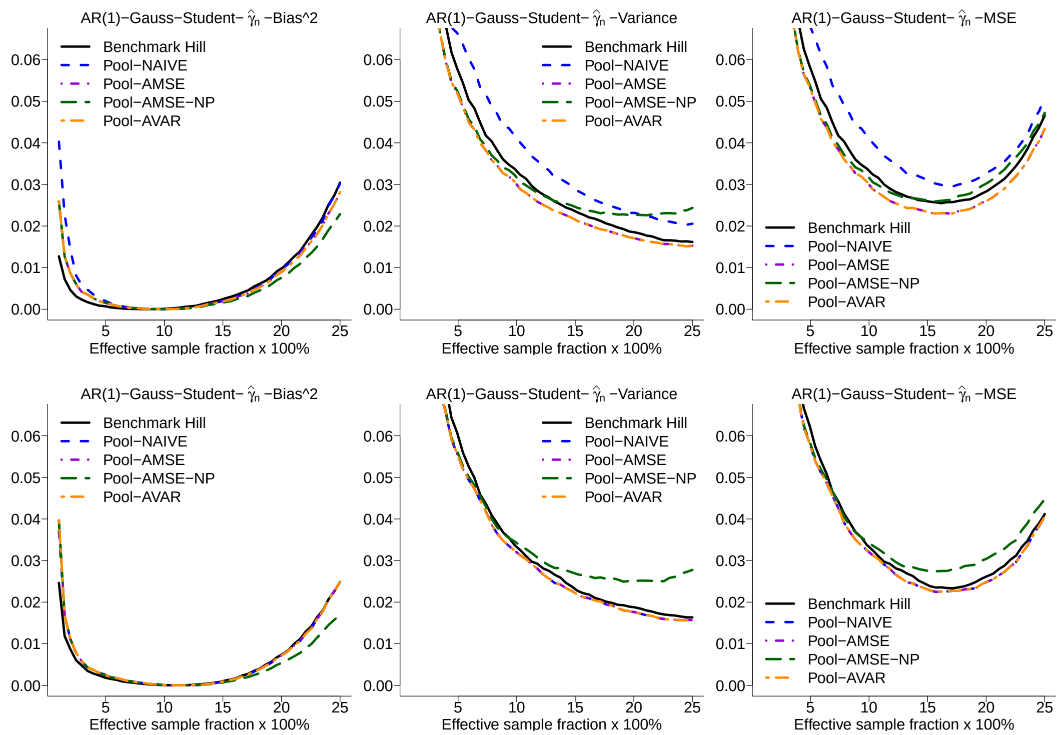

Figure D.17: Simulation results, pooled estimators using residuals, Gaussian-Student innovations model (AR-1-a). All panels relate to tail index estimation. Left panels: squared bias, middle panels: variance, right panels: MSE. All results are represented as functions of the sample fraction k_j/n_j (identical for each j). Top row: $(n_1, n_2) = (250, 750)$, bottom row: balanced sample sizes.

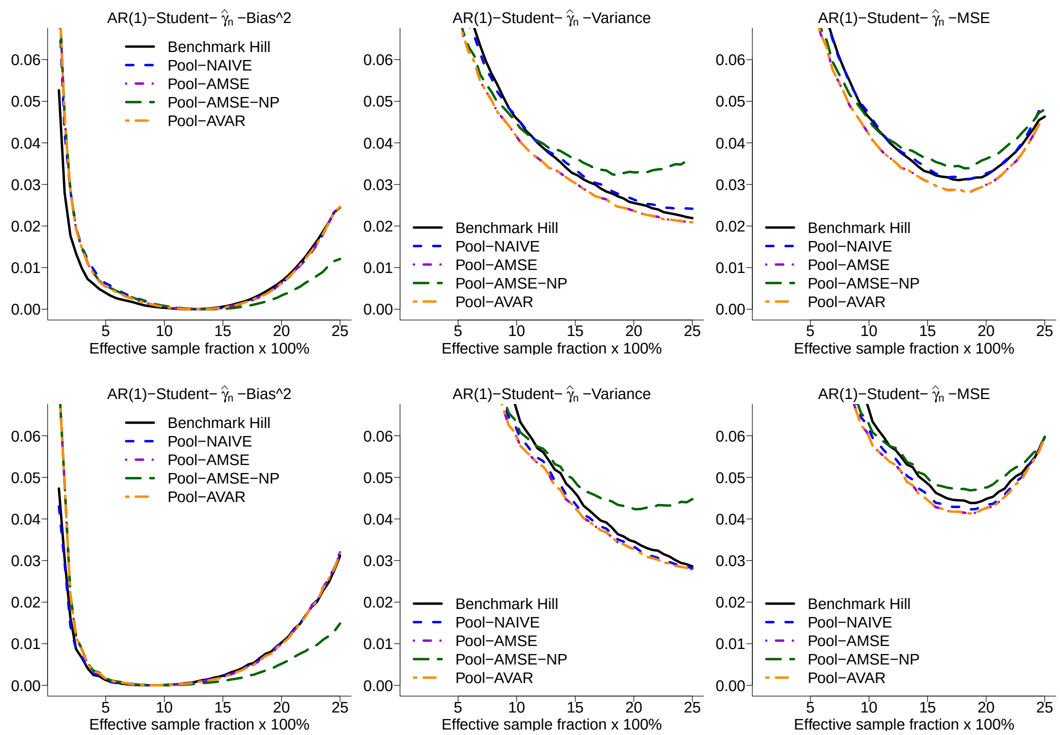

Figure D.18: Simulation results, pooled estimators using residuals, Multivariate-Student innovations model (AR-1-b). All panels relate to tail index estimation. Left panels: squared bias, middle panels: variance, right panels: MSE. All results are represented as functions of the sample fraction k_j/n_j (identical for each j). Top row: $(n_1, n_2) = (200, 400, 400)$, bottom row: balanced sample sizes.

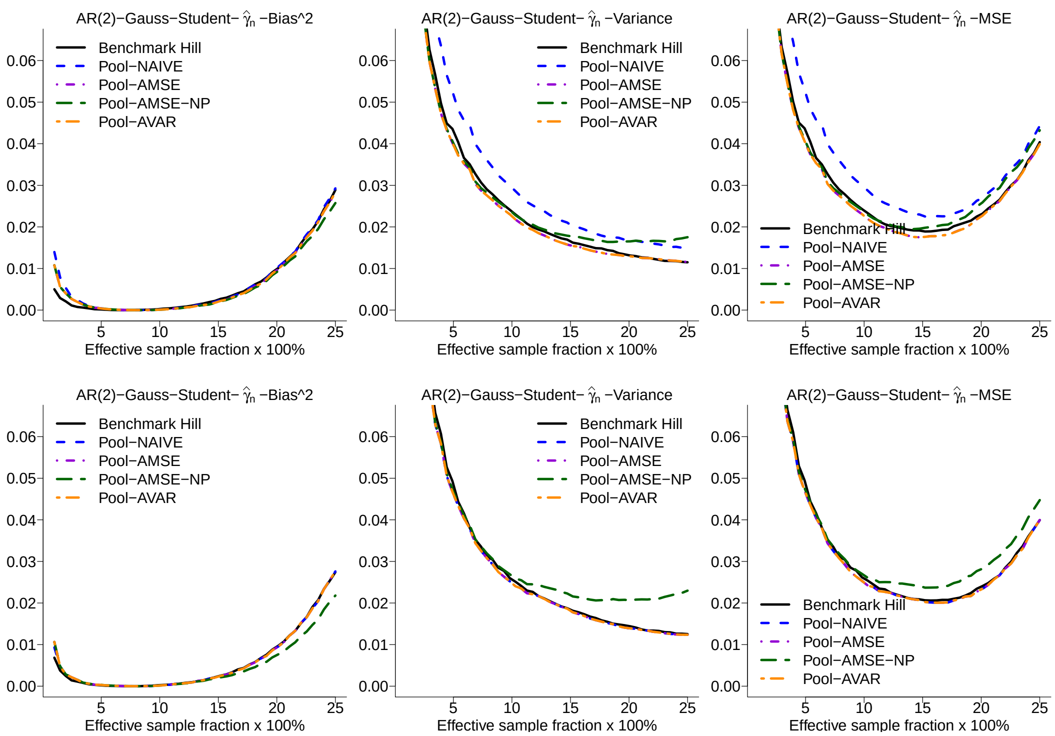

Figure D.19: As in Figure D.17, in the Gaussian-Student innovations model (AR-2-a) with misspecified AR(1) dynamics.

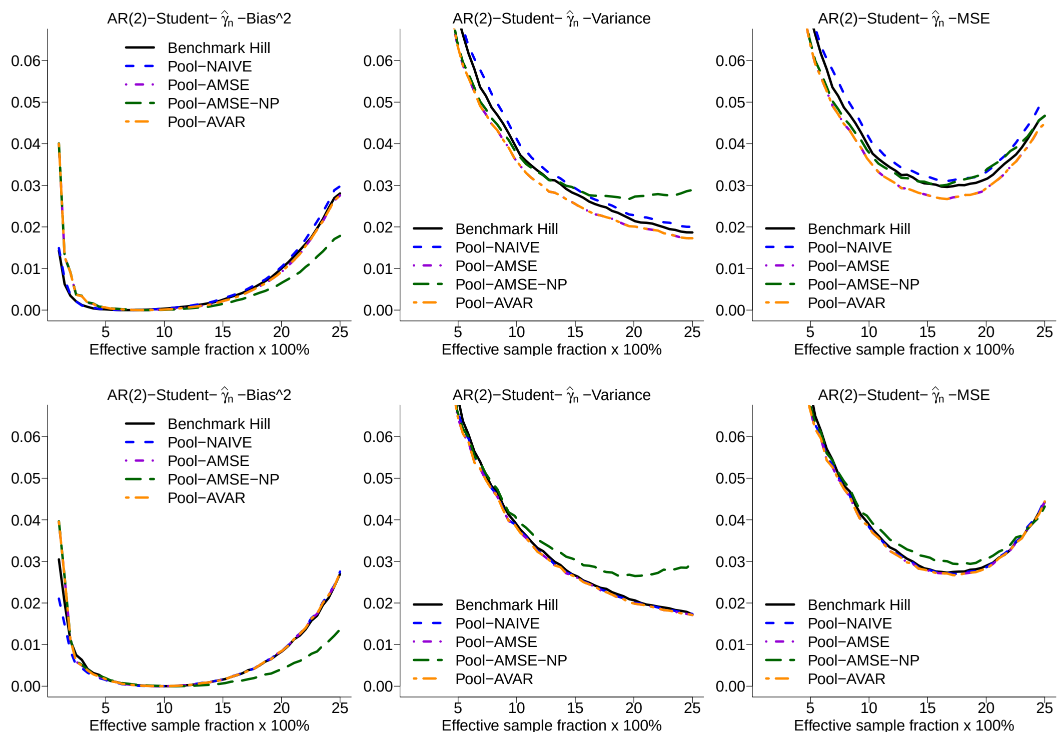

Figure D.20: As in Figure D.18, in the Multivariate-Student innovations model (AR-2-b) with misspecified AR(1) dynamics.

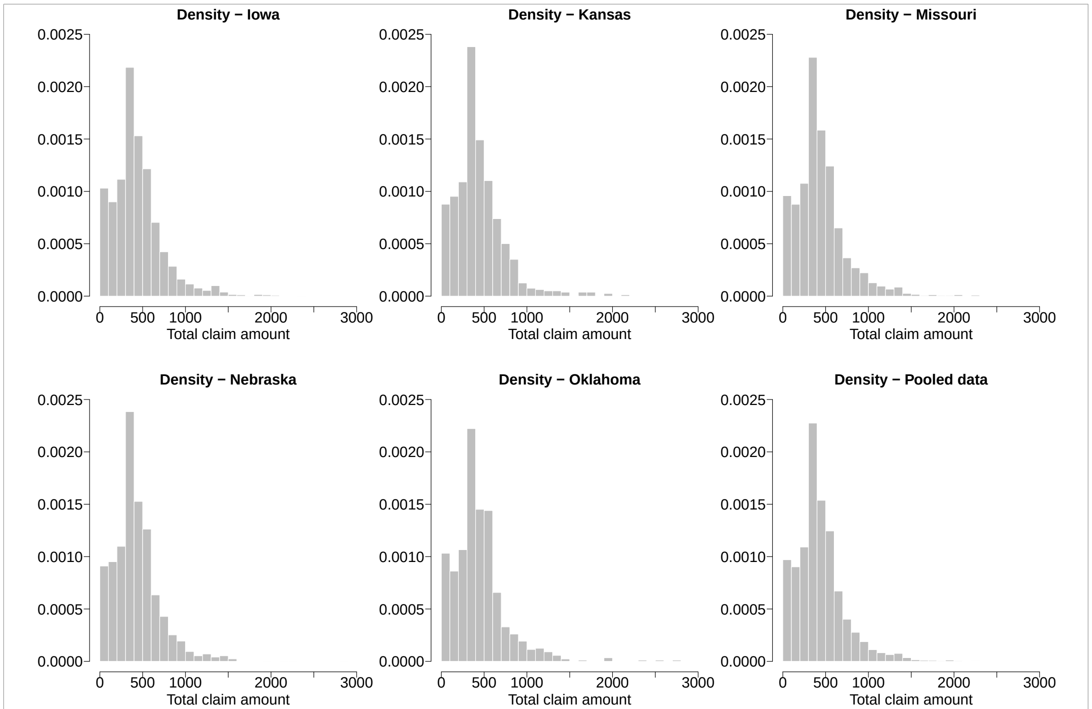

Figure D.21: Car insurance data: Histograms of total claim amounts.

ID	County	Data type	n_j	k_j	$\widehat{\gamma}_j$ [95% CI]
110	Santa Rosa	Raw	226	38	0.344 [0.234, 0.452]
140	Jackson	Raw	225	33	0.330 [0.220, 0.449]
170	Suwanee	Raw	225	31	0.329 [0.212, 0.442]
180	Baker	Residuals	225	15	0.494 [0.244, 0.744]
240	Putnam	Residuals	244	14	0.438 [0.209, 0.668]
290	Volusia	Residuals	278	29	0.442 [0.281, 0.604]
302	Lake	Residuals	281	15	0.514 [0.254, 0.774]
340	Osceola	Residuals	221	14	0.518 [0.247, 0.789]

Table D.1. Florida rainfall data: Information gathered at each individual station. The estimates and confidence intervals reported in the last column correspond to the selected k_j values indicated by the vertical blue lines in Figure D.24.

D.2. Data analysis

D.2.1. Distributed inference for car insurance data

D.2.2. Pooling for regional inference on extreme rainfall

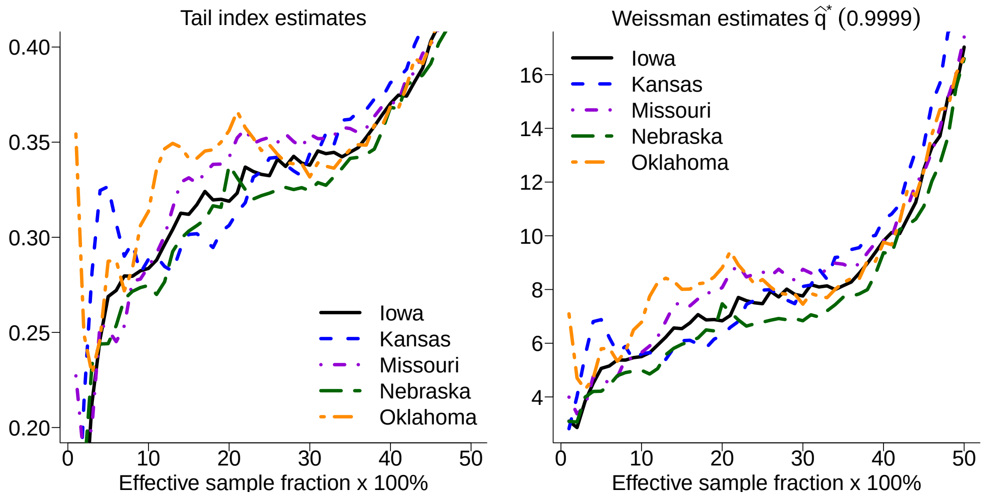

Figure D.22: Car insurance data: Hill estimates $\widehat{\gamma}_j(k_j)$ (left) and Weissman estimates $\widehat{q}_j^{\star}(0.9999|k_j)$ (right) for each state, as functions of the sample fraction k_j/n_j, assumed to be identical in each subsample.

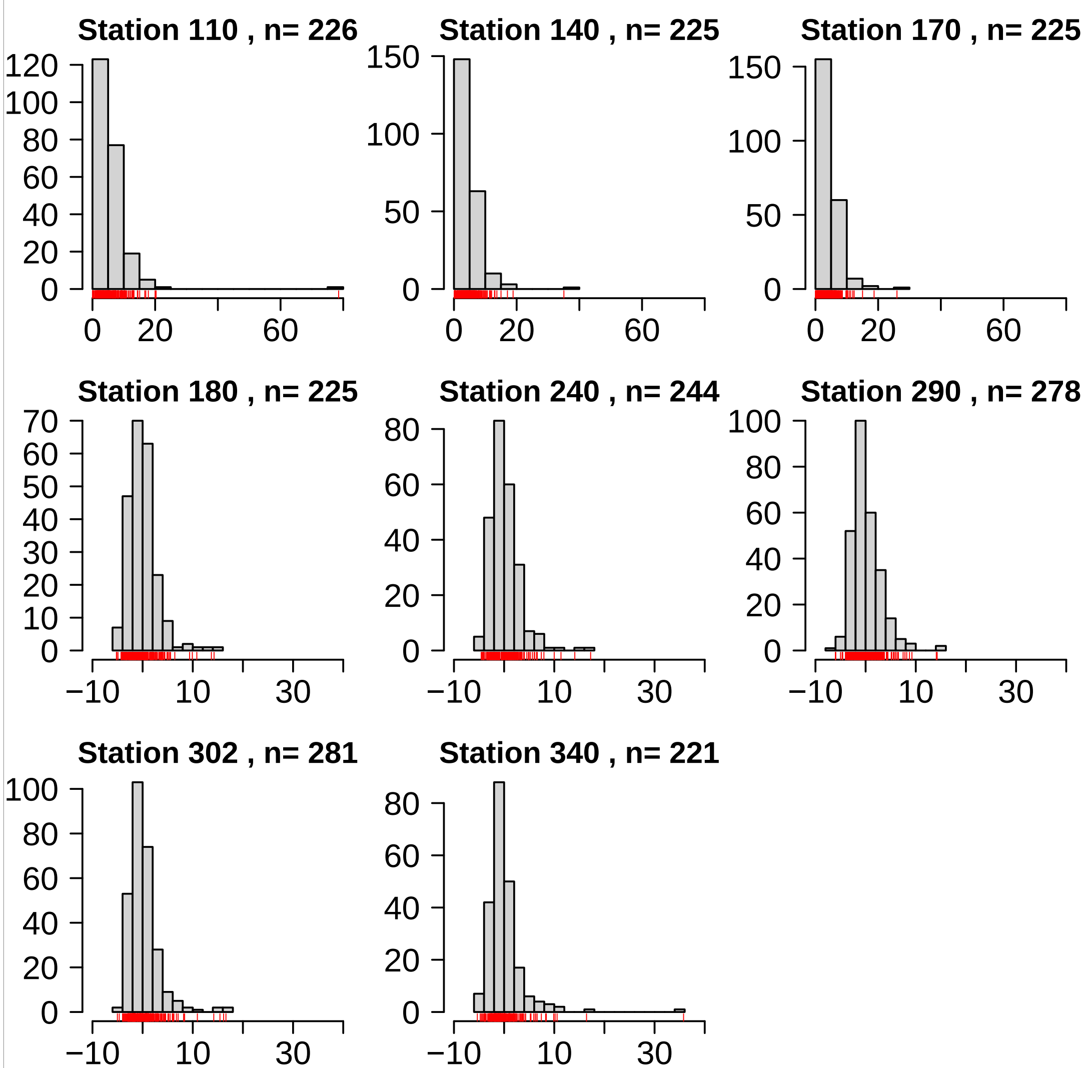

Figure D.23: Florida rainfall data: Histograms of the raw data for the three stations in the top panel (red cluster), and for the residuals obtained from the fitted SARMA models of the five remaining stations (green cluster).

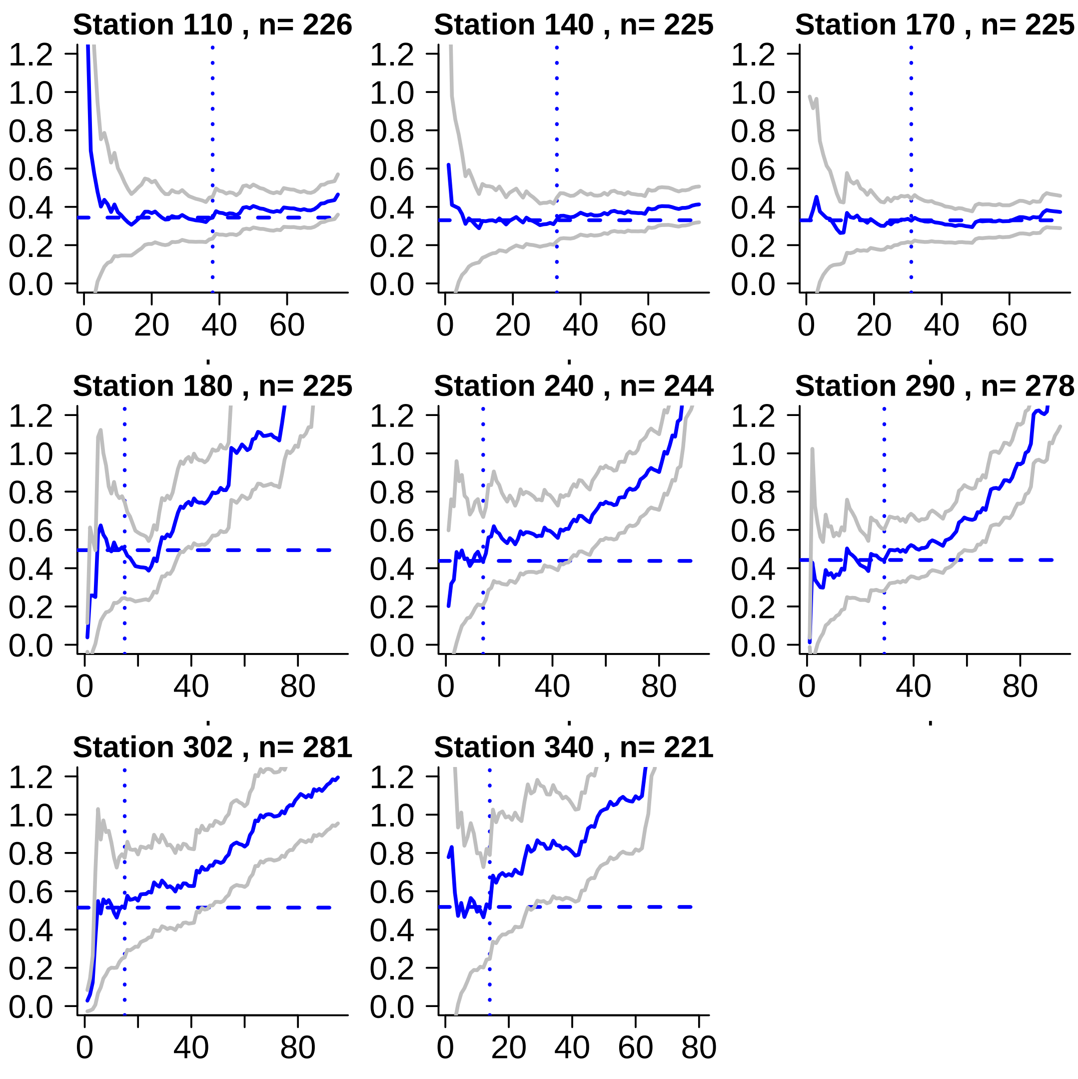

Figure D.24: Florida rainfall data: Individual tail index estimators. For each station j, we represent the Hill estimate $\widehat{\gamma}_j(k_j)$ (solid blue) and its associated asymptotic 95% confidence interval (solid gray), as functions of the effective sample size $k = k_j$. The corresponding Hill point estimate is represented by the horizontal dashed blue line.