1. Introduction

2. Pooling extreme value estimators

2.1. Pooled Hill estimators of the tail index

Theorem 1.

2.2. Optimal choices of weights

Corollary 1.

Remark 1 (On the variance-optimal choice).

Remark 2 (On the AMSE-optimal choice).

Remark 3 (On optimal choices of the k_j in pooled estimators).

Corollary 2.

2.3. Weighted geometric pooling of extreme quantile estimators

Theorem 2.

2.4. Inference using pooled extreme value estimators

Corollary 3.

Remark 4 (Identification of tail homogeneous subgroups).

Remark 5 (With asymptotic independence across subsamples).

Remark 6 (Inference and bias correction).

Remark 7 (Tail homogeneity and tail homoskedasticity).

Corollary 4.

3. The framework of distributed inference

3.1. Distributed estimation of the tail index

Corollary 5.

3.2. Variance-optimal and AMSE-optimal combinations

Corollary 6.

Remark 8 (Asymptotic bias comparison).

Theorem 3.

Theorem 4.

Corollary 7.

3.3. Extreme quantile estimation

Corollary 8.

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

Theorem 5.

3.5. The case of a large number of machines

Theorem 6.

Remark 9 (Comparison with earlier results for fixed m).

Remark 10 (Comparison with earlier results about tail index estimation).

Corollary 9.

Theorem 7.

4. Filtering to handle dependence and covariates

Theorem 8.

Remark 11 (Linking the regression and unconditional settings).

Remark 12 (On the importance of filtering without pooling residuals).

Theorem 9.

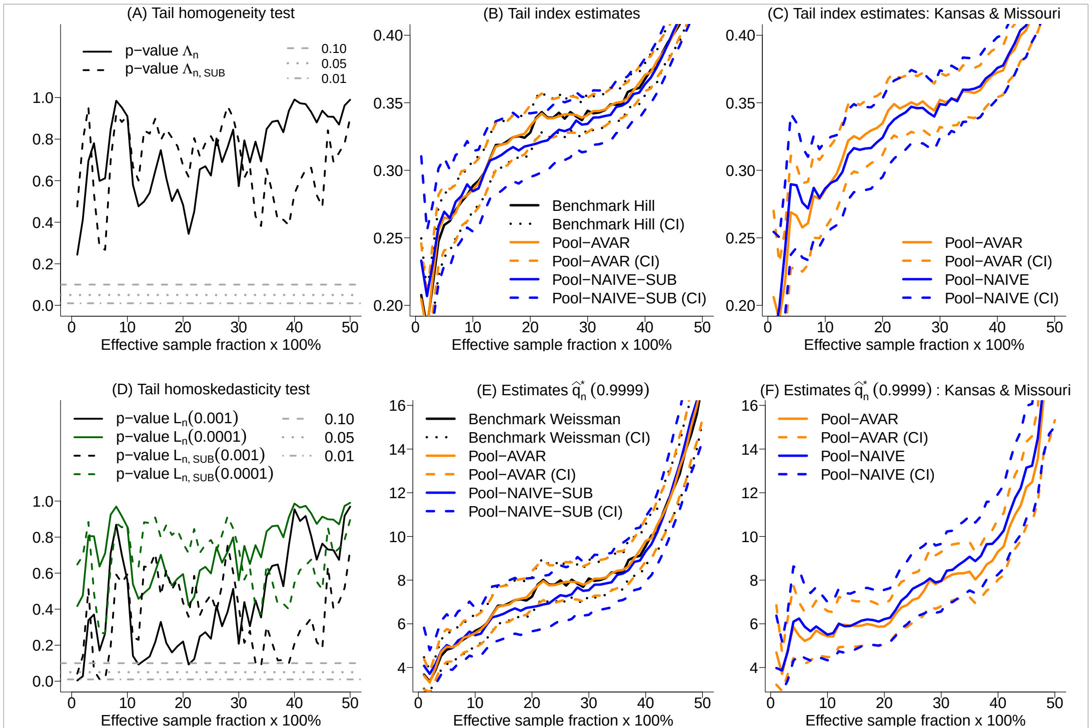

Figure 1: Car insurance data. In (A) and (D), $\Lambda_{n,\mathrm{SUB}}$ and $L_{n,\mathrm{SUB}}(p)$ denote the test statistics Λ_n and $L_n(p)$ calculated from the subsampled data of total size $5 \times 700 = 3{,}500$. In (B) and (C), Pool-AVAR and Pool-NAIVE respectively denote the variance-optimal 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 $\widehat{q}_n^{\star}(1-p|\widehat{\omega}^{(\mathrm{Var})})$ 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.

5. Finite-sample study

5.1. Simulation experiments

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

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

5.1.3. Distributed inference of extreme values

5.2. Data analysis

5.2.1. Distributed inference for car insurance data

5.2.2. Pooling for regional inference on extreme rainfall

6. Discussion

Acknowledgments

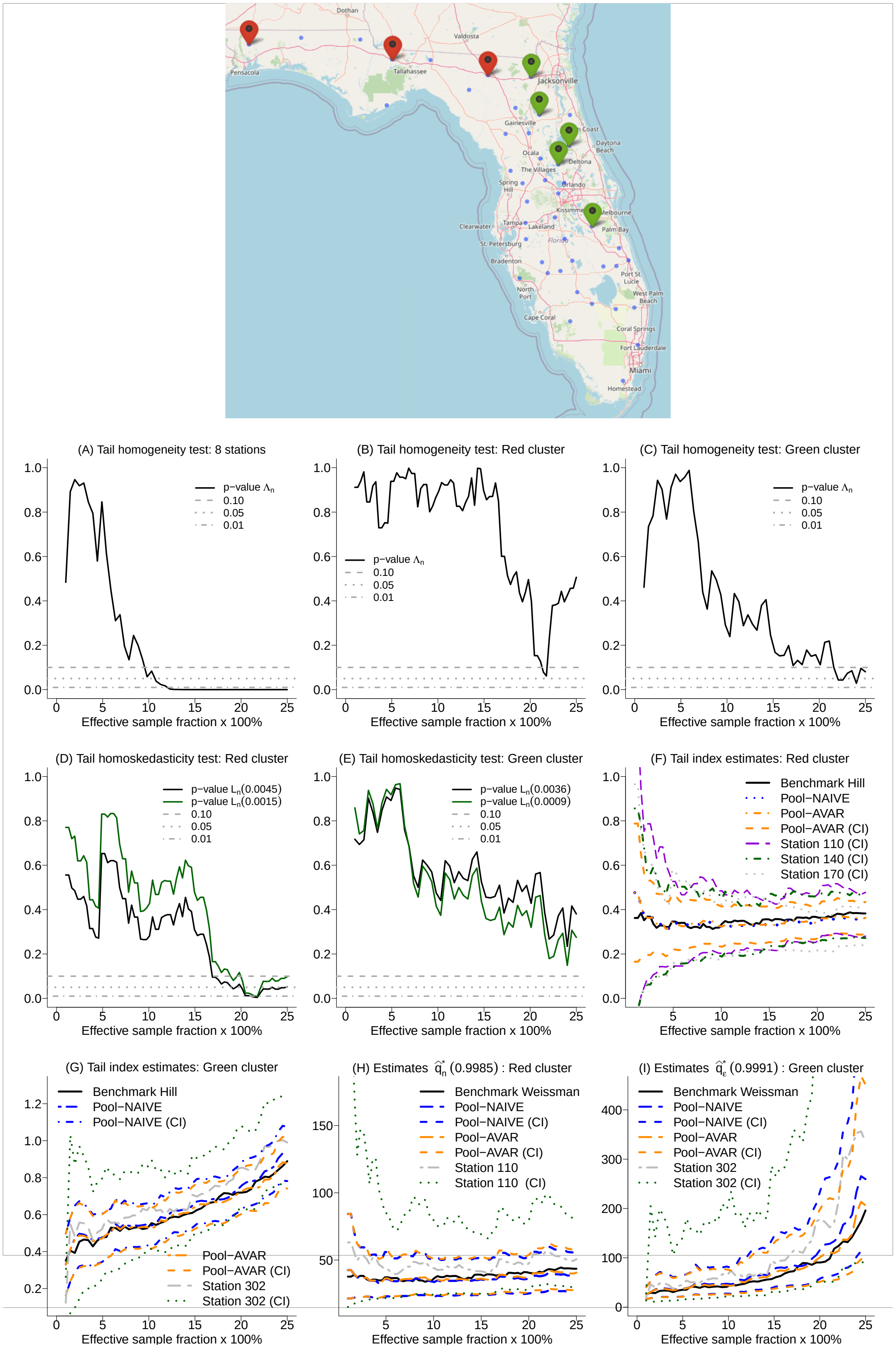

Figure 2: Florida rainfall data. Map of Florida along with its gauge stations and inferential results (notation as in Figure 1).

Supplement to: 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 supplementary material document contains the statement, proof and discussion of a general pooling theorem used as part of the proof of our main results. It also contains the proofs of all theoretical results in the main paper, preceded by auxiliary results and their proofs, in addition to extended remarks, a full description of our simulation study and its results, as well as extra results related to our real data analyses.

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

Appendix A: General pooling theory

Assume that a vector of m estimators $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_{1,n}, \ldots, \widehat{\theta}_{m,n})^\top$ of a common parameter θ is available from n observations. Suppose there exist nonrandom sequences $v_{1,n}, \ldots, v_{m,n} \to \infty$ such that the $m-$dimensional vector $(\sqrt{v_{1,n}}(\widehat{\theta}_{1,n} - \theta), \ldots, \sqrt{v_{m,n}}(\widehat{\theta}_{m,n} - \theta))^\top$ is asymptotically Gaussian with bias vector $\boldsymbol{B}$ and covariance matrix $\mathbf{V}$. We consider the problem of finding an optimal way to pool the estimators $\widehat{\theta}_{j,n}$. In other words, we consider the estimator

$$\widehat{\theta}_n(\boldsymbol{\omega}) = \widehat{\theta}_n(\omega_1, \ldots, \omega_m) = \sum_{j=1}^{m} \omega_j \widehat{\theta}_{j,n} = \boldsymbol{\omega}^\top \widehat{\boldsymbol{\theta}}_n$$

where the weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^\top$ summing up to 1 remain to be selected. This class of estimators contains convex combinations, but also allows for weight vectors with negative elements. There are two obvious criteria for optimality of the weights, either by minimizing the asymptotic variance of $\widehat{\theta}_n(\boldsymbol{\omega})$ or its Asymptotic Mean Squared Error. The following theorem provides the optimality conditions for each criterion, along with the closed form expressions of the optimal weights.

Theorem A.1. *Assume that there exist nonrandom sequences $v_{1,n}, \ldots, v_{m,n} \to \infty$, with $v_{1,n}/v_{j,n} \to c_j \in (0, \infty)$ for any j, such that*

$$(\sqrt{v_{1,n}}(\widehat{\theta}_{1,n} - \theta), \ldots, \sqrt{v_{m,n}}(\widehat{\theta}_{m,n} - \theta))^\top \xrightarrow{d} \mathcal{N}(\boldsymbol{B}, \mathbf{V})$$

where $\boldsymbol{B} \in \mathbb{R}^m$ and $\mathbf{V}$ is an $m \times m$ symmetric positive semidefinite matrix. Set $v_n = \sum_{j=1}^m v_{j,n}$. Then, for any choice of weights $\boldsymbol{\omega}$ such that $\boldsymbol{\omega}^\top \mathbf{1} = \sum_{j=1}^m \omega_j = 1$, one has

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

where $\boldsymbol{B}_{\boldsymbol{c}} = \mathbf{D}_{\boldsymbol{c}}^{1/2}\boldsymbol{B}$ *and* $\mathbf{V}_{\boldsymbol{c}} = \mathbf{D}_{\boldsymbol{c}}^{1/2}\mathbf{V}\mathbf{D}_{\boldsymbol{c}}^{1/2}$ *with* $\mathbf{D}_{\boldsymbol{c}} = (\sum_{i=1}^m c_i^{-1})\,\mathrm{diag}(c_1,\ldots,c_m)$*. The matrix* $\mathbf{V}$ *is positive definite if and only if* $\mathbf{V}_{\boldsymbol{c}}$ *is so, and then we have the following results:*

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

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

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

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

and then

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

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

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

Finally, if $\widehat{\boldsymbol{\omega}}_n^\top\mathbf{1} = 1$ *with* $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$*, then* $\sqrt{v_n}(\widehat{\theta}_n(\widehat{\boldsymbol{\omega}}_n) - \widehat{\theta}_n(\boldsymbol{\omega})) = \mathrm{o}_{\mathbb{P}}(1)$*, and especially*

$$\sqrt{v_n}(\widehat{\theta}_n(\widehat{\boldsymbol{\omega}}_n) - \theta) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\omega}^\top\boldsymbol{B}_{\boldsymbol{c}}, \boldsymbol{\omega}^\top\mathbf{V}_{\boldsymbol{c}}\boldsymbol{\omega}\right).$$

Proof of Theorem A.1. Clearly

$$\sqrt{v_n}(\widehat{\boldsymbol{\theta}}_n - \theta\mathbf{1})^\top = \mathrm{diag}(\sqrt{v_n/v_{1,n}},\ldots,\sqrt{v_n/v_{m,n}})(\sqrt{v_{1,n}}(\widehat{\theta}_{1,n} - \theta),\ldots,\sqrt{v_{m,n}}(\widehat{\theta}_{m,n} - \theta))^\top.$$

Using the fact that $v_n/v_{j,n} = (v_{1,n}/v_{j,n}) \times (v_n/v_{1,n}) \to c_j\sum_{i=1}^m c_i^{-1}$ and the assumption on the joint convergence of the $\widehat{\theta}_{j,n}$, one finds

$$\sqrt{v_n}(\widehat{\boldsymbol{\theta}}_n - \theta\mathbf{1})^\top \xrightarrow{d} \mathcal{N}(\mathbf{D}_{\boldsymbol{c}}^{1/2}\boldsymbol{B}, \mathbf{D}_{\boldsymbol{c}}^{1/2}\mathbf{V}\mathbf{D}_{\boldsymbol{c}}^{1/2}) = \mathcal{N}(\boldsymbol{B}_{\boldsymbol{c}}, \mathbf{V}_{\boldsymbol{c}})$$

as required. The assertion on the limiting distribution of $\widehat{\theta}_n(\omega_1,\ldots,\omega_m)$ now immediately follows from writing $\widehat{\theta}_n(\boldsymbol{\omega}) - \theta = \boldsymbol{\omega}^\top(\widehat{\boldsymbol{\theta}}_n - \theta\mathbf{1})$. The final statements when $\boldsymbol{\omega}$ is replaced by $\widehat{\boldsymbol{\omega}}_n$ satisfying $\widehat{\boldsymbol{\omega}}_n^\top\mathbf{1} = 1$ and $\widehat{\boldsymbol{\omega}}_n^\top \xrightarrow{\mathbb{P}} \boldsymbol{\omega}$ are consequences of the identity

$$\sqrt{v_n}(\widehat{\theta}_n(\widehat{\boldsymbol{\omega}}_n) - \widehat{\theta}_n(\boldsymbol{\omega})) = \sqrt{v_n}(\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega})^\top(\widehat{\boldsymbol{\theta}}_n - \theta\mathbf{1})$$

and Slutsky's lemma.

We turn to solving the minimization problems. The fact that positive definiteness of $\mathbf{V}$ entails positive definiteness of $\mathbf{V}_{\boldsymbol{c}}$ immediately follows from the identity $\boldsymbol{x}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{x} = (\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})^\top \mathbf{V} (\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})$ and the fact that $\mathbf{D}_{\boldsymbol{c}}$ is nonsingular. This means in particular that the quadratic form $\boldsymbol{\omega} \mapsto \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}$ is strictly convex. A standard calculation involving the Lagrange multiplier method then shows that the solution of the problem

$$\min \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega} \text{ subject to } \boldsymbol{\omega}^\top \mathbf{1} = 1$$

satisfies $\mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega} = \mu \mathbf{1}$, where $\mu \in \mathbb{R}$ is a Lagrange multiplier. Taking into account the constraint $\boldsymbol{\omega}^\top \mathbf{1} = 1$ yields the solution

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

The asymptotic normality statement for $\boldsymbol{\omega} = \boldsymbol{\omega}^{(\mathrm{Var})}$ is obvious.

We now solve the AMSE minimization problem, that is,

$$\min\{(\boldsymbol{\omega}^\top \boldsymbol{B}_{\boldsymbol{c}})^2 + \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}\} \text{ subject to } \boldsymbol{\omega}^\top \mathbf{1} = 1.$$

Straightforward calculations show that the gradient of the strictly convex function $\boldsymbol{\omega} \mapsto (\boldsymbol{\omega}^\top \boldsymbol{B}_{\boldsymbol{c}})^2 + \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}$ is $2(\mathbf{V}_{\boldsymbol{c}} + \boldsymbol{B}_{\boldsymbol{c}} \boldsymbol{B}_{\boldsymbol{c}}^\top)\omega$. The optimal solution therefore satisfies $(\mathbf{V}_{\boldsymbol{c}} + \boldsymbol{B}_{\boldsymbol{c}} \boldsymbol{B}_{\boldsymbol{c}}^\top)\omega = \mu \mathbf{1}$ where $\mu \in \mathbb{R}$ is a Lagrange multiplier. Note now that

$$\mathbf{V}_{\boldsymbol{c}} + \boldsymbol{B}_{\boldsymbol{c}} \boldsymbol{B}_{\boldsymbol{c}}^\top = \mathbf{V}_{\boldsymbol{c}}^{1/2} (\mathrm{I}_m + \mathbf{V}_{\boldsymbol{c}}^{-1/2} \boldsymbol{B}_{\boldsymbol{c}} \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1/2}) \mathbf{V}_{\boldsymbol{c}}^{1/2}$$

where I_m is the identity matrix with m columns and $\mathbf{V}_{\boldsymbol{c}}^{-1/2}$ is the unique symmetric positive definite square root of $\mathbf{V}_{\boldsymbol{c}}^{-1}$. This means that $\mathbf{V}_{\boldsymbol{c}} + \boldsymbol{B}_{\boldsymbol{c}} \boldsymbol{B}_{\boldsymbol{c}}^\top$ is invertible, and

$$(\mathbf{V}_{\boldsymbol{c}} + \boldsymbol{B}_{\boldsymbol{c}} \boldsymbol{B}_{\boldsymbol{c}}^\top)^{-1} = \mathbf{V}_{\boldsymbol{c}}^{-1/2} \left(\mathrm{I}_m - \frac{\mathbf{V}_{\boldsymbol{c}}^{-1/2} \boldsymbol{B}_{\boldsymbol{c}} \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1/2}}{1 + \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}}} \right) \mathbf{V}_{\boldsymbol{c}}^{-1/2}.$$

Conclude, by taking the constraint $\boldsymbol{\omega}^\top \mathbf{1} = 1$ into account, that the optimal solution is

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

as announced. Then, if $\boldsymbol{\omega} = \boldsymbol{\omega}^{(\mathrm{AMSE})}$, straightforward but tedious calculations show that

$$(\boldsymbol{\omega}^\top \boldsymbol{B}_{\boldsymbol{c}})^2 = \frac{(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2}{[(1 + \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2]^2}$$
$$\text{and } \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega} = \frac{(1 + \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2 (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (2 + \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2}{[(1 + \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2]^2}.$$

The asymptotic normality of $\widehat{\theta}_n(\boldsymbol{\omega}^{(\mathrm{AMSE})})$ immediately follows, and the optimal value of the AMSE is

$$v_n^{-1} [(\boldsymbol{\omega}^\top \boldsymbol{B}_{\boldsymbol{c}})^2 + \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}] = \frac{1}{v_n} \times \frac{1 + \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}}}{(1 + \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2}.$$

The proof is complete. □

We discuss some implications of these optimal choices in terms of, among others, regularization of bias-optimal weights, sensitivity to uncertainty in weight estimation and gains in asymptotic variance.

Remark A.1 (Asymptotic variance-optimal weights and pseudo-maximum likelihood). *The asymptotic variance-optimal combination $\widehat{\theta}_n(\boldsymbol{\omega}^{(\mathrm{Var})})$ is in fact also a pseudo-maximum likelihood estimator of θ. Indeed, as suggested by Theorem A.1, pretend that $\sqrt{v_n}(\widehat{\boldsymbol{\theta}}_n - \theta\mathbf{1})^\top \stackrel{d}{=} \mathcal{N}(\boldsymbol{B_c}, \mathbf{V_c})$, and assume that $\boldsymbol{B} = \mathbf{0}$ and $\mathbf{V}$ is known and positive definite. Then $\boldsymbol{B_c} = \mathbf{0}$ and $\mathbf{V_c}$ is known and positive definite too, so that*

$$\widehat{\boldsymbol{\theta}}_n^\top \stackrel{d}{=} \mathcal{N}(\theta\mathbf{1}, \mathbf{V_c}/v_n).$$

Considering the vector of estimates $\widehat{\boldsymbol{\theta}}_n$ as a single data point from the multivariate $\mathcal{N}(\theta\mathbf{1}, \mathbf{V_c}/v_n)$ distribution from which θ is to be estimated, the log-likelihood is, as a function of θ,

$$\begin{aligned}\log L(\theta) &= -\frac{v_n}{2}(\widehat{\boldsymbol{\theta}}_n - \theta\mathbf{1})^\top \mathbf{V_c}^{-1}(\widehat{\boldsymbol{\theta}}_n - \theta\mathbf{1}) - \frac{1}{2}\log((2\pi/v_n)^m \det \mathbf{V_c}) \\ &= -\frac{v_n}{2}\left[(\mathbf{1}^\top \mathbf{V_c}^{-1}\mathbf{1})\theta^2 - 2(\mathbf{1}^\top \mathbf{V_c}^{-1}\widehat{\boldsymbol{\theta}}_n)\theta\right] + \text{constant}.\end{aligned}$$

This is obviously a degree 2 strictly concave polynomial maximized at

$$\frac{\mathbf{1}^\top \mathbf{V_c}^{-1}\widehat{\boldsymbol{\theta}}_n}{\mathbf{1}^\top \mathbf{V_c}^{-1}\mathbf{1}} = \left(\frac{\mathbf{V_c}^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{V_c}^{-1}\mathbf{1}}\right)^\top \widehat{\boldsymbol{\theta}}_n = (\boldsymbol{\omega}^{(\mathrm{Var})})^\top \widehat{\boldsymbol{\theta}}_n = \widehat{\theta}_n(\boldsymbol{\omega}^{(\mathrm{Var})}).$$

In other words, the maximum likelihood estimator of θ based on $\widehat{\boldsymbol{\theta}}_n$ in a multivariate Gaussian model is nothing but the variance-optimal pooled estimator $\widehat{\theta}_n(\boldsymbol{\omega}^{(\mathrm{Var})})$.

Remark A.2 (AMSE-optimal weights as a regularization of bias-optimal weights). *In contrast to variance-optimal weights, the AMSE-optimal solution attempts to balance both the bias and variance of the pooled estimator. One may consider instead the more general form $\mathrm{AMSE}(\boldsymbol{\omega}, \lambda) \propto (\boldsymbol{\omega}^\top \boldsymbol{B_c})^2 + \lambda(\boldsymbol{\omega}^\top \mathbf{V_c}\boldsymbol{\omega})$, with a penalty parameter $\lambda > 0$. The optimal set of weights for this generalized AMSE criterion is*

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

As expected, $\boldsymbol{\omega}^{(\mathrm{AMSE})}(\lambda) \to \boldsymbol{\omega}^{(\mathrm{Var})}$ as $\lambda \to \infty$, and we have, with $\boldsymbol{\omega}^{(\mathrm{AMSE})} = \boldsymbol{\omega}^{(\mathrm{AMSE})}(1)$, that

$$\begin{aligned}&[((\boldsymbol{\omega}^{(\mathrm{Var})})^\top \boldsymbol{B_c})^2 + (\boldsymbol{\omega}^{(\mathrm{Var})})^\top \mathbf{V_c}\boldsymbol{\omega}^{(\mathrm{Var})}] - [((\boldsymbol{\omega}^{(\mathrm{AMSE})})^\top \boldsymbol{B_c})^2 + (\boldsymbol{\omega}^{(\mathrm{AMSE})})^\top \mathbf{V_c}\boldsymbol{\omega}^{(\mathrm{AMSE})}] \\ &\qquad = \frac{(\mathbf{1}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})^2[(\boldsymbol{B_c}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})(\mathbf{1}^\top \mathbf{V_c}^{-1}\mathbf{1}) - (\mathbf{1}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})^2]}{(\mathbf{1}^\top \mathbf{V_c}^{-1}\mathbf{1})^2[(1 + \boldsymbol{B_c}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})(\mathbf{1}^\top \mathbf{V_c}^{-1}\mathbf{1}) - (\mathbf{1}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})^2]} \geq 0\end{aligned}$$

by the Cauchy-Schwarz inequality, with equality if and only if the vectors $\mathbf{V_c}^{-1/2}\mathbf{1}$ and $\mathbf{V_c}^{-1/2}\boldsymbol{B_c}$ are either orthogonal (when the two sets of weights lead to the same estimator) or collinear (when $\boldsymbol{B_c}$ is a constant vector).

The other solution, when $\lambda \downarrow 0$ and the vector $\boldsymbol{B_c}$ is not constant, is

$$\boldsymbol{\omega}^{(\mathrm{AMSE})}(0) = \frac{(\boldsymbol{B_c}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})\mathbf{V_c}^{-1}\mathbf{1} - (\mathbf{1}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})\mathbf{V_c}^{-1}\boldsymbol{B_c}}{(\boldsymbol{B_c}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})(\mathbf{1}^\top \mathbf{V_c}^{-1}\mathbf{1}) - (\mathbf{1}^\top \mathbf{V_c}^{-1}\boldsymbol{B_c})^2}.$$

[The denominator is indeed strictly positive, again by the Cauchy-Schwarz inequality.] When $m \geq 3$, this set of weights is the unique minimizer of $\boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}$ subject to the constraints $\boldsymbol{\omega}^\top \boldsymbol{B}_{\boldsymbol{c}} = 0$ and $\boldsymbol{\omega}^\top \mathbf{1} = 1$, that is, the unique set of weights minimizing the variance under the constraint that the bias is 0. This is shown by noting that the solution of this problem satisfies $\mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega} = \mu \mathbf{1} + \mu' \boldsymbol{B}_{\boldsymbol{c}}$, where μ and μ' are two Lagrange multipliers; taking the constraints $\boldsymbol{\omega}^\top \boldsymbol{B}_{\boldsymbol{c}} = 0$ and $\boldsymbol{\omega}^\top \mathbf{1} = 1$ into account and solving the associated system of two linear equations provides the solution $\boldsymbol{\omega}^{(\mathrm{Bias})} = \boldsymbol{\omega}^{(\mathrm{AMSE})}(0)$. In practice, this tends to produce a very unstable pooled estimator, particularly when the asymptotic biases of the individual estimators are relatively close. The AMSE-optimal weights are a powerful way of regularizing such bias-optimal weights and avoiding their inherent instability.

Remark A.3 (On the improvement in asymptotic variance and the sensitivity to uncertainty). *While the naive pooled estimator, obtained for $\boldsymbol{\omega} = \mathbf{1}/(\mathbf{1}^\top \mathbf{1})$, has asymptotic variance $(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}} \mathbf{1})/(\mathbf{1}^\top \mathbf{1})^2$, the variance-optimal pooled estimator has the improved asymptotic variance $1/(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1})$. The improvement factor is*

$$R_{\boldsymbol{c}} = \frac{\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}} \mathbf{1}}{(\mathbf{1}^\top \mathbf{1})^2} \Bigg/ \frac{1}{\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}} = \frac{\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}} \mathbf{1}}{\mathbf{1}^\top \mathbf{1}} \times \frac{\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{1}}.$$

By the Cauchy-Schwarz inequality, the product of Rayleigh quotients $R_{\boldsymbol{c}}$ is greater than or equal to 1, with equality if and only if $\mathbf{1}$ is an eigenvector of $\mathbf{V}_{\boldsymbol{c}}$, meaning that the naive pooled estimator is also variance-optimal if and only if $\mathbf{V}_{\boldsymbol{c}}$ is a positive multiple of a (doubly) stochastic matrix. A sharp upper bound on $R_{\boldsymbol{c}}$ follows from a Kantorovich inequality (see e.g. [14]): if $0 < \lambda_{1,\boldsymbol{c}} \leq \cdots \leq \lambda_{m,\boldsymbol{c}}$ are the eigenvalues of $\mathbf{V}_{\boldsymbol{c}}$,

$$R_{\boldsymbol{c}} \leq \frac{1}{4} \left(2 + \frac{\lambda_{1,\boldsymbol{c}}}{\lambda_{m,\boldsymbol{c}}} + \frac{\lambda_{m,\boldsymbol{c}}}{\lambda_{1,\boldsymbol{c}}} \right).$$

If $0 < \lambda_1 \leq \cdots \leq \lambda_m$ are the eigenvalues of $\mathbf{V}$ and $\mathrm{cond}(\mathbf{V}) = \lambda_m/\lambda_1$ is the condition number of $\mathbf{V}$ (i.e. the ratio between its largest and lowest eigenvalues), this can be further bounded above in a somewhat nicer fashion. Note indeed that for any $\boldsymbol{x} \neq \mathbf{0}$,

$$\frac{\boldsymbol{x}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}} = \frac{(\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})^\top \mathbf{V} (\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})}{(\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})^\top (\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})} \times \frac{(\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})^\top (\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})}{\boldsymbol{x}^\top \boldsymbol{x}}.$$

Now clearly

$$\left(\sum_{i=1}^m c_i^{-1} \right) \min_{1 \leq j \leq m} c_j \leq \frac{(\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})^\top (\mathbf{D}_{\boldsymbol{c}}^{1/2} \boldsymbol{x})}{\boldsymbol{x}^\top \boldsymbol{x}} \leq \left(\sum_{i=1}^m c_i^{-1} \right) \max_{1 \leq j \leq m} c_j$$

and so by the Courant-Fischer min-max characterization of eigenvalues of a symmetric matrix by the Rayleigh quotient (see Theorem 10, p.116 in [19]),

$$\left(\sum_{i=1}^m c_i^{-1} \right) \min_{1 \leq j \leq m} c_j \leq \frac{\lambda_{1,\boldsymbol{c}}}{\lambda_1} \leq \left(\sum_{i=1}^m c_i^{-1} \right) \max_{1 \leq j \leq m} c_j$$
$$\text{and } \left(\sum_{i=1}^m c_i^{-1} \right) \min_{1 \leq j \leq m} c_j \leq \frac{\lambda_{m,\boldsymbol{c}}}{\lambda_m} \leq \left(\sum_{i=1}^m c_i^{-1} \right) \max_{1 \leq j \leq m} c_j.$$

As a consequence,

$$R_{\boldsymbol{c}} \leq \frac{1}{4} \left(2 + \frac{\max_{1 \leq j \leq m} c_j}{\min_{1 \leq j \leq m} c_j} \left\{ \frac{1}{\mathrm{cond}(\mathbf{V})} + \mathrm{cond}(\mathbf{V}) \right\} \right).$$

In finite-samples, however, the uncertainty in the estimation of the weights may play a role in the performance of the pooled estimator. The sensitivity to this uncertainty can be measured using the condition number of the asymptotic variance function $\boldsymbol{\omega} \mapsto \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}$. The (relative) condition number of a nonlinear real-valued differentiable function f at $\boldsymbol{x} \neq \mathbf{0}$ w.r.t. the Euclidean norm $\|\cdot\|$ is (see [22])

$$\operatorname{cond}(f, \boldsymbol{x}) = \frac{\|\nabla f(\boldsymbol{x})\| \|\boldsymbol{x}\|}{|f(\boldsymbol{x})|} = \lim_{\varepsilon \downarrow 0} \sup_{\|\boldsymbol{h}\| \leq \varepsilon} \frac{|f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x})|}{|f(\boldsymbol{x})|} \Big/ \frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}$$

where $\nabla f(\boldsymbol{x})$ is the gradient of f at $\boldsymbol{x}$. For $f : \boldsymbol{\omega} \mapsto \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}$, $\nabla f(\boldsymbol{\omega}) = 2\mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}$, so

$$\operatorname{cond}(\boldsymbol{\omega} \mapsto \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}, \boldsymbol{\omega}^{(\mathrm{Var})}) = 2 \frac{\|\mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}\| \|\mathbf{1}\|}{\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}} = 2 \left(\frac{\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-2} \mathbf{1}}{\mathbf{1}^\top \mathbf{1}} \right)^{1/2} \left(\frac{\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{1}} \right)^{-1}.$$

This condition number is always larger than 2 by the Cauchy-Schwarz inequality. It can also be bounded from above using the Courant-Fischer min-max principle again:

$$\operatorname{cond}(\boldsymbol{\omega} \mapsto \boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega}, \boldsymbol{\omega}^{(\mathrm{Var})}) \leq 2 \frac{\lambda_{m,\boldsymbol{c}}}{\lambda_{1,\boldsymbol{c}}} \leq 2 \frac{\max_{1 \leq j \leq m} c_j}{\min_{1 \leq j \leq m} c_j} \operatorname{cond}(\mathbf{V}).$$

This means that the variance-optimal pooled estimator tends to achieve its best theoretical performance when at least a severe unbalance occurs between sample sizes or when the covariance matrix $\mathbf{V}$ approaches singularity, with the caveat that the resulting pooled estimator might be unstable if the estimated variance-optimal weights are not accurate enough.

Remark A.3 suggests the following practical guidelines for the choice of optimal weights. If the estimated weights are not too large in absolute values, then the associated (variance-optimal or AMSE-optimal) pooling strategy can be favored over both naive pooling and individual estimation. Otherwise, the resulting pooled estimator might be unstable and highly variable: its stability can then be checked by making use of bootstrap or resampling. To see why pooled estimators can have arbitrarily large variances under the sole assumption that $\boldsymbol{\omega}^\top \mathbf{1} = 1$, consider the case when $m = 2$ and $\widehat{\theta}_{1,n} = n^{-1} \sum_{i=1}^n X_i$ and $\widehat{\theta}_{2,n} = n^{-1} \sum_{i=1}^n Y_i$ are sample means of i.i.d. random pairs (X_i, Y_i) having common expectation m, unit marginal variances and correlation $\rho \in (-1, 1)$. In this case, the asymptotic variance of $\sqrt{n}(\widehat{\theta}_n(\omega, 1-\omega) - m)$ is

$$(\omega, 1-\omega) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \omega \\ 1-\omega \end{pmatrix} = 1 - 2(1-\rho)\omega(1-\omega)$$

which is obviously not bounded as $|\omega| \to \infty$.

If the chosen pooled estimator exhibits instability, one solution that keeps the advantages of pooling to some extent is to limit oneself to convex combinations of weights. Indeed, if $\boldsymbol{\omega} \geq \mathbf{0}$ elementwise,

$$\boldsymbol{\omega}^\top \mathbf{V}_{\boldsymbol{c}} \boldsymbol{\omega} = \sum_{i,j=1}^m [\mathbf{V}_{\boldsymbol{c}}]_{i,j} \omega_i \omega_j \leq \sum_{i,j=1}^m \frac{[\mathbf{V}_{\boldsymbol{c}}]_{i,i} + [\mathbf{V}_{\boldsymbol{c}}]_{j,j}}{2} \omega_i \omega_j = \sum_{i=1}^m [\mathbf{V}_{\boldsymbol{c}}]_{i,i} \omega_i.$$

This especially means that convex combinations can never produce a worse asymptotic variance than the individual estimator which contributes the highest variance to the pooling scheme. In fact, the naive pooled estimator itself, whose asymptotic variance is bounded above by the average of the diagonal elements of $\mathbf{V}_{\boldsymbol{c}}$, already brings a very substantial improvement if the individual variances are very different (for instance, due to unbalanced sample sizes).

Appendix B: Results of the main paper and their proofs

B.1. Auxiliary results

Lemma B.1 contains a covariance calculation which strengthens and corrects Lemma 6 in [25] (which is incorrect in the case of asymptotic independence). This generalized result is essential in the proof of Theorem 1.

Lemma B.1. *Suppose that X and Y have continuous distribution functions and satisfy conditions $\mathcal{C}_2(\gamma_X, \rho_X, A_X)$ and $\mathcal{C}_2(\gamma_Y, \rho_Y, A_Y)$, respectively. Assume also that there is a function R on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ such that we have the convergence*

$$\lim_{s\to\infty} s\,\mathbb{P}\left(\overline{F}_X(X) \le \frac{x}{s},\ \overline{F}_Y(Y) \le \frac{y}{s}\right) = R(x,y)$$

for any $(x,y) \in [0,\infty]^2 \setminus \{(\infty,\infty)\}$. Let $k_X = k_X(n), k_Y = k_Y(n)$ be such that:

- *$k_X, k_Y \to \infty$ and $k_X/n_X, k_Y/n_Y \to 0$;*
- *$n_X/n_Y \to b \in (0,\infty)$ and $k_X/k_Y \to c \in (0,\infty)$.*

Assume finally that f, g are continuously differentiable in a neighborhood of infinity, ultimately increasing, and such that f', g' are regularly varying at infinity with indices $a_X - 1$ and $a_Y - 1$, where $0 \le 2a_X\gamma_X < 1$ and $0 \le 2a_Y\gamma_Y < 1$. Then we have

$$\begin{aligned}
&\frac{k_X}{n_X}\operatorname{Cov}\left(\frac{[f(X)-f(U_X(n_X/k_X))]\mathbb{1}_{\{X>U_X(n_X/k_X)\}}}{\mathbb{E}([f(X)-f(U_X(n_X/k_X))]\mathbb{1}_{\{X>U_X(n_X/k_X)\}})}-1,\right.\\
&\qquad\qquad\left.\frac{[g(Y)-g(U_Y(n_Y/k_Y))]\mathbb{1}_{\{Y>U_Y(n_Y/k_Y)\}}}{\mathbb{E}([g(Y)-g(U_Y(n_Y/k_Y))]\mathbb{1}_{\{Y>U_Y(n_Y/k_Y)\}})}-1\right)\\
&\to \frac{(1-a_X\gamma_X)(1-a_Y\gamma_Y)}{\gamma_X\gamma_Y}\int_1^\infty\int_1^\infty x^{a_X-1}y^{a_Y-1}R\left(b^{-1}cx^{-1/\gamma_X}, y^{-1/\gamma_Y}\right)dx\,dy\\
&= (1-a_X\gamma_X)(1-a_Y\gamma_Y)\int_0^1\int_0^1 u^{-a_X\gamma_X}v^{-a_Y\gamma_Y}R\left(b^{-1}cu, v\right)\frac{du}{u}\frac{dv}{v},
\end{aligned}$$

$$\begin{aligned}
&\frac{k_X}{n_X}\operatorname{Cov}\left(\frac{[f(X)-f(U_X(n_X/k_X))]\mathbb{1}_{\{X>U_X(n_X/k_X)\}}}{\mathbb{E}([f(X)-f(U_X(n_X/k_X))]\mathbb{1}_{\{X>U_X(n_X/k_X)\}})}-1, \frac{\mathbb{1}_{\{Y>U_Y(n_Y/k_Y)\}}}{\mathbb{P}(Y>U_Y(n_Y/k_Y))}-1\right)\\
&\to \frac{1-a_X\gamma_X}{\gamma_X}\int_1^\infty x^{a_X-1}R\left(b^{-1}cx^{-1/\gamma_X},1\right)dx = (1-a_X\gamma_X)\int_0^1 u^{-a_X\gamma_X}R\left(b^{-1}cu,1\right)\frac{du}{u},
\end{aligned}$$

$$\begin{aligned}
&\frac{k_X}{n_X}\operatorname{Cov}\left(\frac{[g(Y)-g(U_Y(n_Y/k_Y))]\mathbb{1}_{\{Y>U_Y(n_Y/k_Y)\}}}{\mathbb{E}([g(Y)-g(U_Y(n_Y/k_Y))]\mathbb{1}_{\{Y>U_Y(n_Y/k_Y)\}})}-1, \frac{\mathbb{1}_{\{X>U_X(n_X/k_X)\}}}{\mathbb{P}(X>U_X(n_X/k_X))}-1\right)\\
&\to \frac{1-a_Y\gamma_Y}{\gamma_Y}\int_1^\infty y^{a_Y-1}R\left(b^{-1}c, y^{-1/\gamma_Y}\right)dy = (1-a_Y\gamma_Y)\int_0^1 y^{-a_Y\gamma_Y}R\left(b^{-1}c,v\right)\frac{dv}{v},
\end{aligned}$$

and

$$\frac{k_X}{n_X}\,\mathrm{Cov}\left(\frac{\mathbb{1}_{\{X>U_X(n_X/k_X)\}}}{\mathbb{P}(X>U_X(n_X/k_X))}-1,\ \frac{\mathbb{1}_{\{Y>U_Y(n_Y/k_Y)\}}}{\mathbb{P}(Y>U_Y(n_Y/k_Y))}-1\right)\to R(b^{-1}c,1)$$

as $n\to\infty$.

Proof of Lemma B.1. Write

$$\frac{n_Y}{k_Y}=\frac{n_X}{k_X\times(n_X/n_Y)\times(k_Y/k_X)}=\frac{n_X}{k_Y'}$$

and note that $k_X/k_Y'\to c/b$. Follow then the proof of Lemma 6 in [25], and use the change of variables $u=x^{-1/\gamma_X}$ and $v=y^{-1/\gamma_Y}$; in the specific case when R is everywhere 0, each covariance term in fact converges to 1, so the announced convergences hold because $k_X/n_X\to 0$. □

Next, we provide a general lemma about second-order regular variation with (strictly) negative second order parameter.

Lemma B.2. *Let U satisfy condition $C_2(\gamma,\rho,A)$, F be the associated distribution function and $\overline{F}=1-F$ be the associated survival function. If $\rho<0$, then there exists a constant $C>0$ such that*

$$U(t)=Ct^{\gamma}\left(1+\frac{A(t)}{\rho}+\mathrm{o}(|A(t)|)\right)\ \text{as } t\to\infty$$

$$\text{and}\ \ \overline{F}(x)=C^{1/\gamma}x^{-1/\gamma}\left(1+\frac{A(1/\overline{F}(x))}{\gamma\rho}+\mathrm{o}(|A(1/\overline{F}(x))|)\right)\ \text{as } x\to\infty.$$

Proof of Lemma B.2. The first asymptotic expansion on U is a direct consequence of the equation below Equation (2.3.23) in [7]. For the second one, write

$$U(1/\overline{F}(x))=C[\overline{F}(x)]^{-\gamma}\left(1+\frac{A(1/\overline{F}(x))}{\rho}+\mathrm{o}(|A(1/\overline{F}(x))|)\right)\ \text{as } x\to\infty.$$

Isolating $[\overline{F}(x)]^{-\gamma}$ and using a Taylor expansion then yields

$$\overline{F}(x)=C^{1/\gamma}[U(1/\overline{F}(x))]^{-1/\gamma}\left(1+\frac{A(1/\overline{F}(x))}{\gamma\rho}+\mathrm{o}(|A(1/\overline{F}(x))|)\right)\ \text{as } x\to\infty.$$

Use finally the local inversion relationship $U(1/\overline{F}(x))=x(1+\mathrm{o}(|A(1/\overline{F}(x))|))$ (see for example Lemma 1 in [4]) to complete the proof. □

The following lemma is a stronger version of Proposition A.5 in [21], appropriate for our purpose of showing the consistency of the tests of equality of tail indices and extreme quantiles.

Lemma B.3. *Assume that $(\mathbf{Z}_n)$ is a sequence of $m-$dimensional random vectors such that*

$$v_n(\mathbf{Z}_n-\boldsymbol{u}_n)\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$$

where $v_n \to \infty$, $(\boldsymbol{u}_n)$ is a sequence of $m-$dimensional nonrandom vectors and $\boldsymbol{\Sigma}$ is an $m \times m$ positive definite symmetric matrix. Let $(\widehat{\boldsymbol{\Sigma}}_n)$ be a random sequence of positive semidefinite symmetric matrices such that $\widehat{\boldsymbol{\Sigma}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}$. Set

$$\boldsymbol{D}_n = v_n \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \left(\boldsymbol{Z}_n - \frac{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \boldsymbol{Z}_n}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}} \mathbf{1} \right)$$

which is a well-defined sequence of $m-$dimensional random vectors with arbitrarily high probability as $n \to \infty$. Let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^m$.

(i) *If there is a real sequence (c_n) such that $v_n(\boldsymbol{u}_n - c_n \mathbf{1}) \to \mathbf{0}$, then $\boldsymbol{D}_n^\top \boldsymbol{D}_n \xrightarrow{d} \chi^2_{m-1}$.*

(ii) *If $\liminf_{n\to\infty} \inf_{c \in \mathbb{R}} \|\boldsymbol{u}_n - c\mathbf{1}\| > 0$, then $\boldsymbol{D}_n^\top \boldsymbol{D}_n \xrightarrow{\mathbb{P}} \infty$.*

Proof of Lemma B.3. With arbitrarily high probability as $n \to \infty$,

$$\boldsymbol{D}_n = v_n \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \left(\boldsymbol{Z}_n - \boldsymbol{u}_n - \frac{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} (\boldsymbol{Z}_n - \boldsymbol{u}_n)}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}} \mathbf{1} \right) + v_n \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \left(\boldsymbol{u}_n - \frac{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \boldsymbol{u}_n}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}} \mathbf{1} \right) \tag{B.1}$$

$$= v_n \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \left(\boldsymbol{Z}_n - \boldsymbol{u}_n - \frac{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} (\boldsymbol{Z}_n - \boldsymbol{u}_n)}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}} \mathbf{1} \right) + v_n \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \left(\boldsymbol{u}_n - c\mathbf{1} - \frac{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} (\boldsymbol{u}_n - c\mathbf{1})}{\mathbf{1}^\top \widehat{\boldsymbol{\Sigma}}_n^{-1} \mathbf{1}} \mathbf{1} \right) \tag{B.2}$$

whatever $c \in \mathbb{R}$ is. The first term in (B.1) and (B.2) converges in distribution to the Euclidean projection of a standard Gaussian random vector with independent components onto the orthogonal complement of the line spanned by the vector $\boldsymbol{\Sigma}^{-1/2}\mathbf{1}$. Applying Cochran's theorem, taking $c = c_n$ in (B.2) and noting that the second term therein is asymptotically negligible yields statement (i). To show statement (ii), use (B.1) and the reverse triangle inequality to get

$$\liminf_{n\to\infty} \|\boldsymbol{D}_n\| \geq \liminf_{n\to\infty} \left\{ v_n \inf_{c\in\mathbb{R}} \left\| \widehat{\boldsymbol{\Sigma}}_n^{-1/2} (\boldsymbol{u}_n - c\mathbf{1}) \right\| \right\} + \mathrm{O}_{\mathbb{P}}(1)$$

with arbitrarily high probability as $n \to \infty$. Use the assumption $\ell = \liminf_{n\to\infty} \inf_{c\in\mathbb{R}} \|\boldsymbol{u}_n - c\mathbf{1}\| > 0$ to get

$$\begin{aligned}
\liminf_{n\to\infty} \|\boldsymbol{D}_n\| &\geq \liminf_{n\to\infty} \left\{ v_n \left(\inf_{c\in\mathbb{R}} \|\boldsymbol{u}_n - c\mathbf{1}\| \left\| \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \frac{\boldsymbol{u}_n - c\mathbf{1}}{\|\boldsymbol{u}_n - c\mathbf{1}\|} \right\| \right) \right\} + \mathrm{O}_{\mathbb{P}}(1) \\
&\geq \ell \liminf_{n\to\infty} \left\{ v_n \inf_{\|\boldsymbol{u}\|=1} \left\| \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{u} \right\| \right\} + \mathrm{O}_{\mathbb{P}}(1).
\end{aligned}$$

Now

$$\sup_{\|\boldsymbol{u}\|=1} \left| \left\| \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{u} \right\| - \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{u} \right\| \right| \leq \sup_{\|\boldsymbol{u}\|=1} \left\| \left(\widehat{\boldsymbol{\Sigma}}_n^{-1/2} - \boldsymbol{\Sigma}^{-1/2} \right) \boldsymbol{u} \right\| \leq \left\| \widehat{\boldsymbol{\Sigma}}_n^{-1/2} - \boldsymbol{\Sigma}^{-1/2} \right\| \xrightarrow{\mathbb{P}} 0$$

by continuity of the matrix square root mapping. [On the right-hand side the norm is the matrix norm induced by the Euclidean norm.] Obviously $\inf_{\|\boldsymbol{u}\|=1} \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}\| > 0$ and therefore

$$\inf_{\|\boldsymbol{u}\|=1} \left\| \widehat{\boldsymbol{\Sigma}}_n^{-1/2} \boldsymbol{u} \right\| \xrightarrow{\mathbb{P}} \inf_{\|\boldsymbol{u}\|=1} \left\| \boldsymbol{\Sigma}^{-1/2} \boldsymbol{u} \right\| > 0.$$

Hence $\liminf_{n\to\infty} \|\boldsymbol{D}_n\| = \infty$ in probability, proving statement (ii). □

The next lemma makes it possible to use the second-order condition on U in the distributed inference framework uniformly over all machines in the case $m \to \infty$.

Lemma B.4. *Let $(Y_{i,j})$, $i, j \geq 1$, be a double array of unit Pareto random variables such that for any j, the $Y_{i,j}$, $i \geq 1$, are independent, and denote by $Y_{1:n_j,j} \leq \cdots \leq Y_{n_j:n_j,j}$ the order statistics related to the sample $(Y_{1,j}, \ldots, Y_{n_j,j})$. Assume that $m = m(n) \to \infty$, and $k_j = k_j(n)$, $n_j = n_j(n)$ are such that $\sup_{1\leq j\leq m} k_j/n_j \to 0$ and $\inf_{1\leq j\leq m} n_j/\log m \to \infty$. Then*

$$\inf_{1\leq j\leq m} Y_{n_j-k_j:n_j,j} \xrightarrow{\mathbb{P}} +\infty.$$

Proof of Lemma B.4. Define

$$K_i = \max\left(n_i \sup_{1\leq j\leq m} \frac{k_j}{n_j}, \sqrt{n_i}\right).$$

Then $\inf_{1\leq j\leq m} K_j \to \infty$, $\sup_{1\leq j\leq m} K_j/n_j \to 0$, and $K_i \geq k_i$, so that

$$\inf_{1\leq j\leq m} Y_{n_j-k_j:n_j,j} \geq \inf_{1\leq j\leq m} Y_{n_j-K_j:n_j,j}.$$

It is therefore enough to treat the case when additionally $\inf_{1\leq j\leq m} k_j \to \infty$.

Let first $Y_{n-k:n}$ be the $(n-k)$th largest order statistic of an i.i.d. unit Pareto sample of size n. Then, for any $t > 1$,

$$\begin{aligned}\mathbb{P}(Y_{n-k:n} \leq t) = \mathbb{P}(1/Y_{n-k:n} \geq 1/t) &= \int_{1/t}^{1} \frac{n!}{k!(n-k-1)!} x^k (1-x)^{n-k-1} dx \\ &\leq \binom{n}{k} (1-1/t)^{n-k}.\end{aligned}$$

By Stirling's formula, $N! = (N/e)^N \sqrt{2\pi N}(1+\mathrm{o}(1))$ as $N \to \infty$. A Taylor expansion of $x \mapsto \log(1+x)$ around 0 and straightforward computations then yield

$$\binom{n}{k} = \exp\left(-(1/2)\log(2\pi k) - k\log(k/n) + k(1+\epsilon(n,k,n/k))\right)$$

where the function ϵ (whose value changes from this line to the next) is such that $\epsilon(x,y,z) \to 0$ as $\min(x,y,z) \to \infty$. Consequently, since $x\log x \to 0$ as $x \downarrow 0$,

$$\mathbb{P}(Y_{n-k:n} \leq t) \leq \binom{n}{k} (1-1/t)^{n-k} = \exp(n[\log(1-1/t) + \epsilon(n,k,n/k)]). \tag{B.3}$$

Thus, by our assumptions on the n_j and k_j,

$$\begin{aligned}\mathbb{P}\left(\inf_{1\leq j\leq m} Y_{n_j-k_j:n_j,j} \leq t\right) &\leq \sum_{j=1}^{m} \mathbb{P}(Y_{n_j-k_j:n_j} \leq t) \text{ (the } Y_{i,j},\ i \geq 1\text{, are i.i.d. unit Pareto)} \\ &\leq m \sup_{1\leq j\leq m} \mathbb{P}(Y_{n_j-k_j:n_j} \leq t) \\ &\leq m\exp\left([\log(1-1/t) + \mathrm{o}(1)] \inf_{1\leq j\leq m} n_j\right) \to 0.\end{aligned}$$

This completes the proof. □

Lemma B.5 is a result of general interest on certain expectations of intermediate Pareto order statistics. It will be key when it comes to evaluating the bias term of the pooled tail index estimator in the case $m \to \infty$ of the distributed inference context.

Lemma B.5. *Let φ be a measurable, regularly varying function at infinity with index $\rho \le 0$, and bounded on all compact intervals of the form $[1, x_0]$, $x_0 > 1$. Let $Y_{n-k:n}$ be the $(n-k)$th largest order statistic of an i.i.d. unit Pareto sample of size n. Then, for any $k \ge 1$, $\mathbb{E}(\varphi(Y_{n-k:n}))$ is finite and, for any $\delta > 0$, there exists a positive integer n_0 such that*

$$\min(n, n/k) \ge n_0 \Rightarrow \left| \frac{\mathbb{E}(\varphi(Y_{n-k:n}))}{\varphi(n/k)} - k^\rho \frac{\Gamma(k-\rho+1)}{k!} \right| \le \delta.$$

[The quantity $k^\rho \Gamma(k-\rho+1)/k!$ is bounded in $k \ge 1$.] In particular, if $k = k(n) \to \infty$ with $k/n \to 0$, then $\mathbb{E}(\varphi(Y_{n-k:n}))/\varphi(n/k) \to 1$ as $n \to \infty$, and if k is instead fixed, then $\mathbb{E}(\varphi(Y_{n-k:n}))/\varphi(n) \to \Gamma(k-\rho+1)/k!$ as $n \to \infty$. [The latter convergence also applies when $k = 0$.]

Proof of Lemma B.5. By Equation (2.1.6) on p.10 of [6] and the identity $\Gamma(N+1) = N!$ valid for any nonnegative integer N,

$$\mathbb{E}(\varphi(Y_{n-k:n})) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \int_1^\infty \varphi(x) x^{-k} \left(1 - \frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2}.$$

Since φ is bounded on compact intervals and $x^{-1/2-\rho}\varphi(x)$ is bounded in a neighborhood of infinity, this expectation is finite.

To show the announced convergence, we define, for any $x_0 > 1$,

$$I_n(x_0) = \int_1^{x_0} \varphi(x) x^{-k} \left(1 - \frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2} \text{ and } J_n(x_0) = \int_{x_0}^\infty \varphi(x) x^{-k} \left(1 - \frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2},$$

as well as

$$I'_n(x_0) = \int_1^{x_0} x^{\rho-k} \left(1 - \frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2} \text{ and } J'_n(x_0) = \int_{x_0}^\infty x^{\rho-k} \left(1 - \frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2}.$$

Note first that

$$\left| \frac{J_n(x_0)}{\varphi(n/k)} - \left(\frac{k}{n}\right)^\rho J'_n(x_0) \right| \le \int_{x_0}^\infty \left| \frac{\varphi(x)}{\varphi(n/k)} - \left(\frac{kx}{n}\right)^\rho \right| x^{-k} \left(1 - \frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2}.$$

Choose $\iota \in (0,1)$ arbitrarily small. Then, by Proposition B.1.10 on p.369 of [7], we can fix $x_0 > 1$ such that, if n/k is large enough,

$$\left| \frac{J_n(x_0)}{\varphi(n/k)} - \left(\frac{k}{n}\right)^\rho J'_n(x_0) \right| \le \iota \int_{x_0}^\infty \left(\left(\frac{kx}{n}\right)^{\rho+\iota} + \left(\frac{kx}{n}\right)^{\rho-\iota} \right) x^{-k} \left(1 - \frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2}.$$

Clearly

$$\int_{x_0}^{\infty} \left(\left(\frac{kx}{n}\right)^{\rho+\iota} + \left(\frac{kx}{n}\right)^{\rho-\iota} \right) x^{-k} \left(1-\frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2}$$
$$\leq \Gamma(n-k) \left[\left(\frac{k}{n}\right)^{\rho+\iota} \frac{\Gamma(k-\rho-\iota+1)}{\Gamma(n-\rho-\iota+1)} + \left(\frac{k}{n}\right)^{\rho-\iota} \frac{\Gamma(k-\rho+\iota+1)}{\Gamma(n-\rho+\iota+1)} \right].$$

Hence, if $\psi(a,x) = x^a \Gamma(x-a+1)/\Gamma(x+1)$ for $a < 1$ and $x \geq 0$, the inequality

$$\frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \int_{x_0}^{\infty} \left(\left(\frac{kx}{n}\right)^{\rho+\iota} + \left(\frac{kx}{n}\right)^{\rho-\iota} \right) x^{-k} \left(1-\frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2} \leq \frac{\psi(\rho+\iota,k)}{\psi(\rho+\iota,n)} + \frac{\psi(\rho-\iota,k)}{\psi(\rho-\iota,n)}.$$

Since the Gamma function is continuous on $(0,\infty)$, the function ψ is obviously continuous on $(-\infty,1)\times[0,\infty)$ and, from [26], $\psi(a,x) \to 1$ as $x \to \infty$ uniformly in a belonging to any compact subinterval of $(-\infty,1)$, so there is a fixed constant $C_1 > 0$ such that

$$\frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \left| \frac{J_n(x_0)}{\varphi(n/k)} - \left(\frac{k}{n}\right)^{\rho} J_n'(x_0) \right| \leq C_1 \iota \tag{B.4}$$

when n and n/k are large enough. Besides, since $x^{1-\rho}\varphi(x) \to \infty$ as $x \to \infty$ and φ is bounded on $[1, x_0]$,

$$\frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \left| \frac{I_n(x_0)}{\varphi(n/k)} - \left(\frac{k}{n}\right)^{\rho} I_n'(x_0) \right|$$
$$\leq \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \int_1^{x_0} \left| \frac{\varphi(x)}{\varphi(n/k)} - \left(\frac{kx}{n}\right)^{\rho} \right| x^{-k} \left(1-\frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2}$$
$$\leq \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \left(\frac{n}{k}\right)^{1-\rho} \int_1^{x_0} C_2 \left(1-\frac{1}{x}\right)^{n-k-1} \frac{dx}{x^2} = C_2 \binom{n}{k} \left(\frac{n}{k}\right)^{1-\rho} \left(1-\frac{1}{x_0}\right)^{n-k}$$

when n/k is large enough, where $C_2 = C_2(x_0)$ is a finite positive constant. Stirling's formula, a Taylor expansion of $x \mapsto \log(1+x)$ and straightforward computations then entail

$$\binom{n}{k} = \exp(-\log(k!) + k\log(n/e) + (n-k)\log(1+k/(n-k)) + \varepsilon(n,n/k))$$

where the function ε (whose value may change from one line to the next) is such that $\varepsilon(x,y) \to 0$ as $\min(x,y) \to \infty$. Moreover

$$\log(k!) = \sum_{j=2}^{k} \log j \geq \sum_{j=2}^{k} \int_{j-1}^{j} \log t \, dt = \int_1^k \log t \, dt = k\log k - k$$

and thus simple calculations yield $\binom{n}{k} \leq \exp(n \times \varepsilon(n,n/k))$. Conclude that

$$\frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \left| \frac{I_n(x_0)}{\varphi(n/k)} - \left(\frac{k}{n}\right)^{\rho} I_n'(x_0) \right| \leq C_2 \exp(n[\log(1-1/x_0) + \varepsilon(n,n/k)]) \leq \iota \tag{B.5}$$

when n and n/k are large enough. Combine (B.4) and (B.5) and note that for any $x_0 > 1$,

$$I_n'(x_0) + J_n'(x_0) = \int_1^\infty x^{\rho-k}\left(1-\frac{1}{x}\right)^{n-k-1}\frac{dx}{x^2} = \frac{\Gamma(k-\rho+1)\Gamma(n-k)}{\Gamma(n-\rho+1)}$$

to get

$$\left|\frac{\mathbb{E}(\varphi(Y_{n-k:n}))}{\varphi(n/k)} - n^{-\rho}\frac{\Gamma(n+1)}{\Gamma(n-\rho+1)}\times k^{\rho}\frac{\Gamma(k-\rho+1)}{\Gamma(k+1)}\right| \le \iota(C_1+1)$$

for n and n/k large enough. Use once again the convergence $n^{\rho}\Gamma(n-\rho+1)/\Gamma(n+1)\to 1$ as $n\to\infty$ and the boundedness of $k^{\rho}\Gamma(k-\rho+1)/\Gamma(k+1) = \psi(\rho,k)$ in $k\ge 0$ to obtain

$$\left|\frac{\mathbb{E}(\varphi(Y_{n-k:n}))}{\varphi(n/k)} - k^{\rho}\frac{\Gamma(k-\rho+1)}{\Gamma(k+1)}\right| \le \iota(C_1+2)$$

for n and n/k large enough. This completes the proof since ι is arbitrarily small. □

B.2. Main results

Proof of Theorem 1. It suffices to show that

$$(\sqrt{k_1}(\widehat{\gamma}_1(k_1)-\gamma_1),\ldots,\sqrt{k_m}(\widehat{\gamma}_m(k_m)-\gamma_m))^{\top} \xrightarrow{d} \mathcal{N}(\boldsymbol{B},\mathbf{V}).$$

The second convergence stated in Theorem 1 (in the case $\gamma_j=\gamma$ for all j) and the assertions about optimal weights and the composite estimator with estimated weights will then directly follow from Theorem A.1.

Apply Corollary 1 in [25] to get, for any j,

$$\sqrt{k_j}(\widehat{\gamma}_j(k_j)-\gamma_j) = \sqrt{k_j}(\widetilde{\gamma}_j(k_j)-\gamma_j) + o_{\mathbb{P}}(1)$$
$$\text{where } \widetilde{\gamma}_j(k_j) = \frac{\sum_{i=1}^{n_j}[\log X_{i,j}-\log U_j(n_j/k_j)]\mathbb{1}_{\{X_{i,j}>U_j(n_j/k_j)\}}}{\sum_{i=1}^{n_j}\mathbb{1}_{\{X_{i,j}>U_j(n_j/k_j)\}}}.$$

We then equivalently show the joint asymptotic normality of the $\sqrt{k_j}(\widetilde{\gamma}_j(k_j)-\gamma_j)$, which are ratios of i.i.d. sums. By Lemma 3(ii) in [25] with $f=\log$,

$$\mathbb{E}([\log X_j - \log U_j(n_j/k_j)]|X_j > U_j(n_j/k_j)) = \gamma_j + \frac{1}{1-\rho_j}A(n_j/k_j) + o(1/\sqrt{k_j}).$$

Moreover, by Lemma 4 in [25],

$$\sqrt{k_j}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{[\log X_{i,j}-\log U_j(n_j/k_j)]\mathbb{1}_{\{X_{i,j}>U_j(n_j/k_j)\}}}{\mathbb{E}([\log X_j-\log U_j(n_j/k_j)]\mathbb{1}_{\{X_j>U_j(n_j/k_j)\}})}-1\right) = O_{\mathbb{P}}(1)$$

and

$$\sqrt{k_j}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\frac{\mathbb{1}_{\{X_{i,j}>U_j(n_j/k_j)\}}}{\mathbb{P}(X_j>U_j(n_j/k_j))}-1\right) = O_{\mathbb{P}}(1).$$

Pick $\boldsymbol{a} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$. Using the above three identities and linearizing yields

$$
\begin{aligned}
&\boldsymbol{a}^\top \begin{pmatrix} \sqrt{k_1}(\widetilde{\gamma}_1(k_1) - \gamma_1) \\ \vdots \\ \sqrt{k_m}(\widetilde{\gamma}_m(k_m) - \gamma_m) \end{pmatrix} - \sum_{j=1}^m a_j \frac{\lambda_j}{1-\rho_j} \\
&= \sum_{j=1}^m a_j \gamma_j \left\{ \sqrt{k_j} \left(\frac{n_j^{-1} \sum_{i=1}^{n_j} [\log X_{i,j} - \log U_j(n_j/k_j)] \mathbb{1}_{\{X_{i,j} > U_j(n_j/k_j)\}}}{\mathbb{E}([\log X_j - \log U_j(n_j/k_j)] \mathbb{1}_{\{X_j > U_j(n_j/k_j)\}})} - 1 \right) \right. \\
&\qquad \left. - \sqrt{k_j} \left(\frac{n_j^{-1} \sum_{i=1}^{n_j} \mathbb{1}_{\{X_{i,j} > U_j(n_j/k_j)\}}}{\mathbb{P}(X_j > U_j(n_j/k_j))} - 1 \right) \right\} + \mathrm{o}_{\mathbb{P}}(1).
\end{aligned} \tag{B.6}
$$

We now express the right-hand side as a sum of independent random variables. A straightforward modification of the proof of Lemma 4 in [25] provides, when $n_j' = n_j'(n)$ is such that $n_j'/n_j \to 1$,

$$
\sqrt{k_j} \left(\frac{1}{n_j} \sum_{i=\min(n_j, n_j')+1}^{\max(n_j, n_j')} \frac{[\log X_{i,j} - \log U_j(n_j/k_j)] \mathbb{1}_{\{X_{i,j} > U_j(n_j/k_j)\}}}{\mathbb{E}([\log X_j - \log U_j(n_j/k_j)] \mathbb{1}_{\{X_j > U_j(n_j/k_j)\}})} - 1 \right) = \mathrm{o}_{\mathbb{P}}(1)
$$

and

$$
\sqrt{k_j} \left(\frac{1}{n_j} \sum_{i=\min(n_j, n_j')+1}^{\max(n_j, n_j')} \frac{\mathbb{1}_{\{X_{i,j} > U_j(n_j/k_j)\}}}{\mathbb{P}(X_j > U_j(n_j/k_j))} - 1 \right) = \mathrm{o}_{\mathbb{P}}(1).
$$

Up to reordering the margins, one may assume without loss of generality that $b_m \le b_{m-1} \le \cdots \le b_1$. When $b_{j+1} < b_j$, this means that $n_{j+1}(n) > n_j(n)$ for n large enough; if $b_{j+1} = b_j$, then $n_{j+1}/n_j \to 1$, and according to the previous two identities, one may replace n_j and n_{j+1} by $\min(n_j, n_{j+1})$ and $\max(n_j, n_{j+1})$ respectively, up to $\mathrm{o}_{\mathbb{P}}(1)$ terms in (B.6). In other words, we may in fact also assume that $n_1 \le n_2 \le \cdots \le n_m$ without loss of generality. Define $n_0 = 0$. Then one may reformulate (B.6) as

$$
\begin{aligned}
&\boldsymbol{a}^\top \begin{pmatrix} \sqrt{k_1}(\widetilde{\gamma}_1(k_1) - \gamma_1) \\ \vdots \\ \sqrt{k_m}(\widetilde{\gamma}_m(k_m) - \gamma_m) \end{pmatrix} - \sum_{j=1}^m a_j \frac{\lambda_j}{1-\rho_j} \\
&= \sum_{\ell=1}^m \sum_{i=n_{\ell-1}+1}^{n_\ell} \sum_{j=\ell}^m a_j \gamma_j \left\{ \frac{\sqrt{k_j}}{n_j} \left(\frac{[\log X_{i,j} - \log U_j(n_j/k_j)] \mathbb{1}_{\{X_{i,j} > U_j(n_j/k_j)\}}}{\mathbb{E}([\log X_j - \log U_j(n_j/k_j)] \mathbb{1}_{\{X_j > U_j(n_j/k_j)\}})} - 1 \right) \right. \\
&\qquad \left. - \frac{\sqrt{k_j}}{n_j} \left(\frac{\mathbb{1}_{\{X_{i,j} > U_j(n_j/k_j)\}}}{\mathbb{P}(X_j > U_j(n_j/k_j))} - 1 \right) \right\} + \mathrm{o}_{\mathbb{P}}(1) \\
&= \sum_{\ell=1}^m \sum_{i=n_{\ell-1}+1}^{n_\ell} Z_{i,\ell,n}(\boldsymbol{a}) + \mathrm{o}_{\mathbb{P}}(1).
\end{aligned}
$$

Note that the $\sum_{i=n_{\ell-1}+1}^{n_\ell} Z_{i,\ell,n}(\boldsymbol{a})$, $1 \le \ell \le m$, are independent. It is then enough to obtain the weak convergence of each one of these random quantities. We shall show that

$$\forall \ell \in \{1,\ldots,m\}, \quad \sum_{i=n_{\ell-1}+1}^{n_\ell} Z_{i,\ell,n}(\boldsymbol{a}) \xrightarrow{d} \mathcal{N}(0,\sigma_\ell^2(\boldsymbol{a})) \tag{B.7}$$

$$\text{with } \sigma_\ell^2(\boldsymbol{a}) = \left(\frac{1}{b_\ell} - \frac{1}{b_{\ell-1}}\right)\left[\sum_{j=\ell}^m a_j^2 b_j \gamma_j^2 + \sum_{\substack{j,j'=\ell \\ j\neq j'}}^m a_j a_{j'} \gamma_j \gamma_{j'} \frac{R_{j,j'}(b_j c_{j'}, b_{j'} c_j)}{\sqrt{c_j}\sqrt{c_{j'}}}\right].$$

[Here we make the convention that $b_0 = \infty$ and $1/\infty = 0$, and that if $\sigma^2 = 0$, the Gaussian distribution $\mathcal{N}(0,\sigma^2)$ corresponds to the degenerate distribution at 0.] Note further that for any ℓ, the $Z_{i,\ell,n}(\boldsymbol{a})$, $n_{\ell-1}+1 \le i \le n_\ell$, are independent and centered, and that

$$\sum_{i=n_{\ell-1}+1}^{n_\ell} \mathrm{Var}(Z_{i,\ell,n}(\boldsymbol{a})) \to \sigma_\ell^2(\boldsymbol{a}) \quad \text{and} \quad \sum_{i=n_{\ell-1}+1}^{n_\ell} \mathbb{E}|Z_{i,\ell,n}(\boldsymbol{a})|^3 \to 0.$$

Indeed, the second convergence is a direct consequence of the Hölder inequality, Lemma 4 in [25] and the proportionality assumption on the k_j and n_j, while the first convergence is a consequence of convergences $k_j/k_\ell \to c_\ell/c_j$ and $n_j/n_\ell \to b_\ell/b_j$, Lemma B.1 and the identity

$$\forall \alpha > 0, \quad \int_0^1 \int_0^1 R_{j,\ell}(\alpha u, v) \frac{du}{u}\frac{dv}{v} = \int_0^1 R_{j,\ell}(\alpha u, 1) \frac{du}{u} + \int_0^1 R_{j,\ell}(\alpha, v) \frac{dv}{v}$$

valid by 1-homogeneity of $R_{j,\ell}$ (see Theorem 1(ii) in [23]). Convergence (B.7) now follows from the Lyapunov central limit theorem, see *e.g.* Theorem 27.3 in p.362 of [1] (except if $\ell \ge 2$ is such that $b_{\ell-1} = b_\ell$, in which case it immediately follows from the fact that $\sigma_\ell^2(\boldsymbol{a}) = 0$). Then, by independence of the $\sum_{i=n_{\ell-1}+1}^{n_\ell} Z_{i,\ell,n}(\boldsymbol{a})$ for $1 \le \ell \le m$, we get

$$\boldsymbol{a}^\top \begin{pmatrix} \sqrt{k_1}(\widetilde{\gamma}_1(k_1) - \gamma_1) \\ \vdots \\ \sqrt{k_m}(\widetilde{\gamma}_m(k_m) - \gamma_m) \end{pmatrix} - \sum_{j=1}^m a_j \frac{\lambda_j}{1-\rho_j} \to \mathcal{N}\left(0, \sum_{\ell=1}^m \sigma_\ell^2(\boldsymbol{a})\right).$$

Define $\overline{R}_{j,j'}(u,v) = \min(u,v)$ for $j = j'$ and $R_{j,j'}$ otherwise and note that

$$\begin{aligned}
\sum_{\ell=1}^m \sigma_\ell^2(\boldsymbol{a}) &= \sum_{\ell=1}^m \left(\frac{1}{b_\ell} - \frac{1}{b_{\ell-1}}\right) \sum_{j,j'=\ell}^m a_j a_{j'} \gamma_j \gamma_{j'} \frac{\overline{R}_{j,j'}(b_j c_{j'}, b_{j'} c_j)}{\sqrt{c_j}\sqrt{c_{j'}}} \\
&= \sum_{j,j'=1}^m a_j a_{j'} \gamma_j \gamma_{j'} \frac{\overline{R}_{j,j'}(b_j c_{j'}, b_{j'} c_j)}{\sqrt{c_j}\sqrt{c_{j'}}} \sum_{\ell=1}^{\min(j,j')} \left(\frac{1}{b_\ell} - \frac{1}{b_{\ell-1}}\right) \\
&= \sum_{j,j'=1}^m a_j a_{j'} \gamma_j \gamma_{j'} \frac{\overline{R}_{j,j'}(b_j c_{j'}, b_{j'} c_j)}{b_{\min(j,j')}\sqrt{c_j}\sqrt{c_{j'}}} \\
&= \sum_{j=1}^m a_j^2 \gamma_j^2 + \sum_{1\le j\neq j'\le m} a_j a_{j'} \gamma_j \gamma_{j'} \frac{R_{j,j'}(b_j c_{j'}, b_{j'} c_j)}{\max(b_j, b_{j'})\sqrt{c_j}\sqrt{c_{j'}}}.
\end{aligned}$$

Conclude with the Cramér-Wold device. $\square$

Proof of Corollary 1. The assumptions ensure that each $\widehat{\lambda}_j$ is a consistent estimator of λ_j, and therefore $\widehat{\boldsymbol{B}}_{\boldsymbol{c}}$ is a consistent estimator of $\boldsymbol{B}_{\boldsymbol{c}}$. We have also observed that the $\widehat{R}_{j,\ell}$ are locally uniformly consistent estimators of the $R_{j,\ell}$, and therefore $\widehat{\mathbf{V}}_{\boldsymbol{c}}$ is a consistent estimator of $\mathbf{V}_{\boldsymbol{c}}$. It is then straightforward that $\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})}$ and $\widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})}$ are consistent estimators of $\boldsymbol{\omega}^{(\mathrm{Var})}$ and $\boldsymbol{\omega}^{(\mathrm{AMSE})}$, respectively. The result now directly follows from Theorem 1. $\square$

Proof of Corollary 2. Following the proof of Corollary 1, $\widehat{\boldsymbol{B}}_{\boldsymbol{c}}$ is a consistent estimator of $\boldsymbol{B}_{\boldsymbol{c}}$, and $\widehat{\boldsymbol{\omega}}_n^{(\mathrm{Var})}$ and $\widehat{\boldsymbol{\omega}}_n^{(\mathrm{AMSE})}$ are consistent estimators of $\boldsymbol{\omega}^{(\mathrm{Var})}$ and $\boldsymbol{\omega}^{(\mathrm{AMSE})}$, respectively. Use then Corollary 1 and (for the AMSE-optimal estimator) the statement of Theorem A.1 dedicated to AMSE-optimal weighting. $\square$

Proof of Theorem 2. Recall that n_j/n and k_j/k have finite positive limits. As such,

$$\frac{\log(k_j/(n_j p))}{\log(k/(np))} - 1 = \frac{\log(k_j/k) - \log(n_j/n)}{\log(k/(np))} \to 0.$$

The first convergence is then a direct consequence of Theorem 1 and Theorem 4.3.8 on p.138 of [7]. To show the second asymptotic normality result, let $q = q_\ell$ for a fixed $\ell \in \{1, \ldots, m\}$, and start by writing

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

By Theorems 2.4.8, 3.2.5 and the equation at the top of p.139 in [7],

$$\log \frac{\widehat{q}_j^{\star}(1-p|k_j)}{q(1-p)} = \log\left(\frac{k}{np}\right)(\widehat{\gamma}_j(k_j) - \gamma) + \log\left(\frac{q_j(1-p)}{q(1-p)}\right) + \mathrm{O}_{\mathbb{P}}(1/\sqrt{k}).$$

Use Lemma B.2 to get

$$\begin{aligned}
\log\left(\frac{q_j(1-p)}{q(1-p)}\right) = \log\left(\frac{q_j(1-p)}{q_\ell(1-p)}\right) &= \mathrm{O}(|A_j(1/p)|) + \mathrm{O}(|A_\ell(1/p)|) \\
&= \mathrm{o}(|A_j(n_j/k_j)|) + \mathrm{o}(|A_\ell(n_\ell/k_\ell)|) \\
&= \mathrm{o}(1/\sqrt{k}).
\end{aligned}$$

Here the regular variation property of the A_j with index $\rho_j < 0$ was used along with the assumption $\sqrt{k_j} A(n_j/k_j) \to \lambda_j$, the asymptotic proportionality between the k_j, and the asymptotic proportionality between the n_j. Hence the identity

$$\log \frac{\widehat{q}_j^{\star}(1-p|k_j)}{q(1-p)} = \log\left(\frac{k}{np}\right)(\widehat{\gamma}_j(k_j) - \gamma) + \mathrm{O}_{\mathbb{P}}(1/\sqrt{k})$$

and therefore

$$\log \frac{\widehat{q}_n^\star(1-p|\widehat{\omega}_n)}{q_\ell(1-p)} = \log \frac{\widehat{q}_n^\star(1-p|\widehat{\omega}_n)}{q(1-p)} = \log\left(\frac{k}{np}\right)(\widehat{\gamma}_n(\widehat{\omega}_n)-\gamma) + \mathrm{O}_{\mathbb{P}}(1/\sqrt{k}).$$

Conclude using Theorem 1. □

Proof of Corollary 3. Set $\boldsymbol{Z}_n = \widehat{\boldsymbol{\gamma}}_n$ and note that $\Lambda_n = \boldsymbol{D}_n^\top \boldsymbol{D}_n$, where

$$\boldsymbol{D}_n = \sqrt{k}\,\overline{\mathbf{V}}_{\boldsymbol{c}}^{-1/2}\left(\boldsymbol{Z}_n - \frac{\mathbf{1}^\top \overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\boldsymbol{Z}_n}{\mathbf{1}^\top \overline{\mathbf{V}}_{\boldsymbol{c}}^{-1}\mathbf{1}}\mathbf{1}\right).$$

Theorem 1 yields $\sqrt{k}(\boldsymbol{Z}_n - \boldsymbol{\gamma}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_{\boldsymbol{c}})$. Besides, the fact that $\overline{\mathbf{V}}_{\boldsymbol{c}}$ is a consistent estimator of $\mathbf{V}_{\boldsymbol{c}}$ results from the local uniform consistency of the $\widehat{R}_{j,\ell}$, highlighted in Section 2.2. Conclude by applying Lemma B.3. The assertion on the asymptotic confidence interval is an immediate consequence of Theorem 1 and of the consistency of $\widehat{\mathbf{V}}_{\boldsymbol{c}}$. □

Proof of Corollary 4. Similarly to the proof of Corollary 3, note that $L_n(p) = \boldsymbol{D}_n^\top \boldsymbol{D}_n$, where now

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

We check the assumptions of Lemma B.3. Let $\log \boldsymbol{q}(1-p) = (\log q_1(1-p), \ldots, \log q_m(1-p))$ and note that, by the arguments of the proof of Theorem 2,

$$\frac{\sqrt{k}}{\log(k/(np))}(\boldsymbol{Z}_n(p) - \log \boldsymbol{q}(1-p)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_{\boldsymbol{c}}).$$

Again $\overline{\mathbf{V}}_{\boldsymbol{c}}$ is a consistent estimator of $\mathbf{V}_{\boldsymbol{c}}$. Use now Lemma B.2 to get

$$\log q_j(1-p) - \log q_\ell(1-p) = \mathrm{O}(|A_j(1/p)|) + \mathrm{O}(|A_\ell(1/p)|)$$

under assumption $(\mathcal{H})$, which allows to apply Lemma B.3(i) to get the result under the null hypothesis. Under the alternative hypothesis $(\mathcal{H}')$, if we had

$$\liminf_{n\to\infty} \inf_{c\in\mathbb{R}} \|\log \boldsymbol{q}(1-p(n)) - c\mathbf{1}\| = 0 \text{ (with } \|\cdot\| \text{ the Euclidean norm)}$$

then there would exist a sequence of integers (n_k) and constants c_{n_k} such that $\log q_j(1-p(n_k)) - c_{n_k} \to 0$ for any j. Then one would have $\log q_j(1-p(n_k)) - \log q_\ell(1-p(n_k)) \to 0$ for any pair (j,ℓ), which is a contradiction because, by Lemma B.2, at least one of the $\log q_j(\alpha) - \log q_\ell(\alpha)$ converges to a nonzero quantity as $\alpha \uparrow 1$. Conclude by applying Lemma B.3(ii). □

Proof of Corollary 5. This immediately follows from Theorem 1. □

Proof of Corollary 6. Recall that $k_j/k \to (\sum_{i=1}^m c_i^{-1})^{-1} c_j^{-1}$ and apply Corollary 5. □

Proof of Theorem 3. The first convergence immediately follows from Remark 8. To prove the second result, note first that

$$\forall j \in \{1,\ldots,m\},\ \left| \frac{U(n_j/k_j)}{U(n/k)} - \left(\frac{n_j k}{n k_j}\right)^{\gamma} \right| = \mathrm{O}(|A(n/k)|)$$

by Theorem 2.3.9 in [7]. Use then the assumptions $k_j/n_j = (k/n)(1+\mathrm{O}(1/\sqrt{k}))$ and $\sqrt{k}A(n/k) \to \lambda \in \mathbb{R}$ to get, by a Taylor expansion,

$$\forall j \in \{1,\ldots,m\},\ \frac{U(n_j/k_j)}{U(n/k)} = 1 + \mathrm{O}(1/\sqrt{k}).$$

Combine Corollaries 1 and 3 in [25] to obtain

$$\sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) = \sqrt{k_j}(\overline{\gamma}_j - \gamma) + \mathrm{o}_{\mathbb{P}}(1)$$
$$\text{where } \overline{\gamma}_j = \frac{\sum_{i=1}^{n_j} [\log X_{i,j} - \log U(n/k)] \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\sum_{i=1}^{n_j} \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}.$$

By Lemma 3(ii) in [25] again with $f = \log$,

$$\mathbb{E}([\log X - \log U(n/k)] | X > U(n/k)) = \gamma + \frac{1}{1-\rho} A(n/k) + \mathrm{o}(1/\sqrt{k}).$$

Write $n/k = n_j/k_j'$ with $k_j' = n_j k/n = k_j(1+\mathrm{o}(1))$. By Lemma 4 in [25],

$$\sqrt{k_j}\left(\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{[\log X_{i,j} - \log U(n/k)] \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\mathbb{E}([\log X - \log U(n/k)] \mathbb{1}_{\{X > U(n/k)\}})} - 1 \right) = \mathrm{O}_{\mathbb{P}}(1)$$

and

$$\sqrt{k_j}\left(\frac{1}{n_j} \sum_{i=1}^{n_j} \frac{\mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\mathbb{P}(X > U(n/k))} - 1 \right) = \mathrm{O}_{\mathbb{P}}(1).$$

Linearizing the $\overline{\gamma}_j$ and using the fact that $\sqrt{k}A(n/k) \to \lambda$ then yields

$$\begin{aligned}
&\sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) - \frac{\sqrt{k_j}}{\sqrt{k}} \frac{\lambda}{1-\rho} \\
&= \gamma \left\{ \sqrt{k_j} \left(\frac{n_j^{-1} \sum_{i=1}^{n_j} [\log X_{i,j} - \log U(n/k)] \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\mathbb{E}([\log X - \log U(n/k)] \mathbb{1}_{\{X > U(n/k)\}})} - 1 \right) \right. \\
&\qquad \left. - \sqrt{k_j} \left(\frac{n_j^{-1} \sum_{i=1}^{n_j} \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\mathbb{P}(X > U(n/k))} - 1 \right) \right\} + \mathrm{o}_{\mathbb{P}}(1).
\end{aligned}$$

We now concentrate on obtaining a similar representation for the Hill estimator of the pooled data. Apply again Corollary 1 in [25] to obtain

$$\sqrt{k}(\widehat{\gamma}_n^{(\mathrm{Hill})}(k) - \gamma) = \sqrt{k}(\widetilde{\gamma}_n(k) - \gamma) + \mathrm{o}_{\mathbb{P}}(1)$$

$$\text{where } \widetilde{\gamma}_n(k) = \frac{\sum_{i=1}^n [\log X_i - \log U(n/k)] \mathbb{1}_{\{X_i > U(n/k)\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > U(n/k)\}}}.$$

Linearize similarly $\widetilde{\gamma}_n(k)$ to get

$$\begin{aligned}
&\sqrt{k}(\widehat{\gamma}_n^{(\text{Hill})}(k) - \gamma) - \frac{\lambda}{1-\rho} \\
&= \gamma \left\{ \sqrt{k} \left(\frac{n^{-1} \sum_{i=1}^n [\log X_i - \log U(n/k)] \mathbb{1}_{\{X_i > U(n/k)\}}}{\mathbb{E}([\log X - \log U(n/k)] \mathbb{1}_{\{X > U(n/k)\}})} - 1 \right) \right. \\
&\qquad \left. - \sqrt{k} \left(\frac{n^{-1} \sum_{i=1}^n \mathbb{1}_{\{X_i > U(n/k)\}}}{\mathbb{P}(X > U(n/k))} - 1 \right) \right\} + \mathrm{o}_{\mathbb{P}}(1) \\
&= \gamma \left\{ \sqrt{k} \left(\frac{n^{-1} \sum_{j=1}^m \sum_{i=1}^{n_j} [\log X_{i,j} - \log U(n/k)] \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\mathbb{E}([\log X - \log U(n/k)] \mathbb{1}_{\{X > U(n/k)\}})} - 1 \right) \right. \\
&\qquad \left. - \sqrt{k} \left(\frac{n^{-1} \sum_{j=1}^m \sum_{i=1}^{n_j} \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\mathbb{P}(X > U(n/k))} - 1 \right) \right\} + \mathrm{o}_{\mathbb{P}}(1) \\
&= \sum_{j=1}^m \frac{n_j}{n} \frac{\sqrt{k}}{\sqrt{k_j}} \times \gamma \left\{ \sqrt{k_j} \left(\frac{n_j^{-1} \sum_{i=1}^{n_j} [\log X_{i,j} - \log U(n/k)] \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\mathbb{E}([\log X - \log U(n/k)] \mathbb{1}_{\{X > U(n/k)\}})} - 1 \right) \right. \\
&\qquad \left. - \sqrt{k_j} \left(\frac{n_j^{-1} \sum_{i=1}^{n_j} \mathbb{1}_{\{X_{i,j} > U(n/k)\}}}{\mathbb{P}(X > U(n/k))} - 1 \right) \right\} + \mathrm{o}_{\mathbb{P}}(1).
\end{aligned}$$

Use now the linearized expression of $\widehat{\gamma}_j(k_j)$ to obtain

$$\begin{aligned}
\sqrt{k}(\widehat{\gamma}_n^{(\text{Hill})}(k) - \gamma) - \frac{\lambda}{1-\rho} &= \sum_{j=1}^m \frac{n_j}{n} \frac{\sqrt{k}}{\sqrt{k_j}} \left(\sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) - \frac{\sqrt{k_j}}{\sqrt{k}} \frac{\lambda}{1-\rho} \right) + \mathrm{o}_{\mathbb{P}}(1) \\
&= \sqrt{k} \sum_{j=1}^m \frac{n_j}{n} (\widehat{\gamma}_j(k_j) - \gamma) - \frac{\lambda}{1-\rho} + \mathrm{o}_{\mathbb{P}}(1).
\end{aligned}$$

Since all the k_j/n_j have the same limit, and the $\widehat{\gamma}_j(k_j)$ are $\sqrt{k_j}$–consistent, we find

$$\begin{aligned}
\sqrt{k}(\widehat{\gamma}_n^{(\text{Hill})}(k) - \gamma) - \frac{\lambda}{1-\rho} &= \sqrt{k} \sum_{j=1}^m \frac{k_j}{k} (\widehat{\gamma}_j(k_j) - \gamma) - \frac{\lambda}{1-\rho} + \mathrm{o}_{\mathbb{P}}(1) \\
&= \sqrt{k}(\widehat{\gamma}_n(\widetilde{\omega}_n^{(\text{Var})}) - \gamma) - \frac{\lambda}{1-\rho} + \mathrm{o}_{\mathbb{P}}(1),
\end{aligned}$$

which completes the proof. □

Proof of Theorem 4. By Corollary 1, the optimal AMSE attainable using weighted pooling is

$$\text{AMSE}(\boldsymbol{\omega}^{(\text{AMSE})}) = \frac{1}{k} \times \frac{1 + \boldsymbol{B_c}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B_c}}{(1 + \boldsymbol{B_c}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B_c})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B_c})^2}.$$

We now collect a few further relationships. Define $\mathcal{S}_{\alpha,\beta} = \sum_{j=1}^{m} c_j^{-\alpha} b_j^{-\beta}$. Recall the identity

$$\lambda_j = c_j^{\rho-1/2} b_j^{-\rho} \lambda_1 = c_j^{\rho-1/2} b_j^{-\rho} \left(\sum_{j=1}^{m} \frac{1}{c_j}\right)^{\rho-1/2} \left(\sum_{j=1}^{m} \frac{1}{b_j}\right)^{-\rho} \lambda \tag{B.8}$$

(see Section 3.1). According to (B.8) and Remark 8, we have

$$\begin{aligned}
&\frac{\gamma^2(1-\rho)^2}{\lambda^2} (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2 - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) \\
&= \frac{1}{\gamma^2} \left(\frac{1}{\lambda^2} \left(\sum_{j=1}^{m} \frac{\lambda_j}{\sqrt{c_j}} \right)^2 - \sum_{j=1}^{m} \frac{1}{c_j} \right) \left(\sum_{j=1}^{m} \frac{1}{c_j} \right)^{-1} \\
&= \frac{1}{\gamma^2} \left(\left(\sum_{j=1}^{m} \frac{1}{c_j} \times \left[\frac{c_j}{b_j} \right]^{\rho} \right)^2 \left(\sum_{j=1}^{m} \frac{1}{c_j} \right)^{2\rho-2} \left(\sum_{j=1}^{m} \frac{1}{b_j} \right)^{-2\rho} - 1 \right) \\
&= \frac{1}{\gamma^2} \left(\mathcal{S}_{1-\rho,\rho}^2 \mathcal{S}_{1,0}^{2\rho-2} \mathcal{S}_{0,1}^{-2\rho} - 1 \right) > 0
\end{aligned} \tag{B.9}$$

(see also Section C.4 below). Recall moreover that $(\boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2$ is positive, by the Cauchy-Schwarz inequality and is, according to (B.8), a multiple of λ^2, because

$$\begin{aligned}
&(\boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2 \\
&= \frac{1}{\gamma^4(1-\rho)^2} \left(\sum_{j=1}^{m} \lambda_j^2 \sum_{j=1}^{m} \frac{1}{c_j} - \left(\sum_{j=1}^{m} \frac{\lambda_j}{\sqrt{c_j}} \right)^2 \right) \left(\sum_{j=1}^{m} \frac{1}{c_j} \right)^{-1} \\
&= \frac{\lambda^2}{\gamma^4(1-\rho)^2} \left(\mathcal{S}_{1-2\rho,2\rho} \mathcal{S}_{1,0} - \mathcal{S}_{1-\rho,\rho}^2 \right) \mathcal{S}_{1,0}^{2\rho-2} \mathcal{S}_{0,1}^{-2\rho} > 0.
\end{aligned} \tag{B.10}$$

We then compare $\mathrm{AMSE}(\boldsymbol{\omega}^{(\mathrm{AMSE})})$ to $\mathrm{AMSE}^{(\mathrm{Hill})} = k^{-1}(\gamma^2 + \lambda^2/(1-\rho)^2)$ by calculating

$$\begin{aligned}
&\frac{k(1-\rho)^2}{\lambda^2} (\mathrm{AMSE}(\boldsymbol{\omega}^{(\mathrm{AMSE})}) - \mathrm{AMSE}^{(\mathrm{Hill})}) [(1 + \boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2] \\
&= \left\{ \frac{\gamma^2(1-\rho)^2}{\lambda^2} (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2 - \mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1} \right\} - [(\boldsymbol{B}_{\boldsymbol{c}}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \boldsymbol{B}_{\boldsymbol{c}})^2] \\
&= \frac{1}{\gamma^2} \left[\left(\mathcal{S}_{1-\rho,\rho}^2 \mathcal{S}_{1,0}^{2\rho-2} \mathcal{S}_{0,1}^{-2\rho} - 1 \right) - \frac{\lambda^2}{\gamma^2(1-\rho)^2} \left(\mathcal{S}_{1-2\rho,2\rho} \mathcal{S}_{1,0} - \mathcal{S}_{1-\rho,\rho}^2 \right) \mathcal{S}_{1,0}^{2\rho-2} \mathcal{S}_{0,1}^{-2\rho} \right].
\end{aligned}$$

Here the identity $\gamma^2(\mathbf{1}^\top \mathbf{V}_{\boldsymbol{c}}^{-1} \mathbf{1}) = 1$ was used at the first step, and Equations (B.9) and (B.10) were used at the last step. Hence $\mathrm{AMSE}(\boldsymbol{\omega}^{(\mathrm{AMSE})}) \geq \mathrm{AMSE}^{(\mathrm{Hill})}$ if and only if

$$\lambda^2 \leq \gamma^2(1-\rho)^2 \frac{\mathcal{S}_{1-\rho,\rho}^2 \mathcal{S}_{1,0}^{2\rho-2} \mathcal{S}_{0,1}^{-2\rho} - 1}{\left(\mathcal{S}_{1-2\rho,2\rho} \mathcal{S}_{1,0} - \mathcal{S}_{1-\rho,\rho}^2 \right) \mathcal{S}_{1,0}^{2\rho-2} \mathcal{S}_{0,1}^{-2\rho}}.$$

Note then that $d_j = (c_j/b_j) \times (\mathcal{S}_{1,0}/\mathcal{S}_{0,1})$, and therefore

$$\mathcal{S}_{1-\rho,\rho}\mathcal{S}_{1,0}^{\rho}\mathcal{S}_{0,1}^{-\rho} = \sum_{j=1}^{m} \frac{d_j^{\rho}}{c_j} = S_\rho \quad \text{and} \quad \mathcal{S}_{1-2\rho,2\rho}\mathcal{S}_{1,0}^{2\rho}\mathcal{S}_{0,1}^{-2\rho} = \sum_{j=1}^{m} \frac{d_j^{2\rho}}{c_j} = S_{2\rho}.$$

Recall finally that $\mathcal{S}_{1,0} = S_0$: the proof is complete after some more straightforward algebra. □

Proof of Corollary 7. This follows from Corollary 5 and the consistency of the estimators of the bias components. □

Proof of Corollary 8. This immediately follows from combining Theorem 2, Corollary 5, Theorems 3 and 4, and Corollary 7. □

In the next few proofs we use Rényi's representation of order statistics of an independent standard exponential sample, which states that

$$(E_{k:n})_{1\le k\le n} \stackrel{d}{=} \left(\sum_{j=1}^{k} \frac{E_j}{n-j+1} \right)_{1\le k\le n} \tag{B.11}$$

whenever $E_1, \ldots, E_n$ are independent unit exponential random variables. See for example p.37 of [7].

Proof of Theorem 5. Up to reordering and without loss of generality, let $\ell' \in \{\ell+1, \ldots, m\}$ be such that $k_j \to \infty$ for any $j \in \{\ell+1, \ldots, \ell'\}$, and k_j is bounded for any $j \in \{\ell'+1, \ldots, m\}$. Write

$$\begin{aligned}\sqrt{k}(\widehat{\gamma}_n(\widetilde{\omega}_n^{(\mathrm{Var})}) - \gamma) &= \sum_{j=1}^{\ell} \widetilde{\omega}_{n,j}^{(\mathrm{Var})} \times \sqrt{k}(\widehat{\gamma}_j(k_j) - \gamma) \\ &\quad + \sum_{j=\ell+1}^{\ell'} \widetilde{\omega}_{n,j}^{(\mathrm{Var})} \times \sqrt{k}(\widehat{\gamma}_j(k_j) - \gamma) + \sum_{j=\ell'+1}^{m} \widetilde{\omega}_{n,j}^{(\mathrm{Var})} \times \sqrt{k}(\widehat{\gamma}_j(k_j) - \gamma).\end{aligned}$$

Follow the proof of Theorem 3.2.5 in [7] to find that, for any $j \in \{1, \ldots, \ell'\}$,

$$\sqrt{k_j}\left(\widehat{\gamma}_j(k_j) - \gamma - \frac{A(n_j/k_j)}{1-\rho} \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

Besides, for any $j \in \{\ell'+1, \ldots, m\}$,

$$\widetilde{\omega}_{n,j}^{(\mathrm{Var})} \sqrt{\frac{k}{k_j}} = \sqrt{\frac{k_j}{k}} = \sqrt{\frac{k}{k_1}} \times \frac{\sqrt{k_1}\sqrt{k_j}}{\sum_{i=1}^{m} k_i} \le \sqrt{\frac{k}{k_1}}\sqrt{\frac{k_j}{k_1}} = \mathrm{O}\left(\sqrt{\frac{k_j}{k_1}} \right) \to 0.$$

Conclude that

$$\forall j \in \{\ell+1, \ldots, \ell'\},\ \widetilde{\omega}_{n,j}^{(\mathrm{Var})} \times \sqrt{k}(\widehat{\gamma}_j(k_j) - \gamma) = \mathrm{o}_{\mathbb{P}}(1).$$

We now turn to the case $j \in \{\ell'+1, \ldots, m\}$. Let (Y_i) be independent unit Pareto random variables. Just as in the proof of Theorem 3.2.5 in [7], we have

$$\sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) \stackrel{d}{=} \gamma\sqrt{k_j}\left(\frac{1}{k_j}\sum_{i=1}^{k_j}\log\frac{Y_{n_j-i+1:n_j}}{Y_{n_j-k_j:n_j}} - 1\right)$$
$$+ \sqrt{k_j}A_0(Y_{n_j-k_j:n_j}) \times \frac{1}{k_j}\sum_{i=1}^{k_j}\frac{1}{\rho}\left(\left[\frac{Y_{n_j-i+1:n_j}}{Y_{n_j-k_j:n_j}}\right]^{\rho} - 1\right)$$
$$+ \mathrm{o}_{\mathbb{P}}\left(\sqrt{k_j}|A(Y_{n_j-k_j:n_j})| \times \frac{1}{k_j}\sum_{i=1}^{k_j}\left[\frac{Y_{n_j-i+1:n_j}}{Y_{n_j-k_j:n_j}}\right]^{\rho+\varepsilon}\right)$$

where A_0 is asymptotically equivalent to A and $\varepsilon > 0$ is arbitrarily small. By Rényi's representation (B.11),

$$\frac{1}{k_j}\sum_{i=1}^{k_j}\log\frac{Y_{n_j-i+1:n_j}}{Y_{n_j-k_j:n_j}} - 1 \stackrel{d}{=} \frac{1}{k_j}\sum_{i=1}^{k_j}\log Y_i - 1 = \mathrm{O}_{\mathbb{P}}(1)$$

and similarly

$$\frac{1}{k_j}\sum_{i=1}^{k_j}\frac{1}{\rho}\left(\left[\frac{Y_{n_j-i+1:n_j}}{Y_{n_j-k_j:n_j}}\right]^{\rho} - 1\right) = \mathrm{O}_{\mathbb{P}}(1) \text{ and } \frac{1}{k_j}\sum_{i=1}^{k_j}\left[\frac{Y_{n_j-i+1:n_j}}{Y_{n_j-k_j:n_j}}\right]^{\rho+\varepsilon} = \mathrm{O}_{\mathbb{P}}(1).$$

Since k_j is bounded and $Y_{n_j-k_j:n_j} \to \infty$ with probability 1 (see Lemma 3.2.1 in [7]), we have $A_0(Y_{n_j-k_j:n_j}) \xrightarrow{\mathbb{P}} 0$, resulting in $\sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) = \mathrm{O}_{\mathbb{P}}(1)$. Conclude that

$$\forall j \in \{\ell'+1, \ldots, m\},\ \widetilde{\omega}_{n,j}^{(\mathrm{Var})} \times \sqrt{k}(\widehat{\gamma}_j(k_j) - \gamma) = \mathrm{O}_{\mathbb{P}}\left(\sqrt{\frac{k_j}{k}}\right) = \mathrm{o}_{\mathbb{P}}(1).$$

Use now the convergences

$$\widetilde{\omega}_{n,j}^{(\mathrm{Var})} \to \left(\sum_{i=1}^{\ell}\frac{1}{c_i}\right)^{-1}\frac{1}{c_j} \text{ and } \sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_j}{1-\rho}, \gamma^2\right)$$

valid for any $j \in \{1, \ldots, \ell\}$ and the independence of the $\widehat{\gamma}_j(k_j)$ to obtain

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

The result about the asymptotic normality of the variance-optimal pooled tail index estimator follows due to the convergences $\sqrt{k_j}A(n_j/k_j) \to c_j^{\rho-1/2}b_j^{-\rho}\lambda_1$, valid for any $j \in \{1, \ldots, \ell\}$, and

$$\frac{\lambda}{\lambda_1} = \lim_{n\to\infty}\frac{\sqrt{k}}{\sqrt{k_1}} \times \frac{A(\sum_{i=1}^{\ell} n_i / \sum_{i=1}^{\ell} k_i)}{A(n_1/k_1)} = \left(\sum_{j=1}^{\ell}\frac{1}{c_j}\right)^{1/2-\rho}\left(\sum_{j=1}^{\ell}\frac{1}{b_j}\right)^{\rho}$$

which entail $\lambda_j = c_j^{\rho-1/2} b_j^{-\rho} \left(\sum_{j=1}^{\ell} c_j^{-1}\right)^{\rho-1/2} \left(\sum_{j=1}^{\ell} b_j^{-1}\right)^{-\rho} \lambda$ for any $j \in \{1, \ldots, \ell\}$.

Let us turn to the asymptotic behavior of the extreme quantile estimator. Write

$$\begin{aligned}
&\frac{\sqrt{k}}{\log(k/(np))} \log \frac{\widehat{q}_n^{\star}(1-p|\widetilde{\omega}_n^{(\mathrm{Var})})}{q(1-p)} - \sqrt{k}(\widehat{\gamma}_n(\widetilde{\omega}_n^{(\mathrm{Var})}) - \gamma) \\
&= \sum_{j=1}^{m} \frac{\sqrt{k_j}}{\sqrt{k}} \left\{ \frac{\log(k_j/(n_j p))}{\log(k/(np))} - 1 \right\} \sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) + \frac{1}{\log(k/(np))} \sum_{j=1}^{m} \frac{\sqrt{k_j}}{\sqrt{k}} \sqrt{k_j} \log \frac{X_{n_j-k_j:n_j,j}}{q(1-k_j/n_j)} \\
&+ \frac{1}{\log(k/(np))} \sum_{j=1}^{m} \frac{\sqrt{k_j}}{\sqrt{k}} \sqrt{k_j} \log\left(\left[\frac{k_j}{n_j p}\right]^{\gamma} \frac{q(1-k_j/n_j)}{q(1-p)} \right). \qquad \text{(B.12)}
\end{aligned}$$

In the first part of the proof, we showed that $\sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) = \mathrm{O}_{\mathbb{P}}(1)$ for any j. Control then the first term on the right-hand side as

$$\begin{aligned}
&\left| \sum_{j=1}^{m} \frac{\sqrt{k_j}}{\sqrt{k}} \left\{ \frac{\log(k_j/(n_j p))}{\log(k/(np))} - 1 \right\} \sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) \right| \\
&= \mathrm{O}_{\mathbb{P}}\left(\max_{1 \le j \le m} \frac{\sqrt{k_j}}{\sqrt{k}} \left| \frac{\log(k_j/(n_j p))}{\log(k/(np))} - 1 \right| \right) = \mathrm{o}_{\mathbb{P}}(1). \qquad \text{(B.13)}
\end{aligned}$$

To control the second term, we first apply Theorem 2.4.8 on p.52 of [7] to get

$$\left| \frac{1}{\log(k/(np))} \sum_{j=1}^{\ell'} \frac{\sqrt{k_j}}{\sqrt{k}} \sqrt{k_j} \log \frac{X_{n_j-k_j:n_j,j}}{q(1-k_j/n_j)} \right| = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\log(k/(np))} \right) = \mathrm{o}_{\mathbb{P}}(1). \qquad \text{(B.14)}$$

Then, by Theorems 1.1.6 on p.10, Lemma 1.2.9 on p.22 and Theorem 2.1.1 on p.38, all from [7],

$$\forall j \in \{\ell'+1, \ldots, m\}, \ \frac{X_{n_j-k_j:n_j,j}}{q(1-1/n_j)} \xrightarrow{d} Y_j$$

where Y_j is a positive random variable with probability 1 (to apply Theorem 2.1.1 on p.38 of [7], observe that since for such j, k_j is nondecreasing and bounded, it must be constant eventually). Hence

$$\forall j \in \{\ell'+1, \ldots, m\}, \ \log \frac{X_{n_j-k_j:n_j,j}}{q(1-k_j/n_j)} = \log \frac{X_{n_j-k_j:n_j,j}}{q(1-1/n_j)} + \log \frac{q(1-1/n_j)}{q(1-k_j/n_j)} = \mathrm{O}_{\mathbb{P}}(1)$$

using the regular variation property of $t \mapsto q(1-t^{-1})$. It follows that

$$\left| \frac{1}{\log(k/(np))} \sum_{j=\ell'+1}^{m} \frac{\sqrt{k_j}}{\sqrt{k}} \sqrt{k_j} \log \frac{X_{n_j-k_j:n_j,j}}{q(1-k_j/n_j)} \right| = \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\log(k/(np))} \right) = \mathrm{o}_{\mathbb{P}}(1). \qquad \text{(B.15)}$$

It remains to control the final term in (B.12). Use Lemma B.2 to get, for any j,

$$\log\left(\left[\frac{k_j}{n_j p}\right]^{\gamma} \frac{q(1-k_j/n_j)}{q(1-p)} \right) = \frac{A(n_j/k_j) - A(1/p)}{\rho} + \mathrm{o}(|A(n_j/k_j)|) + \mathrm{o}(|A(1/p)|).$$

For $j \in \{1, \ldots, \ell\}$, the assumptions on k_j, n_j, k, n and p immediately yield $k_j/(n_j p) \to \infty$, resulting in particular in $A(1/p) = \mathrm{o}(|A(n_j/k_j)|)$, and so

$$\left| \frac{1}{\log(k/(np))} \sum_{j=1}^{\ell} \frac{\sqrt{k_j}}{\sqrt{k}} \sqrt{k_j} \log\left(\left[\frac{k_j}{n_j p} \right]^{\gamma} \frac{q(1-k_j/n_j)}{q(1-p)} \right) \right| = \mathrm{O}\left(\frac{1}{\log(k/(np))} \right) = \mathrm{o}(1). \tag{B.16}$$

For $j \in \{\ell+1, \ldots, m\}$, $\sqrt{k_j}|A(1/p)| = \mathrm{o}(\sqrt{k_j}|A(n_1/k_1)|) = \mathrm{o}(\sqrt{k_1}|A(n_1/k_1)|) = \mathrm{o}(1)$, and then

$$\left| \frac{1}{\log(k/(np))} \sum_{j=\ell+1}^{m} \frac{\sqrt{k_j}}{\sqrt{k}} \sqrt{k_j} \log\left(\left[\frac{k_j}{n_j p} \right]^{\gamma} \frac{q(1-k_j/n_j)}{q(1-p)} \right) \right| = \mathrm{O}\left(\frac{1}{\log(k/(np))} \right) = \mathrm{o}(1). \tag{B.17}$$

Combine (B.12), (B.13), (B.14), (B.15), (B.16) and (B.17) to complete the proof. □

Proof of Theorem 6. This proof is different from those of Theorems 1, 2 and 3 in [3], not only because it allows for a general weighted distributed estimator and heterogeneous effective sample sizes k_j, but also because we use the second-order condition (as allowed by Lemma B.4) differently, resulting in a distinct treatment of the bias term (warranted by Lemma B.5) that allows to tackle the cases of bounded and unbounded k_j in a unified way. In particular, we only resort to a convergence in mean square rather than to a (seemingly unspecified) "weak law of large numbers for triangular array" as on p.8 of [3].

It is a consequence of condition $C_2(\gamma, \rho, A)$ that for t and tx large enough,

$$\left| \frac{\log U(tx) - \log U(t) - \gamma \log x}{A_0(t)} - \frac{x^{\rho} - 1}{\rho} \right| \le \varepsilon(\min(t, tx)) x^{\rho} \max(x^{\varepsilon(\min(t,tx))}, x^{-\varepsilon(\min(t,tx))}) \tag{B.18}$$

where A_0 is asymptotically equivalent to A and ε is a positive function such that $\varepsilon(z) \to 0$ as $z \to \infty$. See Theorem B.2.18 in [7], applied to $t \mapsto \log(t^{-\gamma} U(t))$; the function A_0 can, and is chosen here to, have constant sign. Now $\{X_{i,j}, 1 \le j \le m, 1 \le i \le n_j\} \stackrel{d}{=} \{U(Y_{i,j}), 1 \le j \le m, 1 \le i \le n_j\}$ where the $Y_{i,j}$ are i.i.d. unit Pareto, and thus

$$\begin{aligned} \sqrt{k_j}(\widehat{\gamma}_j(k_j) - \gamma) \stackrel{d}{=} \gamma \frac{1}{\sqrt{k_j}} \sum_{i=1}^{k_j} \left(\log \frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} - 1 \right) \\ + \sqrt{k_j} A_0(Y_{n_j-k_j:n_j,j}) \times \frac{1}{k_j} \sum_{i=1}^{k_j} \frac{1}{\rho} \left(\left[\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} \right]^{\rho} - 1 \right) + R_{n,j} \end{aligned}$$

where the equality in distribution holds jointly in $j \in \{1, \ldots, m\}$ and

$$|R_{n,j}| \le \varepsilon(Y_{n_j-k_j:n_j,j}) \sqrt{k_j} |A_0(Y_{n_j-k_j:n_j,j})| \times \frac{1}{k_j} \sum_{i=1}^{k_j} \left(\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} \right)^{\rho + \varepsilon(Y_{n_j-k_j:n_j,j})}.$$

Write then $\widehat{\gamma}_n(\omega) - \gamma = \sum_{j=1}^{m} \omega_j (\widehat{\gamma}_j(k_j) - \gamma)$, set $v_n = (\sum_{j=1}^{m} \omega_j^2/k_j)^{-1/2}$, and note that

$$v_n(\widehat{\gamma}_n(\omega) - \gamma) \stackrel{d}{=} S_{n,1} + S_{n,2} + S_{n,3} \tag{B.19}$$

$$\text{with } S_{n,1} = \gamma v_n \sum_{j=1}^{m} \frac{\omega_j}{\sqrt{k_j}} \left(\frac{1}{\sqrt{k_j}} \sum_{i=1}^{k_j} \left[\log \frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} - 1 \right] \right),$$

$$S_{n,2} = v_n \sum_{j=1}^{m} \omega_j A_0(Y_{n_j-k_j:n_j,j}) \times \frac{1}{k_j} \sum_{i=1}^{k_j} \frac{1}{\rho} \left(\left[\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} \right]^{\rho} - 1 \right)$$

$$\text{and } S_{n,3} = v_n \sum_{j=1}^{m} \frac{\omega_j}{\sqrt{k_j}} \times R_{n,j}.$$

We handle each of these three sums separately.

Asymptotic behavior of $S_{n,1}$: We show that $S_{n,1}$ has a Gaussian limiting distribution. Firstly, by the fact that the $\log Y_{i,j}$ are i.i.d. unit exponential and Rényi's representation (B.11),

$$\left(\sum_{i=1}^{k_j} \log \frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} \right)_{1 \le j \le m} \stackrel{d}{=} \left(\sum_{i=1}^{k_j} E_{i,j} \right)_{1 \le j \le m}$$

where the $E_{i,j}$, $1 \le j \le m$, $1 \le i \le k_j$, are i.i.d. unit exponential. Therefore

$$S_{n,1} \stackrel{d}{=} \gamma v_n \sum_{j=1}^{m} \frac{\omega_j}{\sqrt{k_j}} \left(\frac{1}{\sqrt{k_j}} \sum_{i=1}^{k_j} (E_{i,j} - 1) \right) = \gamma \sum_{j=1}^{m} \sum_{i=1}^{k_j} Z_{m,i,j}.$$

The $Z_{m,i,j}$ are independent and have expectation 0. We wish to apply the Lyapunov central limit theorem. To do so, we start by noting that

$$\mathrm{Var}\left(\sum_{j=1}^{m} \sum_{i=1}^{k_j} Z_{m,i,j} \right) = \sum_{j=1}^{m} \sum_{i=1}^{k_j} \mathrm{Var}(Z_{m,i,j}) = 1.$$

Besides, for any $\delta > 0$, the moment $\mathbb{E}|E_{1,1} - 1|^{2+\delta}$ is obviously finite, and

$$\begin{aligned} \sum_{j=1}^{m} \sum_{i=1}^{k_j} \mathbb{E}|Z_{m,i,j}|^{2+\delta} &= v_n^{2+\delta} \sum_{j=1}^{m} \frac{1}{k_j^{\delta/2}} \left(\frac{\omega_j^2}{k_j} \right)^{1+\delta/2} \mathbb{E}|E_{1,1} - 1|^{2+\delta} \\ &= \mathbb{E}|E_{1,1} - 1|^{2+\delta} \frac{\sum_{j=1}^{m} k_j^{-\delta/2} (\omega_j^2/k_j)^{1+\delta/2}}{(\sum_{j=1}^{m} \omega_j^2/k_j)^{1+\delta/2}}. \end{aligned}$$

Conclude that, if δ is chosen as in assumption $(\mathcal{W})$,

$$\frac{\sum_{j=1}^{m} \sum_{i=1}^{k_j} \mathbb{E}|Z_{m,i,j}|^{2+\delta}}{[\mathrm{Var}(\sum_{j=1}^{m} \sum_{i=1}^{k_j} Z_{m,i,j})]^{1+\delta/2}} = \mathrm{O}\left(\frac{\sum_{j=1}^{m} k_j^{-\delta/2} (\omega_j^2/k_j)^{1+\delta/2}}{(\sum_{j=1}^{m} \omega_j^2/k_j)^{1+\delta/2}} \right) \to 0 \text{ as } n \to \infty.$$

By Lyapunov's central limit theorem then,

$$S_{n,1} \xrightarrow{d} \mathcal{N}(0, \gamma^2). \tag{B.20}$$

Asymptotic behavior of $S_{n,2}$: We show that $S_{n,2} - \mathbb{E}(S_{n,2})$ converges to 0 in mean square (and therefore also in probability), and we calculate an asymptotic equivalent of $\mathbb{E}(S_{n,2})$ that we will interpret as a bias term. The derivation again uses Rényi's representation (B.11): the arguments of the proof of Lemma 3.2.3 in p.71 of [7] show that the random vectors

$$(Y_{n_j-k_j:n_j,j})_{1\le j\le m} \text{ and } \left(T_{n,j} = \frac{1}{k_j}\sum_{i=1}^{k_j}\frac{1}{\rho}\left(\left[\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}}\right]^{\rho} - 1\right)\right)_{1\le j\le m}$$

are independent, and

$$T_{n,j} \overset{d}{=} \frac{1}{k_j}\sum_{i=1}^{k_j}\frac{1}{\rho}(Y_i^{\rho} - 1),\ Y_i \text{ i.i.d. unit Pareto.}$$

In particular $\mathbb{E}(T_{n,j}) = 1/(1-\rho)$ and $\mathrm{Var}(T_{n,j}) = v_\rho/k_j$ where $v_\rho = \mathrm{Var}(\frac{1}{\rho}(Y_1^{\rho}-1)) < \infty$. By Lemma B.5 and the fact that A_0 and A are asymptotically equivalent,

$$\sup_{1\le j\le m}\left|\frac{\mathbb{E}(A_0(Y_{n_j-k_j:n_j,j}))}{A(n_j/k_j)} - k_j^{\rho}\frac{\Gamma(k_j-\rho+1)}{k_j!}\right| \to 0 \text{ as } n\to\infty. \tag{B.21}$$

Moreover, by the Cauchy-Schwarz inequality,

$$\limsup_{n\to\infty} v_n\sum_{j=1}^m |\omega_j||A(n_j/k_j)| \le \limsup_{n\to\infty}\left(\sum_{j=1}^m\left\{\sqrt{k_j}A(n_j/k_j)\right\}^2\right)^{1/2} < \infty. \tag{B.22}$$

This yields

$$\begin{aligned}\mathbb{E}(S_{n,2}) &= v_n\sum_{j=1}^m \omega_j\mathbb{E}(A_0(Y_{n_j-k_j:n_j,j}))\mathbb{E}(T_{n,j})\\ &= \frac{v_n}{1-\rho}\sum_{j=1}^m \omega_j k_j^{\rho}\frac{\Gamma(k_j-\rho+1)}{k_j!}A(n_j/k_j) + \mathrm{o}\left(v_n\sum_{j=1}^m|\omega_j||A(n_j/k_j)|\right)\\ &= \frac{v_n}{1-\rho}\sum_{j=1}^m \omega_j k_j^{\rho}\frac{\Gamma(k_j-\rho+1)}{k_j!}A(n_j/k_j) + \mathrm{o}(1).\end{aligned} \tag{B.23}$$

Let us now control the variance of $S_{n,2}$. Recall the obvious formula

$$\mathrm{Var}(XY) = \mathbb{E}(X^2)\mathbb{E}(Y^2) - [\mathbb{E}(X)\mathbb{E}(Y)]^2 = \mathbb{E}(X^2)\,\mathrm{Var}(Y) + \mathrm{Var}(X)\,[\mathbb{E}(Y)]^2$$

valid for any pair of square-integrable independent random variables X and Y. Using this formula gives

$$\begin{aligned}\mathrm{Var}(S_{n,2}) &= v_n^2\sum_{j=1}^m \omega_j^2\mathbb{E}(A_0^2(Y_{n_j-k_j:n_j,j}))\,\mathrm{Var}(T_{n,j}) + v_n^2\sum_{j=1}^m\omega_j^2\,\mathrm{Var}(A_0(Y_{n_j-k_j:n_j,j}))\,[\mathbb{E}(T_{n,j})]^2\\ &= v_n^2 v_\rho\sum_{j=1}^m\frac{\omega_j^2}{k_j}\mathbb{E}(A_0^2(Y_{n_j-k_j:n_j,j})) + \frac{v_n^2}{(1-\rho)^2}\sum_{j=1}^m\omega_j^2\,\mathrm{Var}(A_0(Y_{n_j-k_j:n_j,j})).\end{aligned}$$

By Lemma B.5,

$$v_n^2 \sum_{j=1}^m \frac{\omega_j^2}{k_j} \mathbb{E}(A_0^2(Y_{n_j-k_j:n_j,j})) = \mathrm{O}\left(v_n^2 \sum_{j=1}^m \frac{\omega_j^2}{k_j} A^2(n_j/k_j) \right) = \mathrm{O}\left(\sup_{1\le j\le m} |A(n_j/k_j)|^2 \right) = \mathrm{o}(1).$$

Besides, writing $\mathrm{Var}(A_0(Y_{n_j-k_j:n_j,j})) = \mathbb{E}(A_0^2(Y_{n_j-k_j:n_j,j})) - [\mathbb{E}(A_0(Y_{n_j-k_j:n_j,j}))]^2$ and using again Lemma B.5 entails

$$\sup_{1\le j\le m} \left| \frac{\mathrm{Var}(A_0(Y_{n_j-k_j:n_j,j}))}{A^2(n_j/k_j)} - \left\{ k_j^{2\rho} \frac{\Gamma(k_j-2\rho+1)}{k_j!} - \left[k_j^{\rho} \frac{\Gamma(k_j-\rho+1)}{k_j!} \right]^2 \right\} \right| \to 0$$

as $n \to \infty$. It follows that

$$\begin{aligned}
& v_n^2 \sum_{j=1}^m \omega_j^2 \, \mathrm{Var}(A_0(Y_{n_j-k_j:n_j,j})) \\
&= v_n^2 \sum_{j=1}^m \omega_j^2 \left\{ k_j^{2\rho} \frac{\Gamma(k_j-2\rho+1)}{k_j!} - \left[k_j^{\rho} \frac{\Gamma(k_j-\rho+1)}{k_j!} \right]^2 \right\} A^2(n_j/k_j) + \mathrm{o}\left(v_n^2 \sum_{j=1}^m \omega_j^2 A^2(n_j/k_j) \right) \\
&= v_n^2 \sum_{j=1}^m \frac{\omega_j^2}{k_j} \times k_j \left\{ k_j^{2\rho} \frac{\Gamma(k_j-2\rho+1)}{k_j!} - \left[k_j^{\rho} \frac{\Gamma(k_j-\rho+1)}{k_j!} \right]^2 \right\} A^2(n_j/k_j) + \mathrm{o}(1)
\end{aligned}$$

because $v_n^2 \sum_{j=1}^m \omega_j^2 A^2(n_j/k_j) = v_n^2 \sum_{j=1}^m (\omega_j^2/k_j) \times \{\sqrt{k_j} A(n_j/k_j)\}^2 = \mathrm{O}(1)$ by virtue of

$$\sup_{1\le j\le m} \{\sqrt{k_j} A(n_j/k_j)\}^2 \le \sum_{j=1}^m \{\sqrt{k_j} A(n_j/k_j)\}^2 = \mathrm{O}(1).$$

Following [26], for any fixed $a \in \mathbb{R}$, $\Gamma(x+a)/\Gamma(x) = x^{-a}(1+\mathrm{O}(1/x))$ as $x \to \infty$, and thus

$$\limsup_{k\to\infty} k \left| k^{2\rho} \frac{\Gamma(k-2\rho+1)}{k!} - \left[k^{\rho} \frac{\Gamma(k-\rho+1)}{k!} \right]^2 \right| < \infty.$$

This leads to

$$v_n^2 \sum_{j=1}^m \omega_j^2 \, \mathrm{Var}(A_0(Y_{n_j-k_j:n_j,j})) = \mathrm{O}\left(v_n^2 \sum_{j=1}^m \frac{\omega_j^2}{k_j} A^2(n_j/k_j) \right) + \mathrm{o}(1) = \mathrm{o}(1).$$

Conclude that $\mathrm{Var}(S_{n,2}) \to 0$ and use (B.23) to obtain

$$\left| S_{n,2} - \frac{v_n}{1-\rho} \sum_{j=1}^m \omega_j k_j^{\rho} \frac{\Gamma(k_j-\rho+1)}{k_j!} A(n_j/k_j) \right| \xrightarrow{\mathbb{P}} 0. \tag{B.24}$$

Asymptotic behavior of $S_{n,3}$: Pick an arbitrary $\iota \in (0,1)$. Use Lemma B.4 to get, with arbitrarily high probability as $n \to \infty$,

$$|S_{n,3}| \leq \iota v_n \sum_{j=1}^{m} |\omega_j| |A_0(Y_{n_j-k_j:n_j,j})| \times \frac{1}{k_j} \sum_{i=1}^{k_j} \left(\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} \right)^{\rho+\iota}. \tag{B.25}$$

Use once again Rényi's representation (B.11) to obtain that the random vectors

$$(Y_{n_j-k_j:n_j,j})_{1\leq j \leq m} \text{ and } \left(\frac{1}{k_j} \sum_{i=1}^{k_j} \left(\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} \right)^{\rho+\iota} \right)_{1\leq j\leq m}$$

are independent, and

$$\frac{1}{k_j} \sum_{i=1}^{k_j} \left(\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}} \right)^{\rho+\iota} \stackrel{d}{=} \frac{1}{k_j} \sum_{i=1}^{k_j} Y_i^{\rho+\iota}, \ Y_i \text{ i.i.d. unit Pareto.}$$

In particular the right-hand side in (B.25) has a finite expectation, which we control as

$$\begin{aligned} &\iota v_n \sum_{j=1}^{m} |\omega_j| \mathbb{E}(|A_0(Y_{n_j-k_j:n_j,j})|) \mathbb{E}(Y_1^{\rho+\iota}) \\ &\leq \iota \mathbb{E}(Y_1^{\rho+\iota}) \left\{ \sup_{k\geq 1} k^{\rho} \frac{\Gamma(k-\rho+1)}{k!} \right\} \left(v_n \sum_{j=1}^{m} |\omega_j| |A(n_j/k_j)| \right) (1+\mathrm{o}(1)) \end{aligned}$$

by (B.21) and the boundedness of $k^{\rho}\Gamma(k-\rho+1)/k!$ for $k \geq 1$. This is arbitrarily small thanks to (B.22). Conclude by (B.25) that

$$S_{n,3} \xrightarrow{\mathbb{P}} 0. \tag{B.26}$$

Combine (B.20), (B.24) and (B.26) to complete the proof. □

Proof of Theorem 7. The proof starts by writing a more general version of Equation (B.12) in the proof of Theorem 5:

$$\begin{aligned} &\frac{\sqrt{k}}{\log(k/(np))} \log \frac{\widehat{q}_n^{\star}(1-p|\omega)}{q(1-p)} - \sqrt{k}(\widehat{\gamma}_n(\omega)-\gamma) = \sqrt{k} \sum_{j=1}^{m} \omega_j \left\{ \frac{\log(k_j/(n_j p))}{\log(k/(np))} - 1 \right\} (\widehat{\gamma}_j(k_j)-\gamma) \\ &+ \frac{\sqrt{k}}{\log(k/(np))} \sum_{j=1}^{m} \omega_j \log \frac{X_{n_j-k_j:n_j,j}}{q(1-k_j/n_j)} + \frac{\sqrt{k}}{\log(k/(np))} \sum_{j=1}^{m} \omega_j \log \left(\left[\frac{k_j}{n_j p} \right]^{\gamma} \frac{q(1-k_j/n_j)}{q(1-p)} \right). \end{aligned} \tag{B.27}$$

We treat each of the three terms on the right-hand side separately.

(i) The structure of the proof will be used again in the proof of (ii), so we only emphasize the use of the extra condition $\omega_j \geq 0$ when it is necessary.

Let $Y_{i,j}$, $i,j \geq 1$, be i.i.d. unit Pareto random variables. First, one can use the uniform second-order inequality (B.18) to obtain, just as in the proof of Theorem 6 and with the notation therein,

$$\sqrt{k}\sum_{j=1}^{m}\omega_j\left\{\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right\}(\widehat{\gamma}_j(k_j)-\gamma)\overset{d}{=}\mathcal{S}_{n,1}+\mathcal{S}_{n,2}+\mathcal{S}_{n,3} \tag{B.28}$$

with

$$\mathcal{S}_{n,1}=\gamma\sum_{j=1}^{m}\frac{\sqrt{k}}{\sqrt{k_j}}\omega_j\left\{\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right\}\left(\frac{1}{\sqrt{k_j}}\sum_{i=1}^{k_j}\left[\log\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}}-1\right]\right),$$

$$\mathcal{S}_{n,2}=\sum_{j=1}^{m}\frac{\sqrt{k}}{\sqrt{k_j}}\omega_j\left\{\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right\}\sqrt{k_j}A_0(Y_{n_j-k_j:n_j,j})\times\frac{1}{k_j}\sum_{i=1}^{k_j}\frac{1}{\rho}\left(\left[\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}}\right]^{\rho}-1\right),$$

$$\text{and } \mathcal{S}_{n,3}=\sum_{j=1}^{m}\frac{\sqrt{k}}{\sqrt{k_j}}\omega_j\left\{\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right\}R_{n,j}.$$

Follow the steps of the control of $S_{n,1}$, $S_{n,2}$ and $S_{n,3}$ in the proof of Theorem 6 to find that

- $\mathcal{S}_{n,1}$ has expectation 0 and variance

$$\operatorname{Var}(\mathcal{S}_{n,1})=\gamma^2k\sum_{j=1}^{m}\frac{\omega_j^2}{k_j}\left\{\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right\}^2=\mathrm{O}\left(k\sum_{j=1}^{m}\frac{\omega_j^2}{k_j}\times\sup_{1\leq j\leq m}\left|\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right|^2\right).$$

This converges to 0, and therefore $\mathcal{S}_{n,1}\xrightarrow{\mathbb{P}}0$.
- $\mathcal{S}_{n,2}$ is such that

$$\begin{aligned}\mathbb{E}|\mathcal{S}_{n,2}| &\leq \frac{\sqrt{k}}{1-\rho}\sum_{j=1}^{m}|\omega_j|\left|\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right|\mathbb{E}|A_0(Y_{n_j-k_j:n_j,j})| \\ &=\mathrm{O}\left(\sup_{1\leq j\leq m}\left|\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right|\times\left(k\sum_{j=1}^{m}\frac{\omega_j^2}{k_j}\right)^{1/2}\times\left(\sum_{j=1}^{m}\left\{\sqrt{k_j}A(n_j/k_j)\right\}^2\right)^{1/2}\right)\to 0.\end{aligned}$$

Consequently $\mathcal{S}_{n,2}\xrightarrow{\mathbb{P}}0$.
- The quantity $\mathcal{S}_{n,3}$ satisfies, for any fixed $\iota\in(0,1)$,

$$|\mathcal{S}_{n,3}|\leq\iota\sup_{1\leq j\leq m}\left|\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right|\sqrt{k}\sum_{j=1}^{m}|\omega_j||A_0(Y_{n_j-k_j:n_j,j})|\frac{1}{k_j}\sum_{i=1}^{k_j}\left(\frac{Y_{n_j-i+1:n_j,j}}{Y_{n_j-k_j:n_j,j}}\right)^{\rho+\iota}$$

with arbitrarily high probability as $n\to\infty$, and hence $\mathcal{S}_{n,3}\xrightarrow{\mathbb{P}}0$ by bounding the expectation of the random variable on the right-hand side, see the control of $\mathcal{S}_{n,2}$ above.

Consequently

$$\sqrt{k}\sum_{j=1}^{m}\omega_j\left\{\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right\}(\widehat{\gamma}_j(k_j)-\gamma)\xrightarrow{\mathbb{P}}0. \tag{B.29}$$

Second, use the fact that $\inf_{1\le j\le m} n_j/k_j \to \infty$ together with Lemma B.4 and (B.18) to find, for any $\iota \in (0,|\rho|)$, that with arbitrarily high probability as $n\to\infty$,

$$\sum_{j=1}^m \omega_j \log \frac{X_{n_j-k_j:n_j,j}}{q(1-k_j/n_j)} \stackrel{d}{=} \gamma \sum_{j=1}^m \omega_j \log\left(\frac{k_j}{n_j} Y_{n_j-k_j:n_j,j}\right) + \sum_{j=1}^m \omega_j A_0(n_j/k_j)\frac{1}{\rho}\left(\left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^\rho - 1\right) + \sum_{j=1}^m \omega_j A_0(n_j/k_j)\mathcal{R}_{n,j} \tag{B.30}$$

where

$$|\mathcal{R}_{n,j}| \le \iota \max\left(\left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^{\rho+\iota}, \left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^{\rho-\iota}\right).$$

Rényi's representation (B.11) and straightforward calculations entail

$$\mathbb{E}\left(\log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right)\right) = \left\{\sum_{i=1}^{n_j}\frac{1}{i} - \log n_j\right\} - \left\{\sum_{i=1}^{k_j}\frac{1}{i} - \log k_j\right\} = c_{\text{Euler}} - \left\{\sum_{i=1}^{k_j}\frac{1}{i} - \log k_j\right\} + \mathrm{o}(1)$$

where $c_{\text{Euler}} = \lim_{N\to\infty}(\sum_{i=1}^N 1/i - \log N)$ denotes the Euler-Mascheroni constant (here the assumption $\inf_{1\le j\le m} n_j \to \infty$ was used). Observe that the sequence $(\sum_{i=1}^N 1/i - \log N)$ is (strictly) decreasing. Therefore, letting $k^\star = \limsup_{n\to\infty}\sup_{1\le j\le m} k_j < \infty$ and using the assumption that $\omega_j \ge 0$,

$$\limsup_{n\to\infty} \mathbb{E}\left(\sum_{j=1}^m \omega_j \log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right)\right) \le c_{\text{Euler}} - \left\{\sum_{i=1}^{k^\star}\frac{1}{i} - \log k^\star\right\} < 0. \tag{B.31}$$

Moreover, by Rényi's representation (B.11) and straightforward calculations again,

$$\mathrm{Var}\left(\sum_{j=1}^m \omega_j \log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right)\right) = \sum_{j=1}^m \omega_j^2 \sum_{i=k_j+1}^{n_j}\frac{1}{i^2} \le \sum_{j=1}^m \omega_j^2 \int_{k_j}^{n_j}\frac{dx}{x^2} \le \sum_{j=1}^m \frac{\omega_j^2}{k_j}.$$

As a consequence

$$\mathrm{Var}\left(\sum_{j=1}^m \omega_j \log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right)\right) = \mathrm{O}\left(\frac{1}{k}\right) \to 0. \tag{B.32}$$

Combine Equations (B.31) and (B.32) and the Chebyshev inequality to get

$$\sum_{j=1}^m \omega_j \log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right) \le \frac{1}{2}\left(c_{\text{Euler}} - \left\{\sum_{i=1}^{k^\star}\frac{1}{i} - \log k^\star\right\}\right) < 0 \quad \text{with arbitrarily high probability as } n\to\infty. \tag{B.33}$$

Now, since $\rho < 0$,

$$\left|\sum_{j=1}^m \omega_j A_0(n_j/k_j)\frac{1}{\rho}\left(\left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^\rho - 1\right) + \sum_{j=1}^m \omega_j A_0(n_j/k_j)\mathcal{R}_{n,j}\right|$$

$$\leq \left(\frac{1}{|\rho|}+\iota\right)\sum_{j=1}^{m}|\omega_j||A_0(n_j/k_j)|\left(1+\left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^{\rho+\iota}+\left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^{\rho-\iota}\right).$$

Using (B.22) and the asymptotic equivalence between $|A_0|$ and $|A|$ together with the assumptions that $\limsup_{n\to\infty}\sup_{1\leq j\leq m}k_j<\infty$ and $\inf_{1\leq j\leq m}n_j\to\infty$, we get

$$\sum_{j=1}^{m}|\omega_j||A_0(n_j/k_j)|=\mathrm{O}\left(\frac{1}{\sqrt{k}}\right).$$

Besides, by Lemma B.5 applied to $\varphi: x\mapsto x^{\rho\pm\iota}$ and again the assumptions that $\inf_{1\leq j\leq m}n_j\to\infty$ and $\limsup_{n\to\infty}\sup_{1\leq j\leq m}k_j<\infty$,

$$\sup_{1\leq j\leq m}\mathbb{E}\left(\left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^{\rho+\iota}+\left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^{\rho-\iota}\right)=\mathrm{O}(1).$$

Consequently

$$\frac{\sqrt{k}}{\log(k/(np))}\left|\sum_{j=1}^{m}\omega_jA_0(n_j/k_j)\frac{1}{\rho}\left(\left[\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right]^{\rho}-1\right)+\sum_{j=1}^{m}\omega_jA_0(n_j/k_j)\mathcal{R}_{n,j}\right|=\mathrm{o}_{\mathbb{P}}(1). \quad \text{(B.34)}$$

Third, apply (B.18) to get

$$\begin{aligned}&\sup_{1\leq j\leq m}\frac{1}{|A_0(n_j/k_j)|}\left|\log\left(\left[\frac{k_j}{n_jp}\right]^{\gamma}\frac{q(1-k_j/n_j)}{q(1-p)}\right)+A_0(n_j/k_j)\frac{1}{\rho}\left(\left[\frac{k_j}{n_jp}\right]^{\rho}-1\right)\right|\\&\leq\iota\sup_{1\leq j\leq m}\max\left(\left[\frac{k_j}{n_jp}\right]^{\rho+\iota},\left[\frac{k_j}{n_jp}\right]^{\rho-\iota}\right)\end{aligned} \quad \text{(B.35)}$$

for n large enough (use once again that $\inf_{1\leq j\leq m}n_j/k_j\to\infty$). Meanwhile

$$u_n=\sup_{1\leq j\leq m}\left|\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right|\to 0\Rightarrow\forall j\in\{1,\ldots,m\},\ \left(\frac{k}{np}\right)^{1-u_n}\leq\frac{k_j}{n_jp}\leq\left(\frac{k}{np}\right)^{1+u_n}$$

for n large enough, because $k/(np)\to\infty$. Thus, using this convergence again and the assumption that $\rho<0$, we arrive at

$$\sup_{1\leq j\leq m}\frac{1}{|A_0(n_j/k_j)|}\left|\log\left(\left[\frac{k_j}{n_jp}\right]^{\gamma}\frac{q(1-k_j/n_j)}{q(1-p)}\right)\right|=\mathrm{O}(1).$$

Conclude, by (B.22), that

$$\left|\frac{\sqrt{k}}{\log(k/(np))}\sum_{j=1}^{m}\omega_j\log\left(\left[\frac{k_j}{n_jp}\right]^{\gamma}\frac{q(1-k_j/n_j)}{q(1-p)}\right)\right|=\mathrm{O}\left(\frac{\sqrt{k}}{\log(k/(np))}\sum_{j=1}^{m}|\omega_j||A_0(n_j/k_j)|\right)=\mathrm{o}(1). \quad \text{(B.36)}$$

Combine (B.27), (B.29) and (B.36) and Theorem 6 to obtain

$$\log\frac{\widehat{q}_n^{\star}(1-p|\omega)}{q(1-p)}=\sum_{j=1}^{m}\omega_j\log\frac{X_{n_j-k_j:n_j,j}}{q(1-k_j/n_j)}+\mathrm{O}_{\mathbb{P}}\left(\frac{\log(k/(np))}{\sqrt{k}}\right).$$

Combine then (B.30), (B.33) and (B.34) to get the announced non-consistency result.

(ii) Recall the decomposition in (B.27). To prove the asymptotic normality, we show that the three terms on the right-hand side are $o_{\mathbb{P}}(1)$ separately.

The first term is controlled using (B.29), which is still valid.

The control of the second term starts by recalling Equation (B.30). Use first Rényi's representation (B.11) to get

$$\begin{aligned}\log(1+1/n_j)-\log(1+1/k_j) &= \int_{k_j+1}^{n_j+1}\frac{dx}{x}-\log(n_j/k_j)\le \mathbb{E}\left(\log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right)\right)\\ &= \sum_{i=k_j+1}^{n_j}\frac{1}{i}-\log(n_j/k_j)\le\int_{k_j}^{n_j}\frac{dx}{x}-\log(n_j/k_j)=0.\end{aligned}$$

Combine this double inequality with the elementary inequality $\log(1+x)\le x$ (valid for $x>0$) and the Cauchy-Schwarz inequality to find, for n large enough,

$$\sqrt{k}\left|\mathbb{E}\left(\sum_{j=1}^m \omega_j\log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right)\right)\right| = \mathrm{O}\left(\sqrt{k}\sum_{j=1}^m\frac{|\omega_j|}{k_j}\right)=\mathrm{O}\left(\left\{\sum_{j=1}^m\frac{1}{k_j}\right\}^{1/2}\right)=\mathrm{o}(1). \tag{B.37}$$

Recall finally Equation (B.32) and use the Chebyshev inequality to obtain

$$\sum_{j=1}^m\omega_j\log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right)=\mathrm{O}_{\mathbb{P}}(1/\sqrt{k}). \tag{B.38}$$

Equation (B.34) is still valid. This concludes the proof that the second term in the right-hand side of Equation (B.27) is $o_{\mathbb{P}}(1)$.

Finally, Equation (B.35) is valid under the assumption that $\sup_{1\le j\le m}k_j/n_j\to 0$, and then (B.36) is still valid, thus showing that the third term is a $\mathrm{o}(1)$. Conclude that

$$\frac{\sqrt{k}}{\log(k/(np))}\log\frac{\widehat{q}_n^{\star}(1-p|\omega)}{q(1-p)}=\sqrt{k}(\widehat{\gamma}_n(\omega)-\gamma)+o_{\mathbb{P}}(1).$$

Use now Theorem 6 in conjunction with the convergence $x^{\rho}\Gamma(x-\rho+1)/\Gamma(x+1)\to 1$ (as $x\to\infty$) and the fact that $\sqrt{k}\sum_{j=1}^m|\omega_j||A(n_j/k_j)|=\mathrm{O}(1)$ to complete the proof of the asymptotic normality result.

For a proof of the consistency statement under the weaker condition $\inf_{1\le j\le m}k_j\to\infty$, write the following simpler version of (B.27):

$$\begin{aligned}\log\frac{\widehat{q}_n^{\star}(1-p|\omega)}{q(1-p)} &= \log(k/(np))\left((\widehat{\gamma}_n(\omega)-\gamma)+\sum_{j=1}^m\omega_j\left\{\frac{\log(k_j/(n_jp))}{\log(k/(np))}-1\right\}(\widehat{\gamma}_j(k_j)-\gamma)\right)\\ &+\sum_{j=1}^m\omega_j\log\frac{X_{n_j-k_j:n_j,j}}{q(1-k_j/n_j)}+\sum_{j=1}^m\omega_j\log\left(\left[\frac{k_j}{n_jp}\right]^{\gamma}\frac{q(1-k_j/n_j)}{q(1-p)}\right).\end{aligned}$$

The first term is a $\mathrm{O}_{\mathbb{P}}(\log(k/(np))/\sqrt{k}) = \mathrm{o}_{\mathbb{P}}(1)$, by Theorem 6 and Equation (B.29). The third term is a $\mathrm{O}(\sum_{j=1}^{m}|\omega_j||A_0(n_j/k_j)|) = \mathrm{o}(1)$, by Equation (B.36). The second term is controlled by noting, according to Equation (B.37), that

$$\left|\mathbb{E}\left(\sum_{j=1}^{m}\omega_j \log\left(\frac{k_j}{n_j}Y_{n_j-k_j:n_j,j}\right)\right)\right| = \mathrm{O}\left(\frac{1}{\sqrt{k}}\left\{\sum_{j=1}^{m}\frac{1}{k_j}\right\}^{1/2}\right) = \mathrm{O}\left(\frac{1}{\inf_{1\le j\le m} k_j}\right) = \mathrm{o}(1).$$

Equations (B.30), (B.34) and (B.38) then ensure the desired convergence in probability to 0, which completes the proof. □

Proof of Theorem 8. Define

$$\overline{\gamma}_j(k_j) = \frac{1}{k_j}\sum_{i=1}^{k_j}\log(\varepsilon_{n_j-i+1:n_j,j}) - \log(\varepsilon_{n_j-k_j:n_j,j}) \text{ and } \overline{\gamma}_n(\omega) = \sum_{j=1}^{m}\omega_j\overline{\gamma}_j(k_j).$$

These are the pseudo-estimator counterparts of $\widehat{\gamma}_j(k_j)$ and $\widehat{\gamma}_n(\omega)$, where the unobserved errors replace the residuals. Apply Lemma A.3 in [11] to get

$$\max_{1\le j\le m}\sqrt{k_j}\sup_{0<s\le 1}\left|\log\left(\frac{\widehat{\varepsilon}^{(n_j)}_{n_j-\lfloor k_j s\rfloor:n_j,j}}{\varepsilon_{n_j-\lfloor k_j s\rfloor:n_j,j}}\right)\right| = \mathrm{o}_{\mathbb{P}}(1). \tag{B.39}$$

This immediately entails $\sqrt{k_j}(\widehat{\gamma}_j(k_j) - \overline{\gamma}_j(k_j)) = \mathrm{o}_{\mathbb{P}}(1)$ for any $j \in \{1,\dots,m\}$. As a consequence $\sqrt{k}(\widehat{\gamma}_n(\omega) - \overline{\gamma}_n(\omega)) = \mathrm{o}_{\mathbb{P}}(1)$, and the desired result now follows from Theorem 1. □

Proof of Theorem 9. Let q_j be the quantile function of ε_j. Set

$$\widehat{q}^{\star}_j(1-p|k_j,\omega) = \left(\frac{k_j}{n_j p}\right)^{\widehat{\gamma}_n(\omega)}\widehat{\varepsilon}^{(n_j)}_{n_j-k_j:n_j,j} \text{ and } \overline{q}^{\star}_j(1-p|k_j,\omega) = \left(\frac{k_j}{n_j p}\right)^{\overline{\gamma}_n(\omega)}\varepsilon_{n_j-k_j:n_j,j}.$$

[In the latter quantity, the notation of the proof of Theorem 8 was used.] Then following (B.39), the equation $\sqrt{k}(\widehat{\gamma}_n(\omega) - \overline{\gamma}_n(\omega)) = \mathrm{o}_{\mathbb{P}}(1)$ (see the proof of Theorem 8 again) and the fact that $\log(k_j/(n_j p)) = \log(k/(np))(1+\mathrm{o}(1))$ for any j, we obtain

$$\max_{1\le j\le m}\frac{\sqrt{k}}{\log(k/(np))}\left|\log\frac{\widehat{q}^{\star}_j(1-p|k_j,\widehat{\omega}_n)}{\overline{q}^{\star}_j(1-p|k_j,\widehat{\omega}_n)}\right| = \mathrm{o}_{\mathbb{P}}(1).$$

Applying the first convergence in Theorem 2 to $\overline{q}^{\star}_j(1-p|k_j,\widehat{\omega}_n)$ and using again $\sqrt{k}(\widehat{\gamma}_n(\omega) - \overline{\gamma}_n(\omega)) = \mathrm{o}_{\mathbb{P}}(1)$ and the delta-method then yields, for any $j \in \{1,\dots,m\}$,

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

The convergence of $\widehat{q}^{\star}_{X_j|\mathbf{Z}_j=z_j}(1-p|k_j,\widehat{\omega}_n)$ now follows from the identity

$$\frac{\widehat{q}^{\star}_{X_j|\mathbf{Z}_j=z_j}(1-p|k_j,\widehat{\omega}_n)}{q_{X_j|\mathbf{Z}_j=z_j}(1-p)}-1=\frac{\widehat{g}_j(z_j)-g_j(z_j)}{g_j(z_j)+\sigma_j(z_j)q_j(1-p)}$$
$$+\frac{\widehat{\sigma}_j(z_j)-\sigma_j(z_j)}{g_j(z_j)+\sigma_j(z_j)q_j(1-p)}q_j(1-p)+\frac{\widehat{\sigma}_j(z_j)}{g_j(z_j)/q_j(1-p)+\sigma_j(z_j)}\left(\frac{\widehat{q}^{\star}_j(1-p|k_j,\widehat{\omega}_n)}{q_j(1-p)}-1\right),$$

the assumptions on $\widehat{g}_j$ and $\widehat{\sigma}_j$, and the fact that $q_j(1-p)\to\infty$ as $p\downarrow 0$ by the heavy-tailed assumption on the ε_j.

Note then

$$\widetilde{q}^{\star}_n(1-p|\omega)=\prod_{j=1}^m\left[\left(\frac{k_j}{n_jp}\right)^{\widehat{\gamma}_j(k_j)}\widehat{\varepsilon}^{(n_j)}_{n_j-k_j:n_j,j}\right]^{\omega_j}$$
$$\text{and } \breve{q}^{\star}_n(1-p|\omega)=\prod_{j=1}^m\left[\left(\frac{k_j}{n_jp}\right)^{\overline{\gamma}_j(k_j)}\varepsilon_{n_j-k_j:n_j,j}\right]^{\omega_j}.$$

By (B.39) again and since $\sqrt{k_j}(\widehat{\gamma}_j(k_j)-\overline{\gamma}_j(k_j))=\mathrm{o}_{\mathbb{P}}(1)$ for any $j\in\{1,\ldots,m\}$,

$$\frac{\sqrt{k}}{\log(k/(np))}\left|\log\frac{\widetilde{q}^{\star}_n(1-p|\widehat{\omega}_n)}{\breve{q}^{\star}_n(1-p|\widehat{\omega}_n)}\right|=\mathrm{o}_{\mathbb{P}}(1).$$

Applying the second convergence in Theorem 2 to $\breve{q}^{\star}_n(1-p|\widehat{\omega}_n)$ and using again $\sqrt{k}(\widehat{\gamma}_n(\omega)-\overline{\gamma}_n(\omega))=\mathrm{o}_{\mathbb{P}}(1)$ and the delta-method then yields, for any $j\in\{1,\ldots,m\}$,

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

The convergence of $\widetilde{q}^{\star}_{X_j|\mathbf{Z}_j=z_j}(1-p|\widehat{\omega}_n)$ follows by the same calculation used to handle $\widehat{q}^{\star}_j(1-p|k_j,\widehat{\omega}_n)$. □

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

In the context of Section 2, assume that $A_j(t)=\gamma\beta_jt^{\rho_j}$; all commonly used heavy-tailed models satisfy this kind of proportionality assumption between $A_j(t)$ and t^{ρ_j}, see Table 1 in [12]. Define then

$$M^{(\ell)}_j(\kappa_n)=\frac{1}{\kappa_n}\sum_{i=1}^{\kappa_n}\left(\log X_{n_j-i+1:n_j,j}-\log X_{n_j-\kappa_n:n_j,j}\right)^{\ell},\ \text{for } \ell=1,2,3.$$

A reasonably well-performing estimator of ρ_j is

$$\widehat{\rho}^{(\tau)}_j(\kappa_n)=-\left|\frac{3(T^{(\tau)}_j(\kappa_n)-1)}{T^{(\tau)}_j(\kappa_n)-3}\right|,$$

$$\text{with } T_j^{(\tau)}(\kappa_n) = \begin{cases} \dfrac{\left(M_j^{(1)}(\kappa_n)\right)^{\tau} - \left(M_j^{(2)}(\kappa_n)/2\right)^{\tau/2}}{\left(M_j^{(2)}(\kappa_n)/2\right)^{\tau/2} - \left(M_j^{(3)}(\kappa_n)/6\right)^{\tau/3}} & \text{if } \tau > 0, \\[2ex] \dfrac{\log\left(M_j^{(1)}(\kappa_n)\right) - \frac{1}{2}\log\left(M_j^{(2)}(\kappa_n)/2\right)}{\frac{1}{2}\log\left(M_j^{(2)}(\kappa_n)/2\right) - \frac{1}{3}\log\left(M_j^{(3)}(\kappa_n)/6\right)} & \text{if } \tau = 0. \end{cases}$$

This estimator is implemented in the R function `mop` from the package `evt0`. In this package, $\kappa_n = \kappa_{n,j} = \lfloor n_j^{0.999} \rfloor$, and a choice of τ is made based on a stability criterion for $\kappa \mapsto \widehat{\rho}_j^{(\tau)}(\kappa)$ for large κ (see Section 3.2 in [13] for more details). According to Proposition 2.1 in [2], these choices ensure, if $\rho_j > -249.75$ (which will cover all practical applications), that $(\widehat{\rho}_j^{(\tau)}(\kappa_{n,j}) - \rho_j)\log(n_j) = o_{\mathbb{P}}(1)$. An estimator of β_j is then

$$\widehat{\beta}_j(\kappa_{n,j}) = \left(\frac{\kappa_{n,j}}{n_j}\right)^{\widehat{\rho}_j} \frac{T_j^{(1,0)}(\kappa_{n,j})T_j^{(0,1)}(\kappa_{n,j}) - T_j^{(1,1)}(\kappa_{n,j})}{T_j^{(1,0)}(\kappa_{n,j})T_j^{(1,1)}(\kappa_{n,j}) - T_j^{(2,1)}(\kappa_{n,j})},$$

$$\text{with } T_j^{(\ell,\ell')}(\kappa_{n,j}) = \frac{1}{\kappa_{n,j}} \sum_{i=1}^{\kappa_{n,j}} \left(\frac{i}{\kappa_{n,j}}\right)^{-\ell\widehat{\rho}_j} \left[i \log \frac{X_{n_j-i+1:n_j,j}}{X_{n_j-i:n_j,j}} \right]^{\ell'}, \ \widehat{\rho}_j = \widehat{\rho}_j^{(\tau)}(\kappa_{n,j}).$$

This estimator is also available from the R function `mop`. The aforementioned choice of $\kappa_{n,j}$ ensures that $\widehat{\beta}_j = \widehat{\beta}_j(\kappa_{n,j})$ is consistent, see Proposition 2.2 in [2].

C.2. About Remark 1

In the case $m = 2$ and $b_2 = \lim_{n\to\infty} n_1/n_2 \le 1$,

$$\mathbf{V}_{\boldsymbol{c}} = \gamma^2 \left(1 + \frac{1}{c_2}\right) \begin{pmatrix} 1 & R(c_2, b_2) \\ R(c_2, b_2) & c_2 \end{pmatrix}.$$

Then

$$\mathbf{V}_{\boldsymbol{c}}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{c_2}{\gamma^2(c_2+1)(c_2 - R^2(c_2, b_2))} \begin{pmatrix} c_2 - R(c_2, b_2) \\ 1 - R(c_2, b_2) \end{pmatrix}.$$

Both elements of this vector are nonnegative because $R(c_2, b_2) \le \min(c_2, b_2) \le \min(c_2, 1)$. This extends the calculation leading to Corollary 4 in [25] in the case of equal subsample sizes. This calculation also shows that for $c_2 = 1$, the weights are equal, which means that in fact $\boldsymbol{\omega}^{(\mathrm{Var})} = (1/2, 1/2)^\top$ irrespective of the value of the tail copula R.

We turn to the case $m = 3$ and $c_2 = c_3 = 1$, and we further assume that $b_2 = b_3 = 1$ to simplify the discussion. Set $r_{12} = R_{1,2}(1,1)$, $r_{13} = R_{1,3}(1,1)$ and $r_{23} = R_{2,3}(1,1)$, so that the relevant asymptotic covariance matrix is

$$\mathbf{V} = \gamma^2 \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix}.$$

By Sylvester's criterion, $\mathbf{V}$ is positive definite if and only if $0 \le r_{12} < 1$ and $\det(\mathbf{V}) = 1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23} > 0$. If these were the only conditions on the r_{ij} then (perhaps surprisingly) the asymptotic variance of the variance-optimal estimator, which is $3/(\mathbf{1}^\top \mathbf{V}^{-1}\mathbf{1})$, could be arbitrarily close to 0 depending on the values of the r_{ij}; consider for example the case $r_{12} = r_{23} = r \in [0, 1/\sqrt{2})$ and $r_{13} = 0$. Then $\mathbf{V}$ is positive definite and

$$\mathbf{V}^{-1} = \frac{1}{\gamma^2(1-2r^2)} \begin{pmatrix} 1-r^2 & -r & r^2 \\ -r & 1 & -r \\ r^2 & -r & 1-r^2 \end{pmatrix} \Rightarrow \frac{3}{\mathbf{1}^\top \mathbf{V}^{-1}\mathbf{1}} = \gamma^2 \frac{3(1-2r^2)}{3-4r}.$$

This gets arbitrarily close to 0 as $r \uparrow 1/\sqrt{2}$. The significance of this calculation is that, when $m \ge 3$, pooling together sample means constructed on observations that have certain positive correlation structures can be (much) more effective than calculating the sample mean of a full sample of independent data, even though the intuition would dictate that the positive correlation would always make the variance of the pooled estimator larger. That this intuition is incorrect follows from the identity

$$\mathbf{V}^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{\gamma^2(1-2r^2)} \begin{pmatrix} 1-r \\ 1-2r \\ 1-r \end{pmatrix}$$

from which it is seen that the second weight will be negative when $r \in (0, 1/2)$. In other words, in general pooling problems, the variance-optimal estimator is in fact not a convex combination of individual estimators, and thus one cannot rely on the intuition that positive correlations between estimators will always increase the variance in the pooled estimator.

Recall, however, that the multivariate extreme value setup imposes further restrictions on the r_{ij}. Specifically, it is a consequence of results from [15] (see Theorem 3.14 therein) and [20] that, in dimension m, the r_{ij} must satisfy

$$\max(0, r_{ij} + r_{jk} - 1) \le r_{ik} \le 1 - |r_{ij} - r_{jk}|, \quad \text{for } j \neq i, k.$$

These inequalities are sharp, in the sense that one can indeed find random vectors whose tail dependence coefficients r_{ij} satisfy the lower or upper bounds above, see Section 3 of [20]. As noted in [24], in dimension $m = 3$, these inequalities can also be rewritten as

$$r_{12}, r_{13}, r_{23} \ge 0, \ r_{12} + r_{13} - r_{23} \le 1, \ r_{12} + r_{23} - r_{13} \le 1, \ r_{13} + r_{23} - r_{12} \le 1. \tag{C.40}$$

The last three inequalities are interpreted as triangle inequalities on the $1 - r_{ij}$. The set of all such matrices can be identified with a convex stacked polytope $\mathcal{P}$ in $\mathbb{R}^3$ formed by gluing together two tetrahedra along the unit simplex, see Figure 3.1 in p.62 of [24]. This set $\mathcal{P}$ is a proper subset of the positive part $[0,1]^3 \cap \mathcal{E}$ of the Riemannian (quotient) manifold $\mathcal{E}$ called the *elliptope* (see [5], [16] and [18]) representing the set of standard correlation matrices, and defined by the inequalities

$$0 \le r_{12}, r_{13}, r_{23} \le 1, \ 1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23} \ge 0.$$

As we show in the next result, the set $\mathcal{P}$ touches the boundary of $\mathcal{E}$ exclusively at points (r_{12}, r_{13}, r_{23}) of the form $(1, r, r)$, $(r, 1, r)$ or $(r, r, 1)$ for some $r \in [0, 1]$. In other words, any vector of tail dependence coefficients that satisfies (C.40) and is not of the form $(1, r, r)$, $(r, 1, r)$ or $(r, r, 1)$ realizes a positive definite covariance matrix $\mathbf{V}$. Let $x = r_{12}$, $y = r_{13}$ and $z = r_{23}$.

Proposition C.1. *Let*

$$\mathcal{P} = \{(x, y, z) \in [0, \infty)^3 \mid x + y - z \leq 1, y + z - x \leq 1, z + x - y \leq 1\}$$

be the convex stacked polytope representing the set of admissible covariance matrices made of tail dependence coefficients, and

$$\mathcal{E} = \{(x, y, z) \in [-1, 1]^3 \mid 1 - x^2 - y^2 - z^2 + 2xyz \geq 0\}$$

be the elliptope representing the set of admissible correlation matrices. Let also

$$\operatorname{int}\mathcal{E} = \{(x, y, z) \in [-1, 1]^3 \mid 1 - x^2 - y^2 - z^2 + 2xyz > 0\}$$

denote the interior of $\mathcal{E}$, representing the set of positive definite correlation matrices, and $\partial\mathcal{E} = \mathcal{E} \setminus \operatorname{int}\mathcal{E}$ be the boundary of $\mathcal{E}$. Then $\mathcal{P} \subset \mathcal{E}$, and the intersection $\mathcal{P} \cap \partial\mathcal{E}$ is exclusively made of the points of the form $(1, r, r)$, $(r, 1, r)$ and $(r, r, 1)$, for some $r \in [0, 1]$.

Proof of Proposition C.1. Define a function d by

$$d(x, y, z) = 1 - x^2 - y^2 - z^2 + 2xyz.$$

We show that the minimum of d over $\mathcal{P}$ is 0, which will prove the inclusion $\mathcal{P} \subset \mathcal{E}$; this will be done by considering the behavior of d on the boundary of $\mathcal{P}$ and then on its interior. We prove then that $d(x, y, z) = 0$ on $\mathcal{P}$ if and only if (x, y, z) is of the form $(1, r, r)$, $(r, 1, r)$ and $(r, r, 1)$, for some $r \in [0, 1]$, which will complete the proof. Let $(x, y, z) \in \mathcal{P}$. Consider first the case $x = 0$. Then

$$d(0, y, z) = 1 - y^2 - z^2 = 1 - (y + z)^2 + 2yz \geq 0$$

because $y + z = y + z - x \leq 1$, and equality happens if and only if $(y, z) = (0, 1)$ or $(1, 0)$. The function d is similarly nonnegative on the planes defined by $y = 0$ and $z = 0$, with equality if and only if $(x, y, z) = (1, 0, 0)$, $(0, 1, 0)$ or $(0, 0, 1)$. Consider then the case $x + y - z = 1$, namely $z = x + y - 1$. Then

$$\begin{aligned} d(x, y, x + y - 1) &= 1 - x^2 - y^2 - (x + y - 1)^2 + 2xy(x + y - 1) \\ &= -2x^2(1 - y) + 2x(y^2 - 2y + 1) - 2(y^2 - y) \\ &= -2(1 - y)(x^2 - x(1 - y) - y) \\ &= 2(1 - y)(1 - x)(x + y) \geq 0. \end{aligned}$$

Equality happens if and only if $x = 1$, yielding $y = z$ and therefore $(x, y, z) = (1, r, r)$ for some $r \in [0, 1]$, or $y = 1$, yielding $x = z$ and therefore $(x, y, z) = (r, 1, r)$ for some $r \in [0, 1]$. A similar conclusion is reached on the planes defined by $y + z - x = 1$ and $z + x - y = 1$, with d being zero if and only if $(x, y, z) = (1, r, r)$, $(r, 1, r)$ or $(r, r, 1)$ for some $r \in [0, 1]$. We conclude that $\min_{\partial\mathcal{P}} d = 0$, and we finally examine the behavior of d when (x, y, z) belongs to the interior of $\mathcal{P}$. The gradient of d is

$$\nabla d(x, y, z) = \begin{pmatrix} -2x + 2yz \\ -2y + 2xz \\ -2z + 2xy \end{pmatrix}.$$

Cancelling this gradient leads to the equations $x = yz$, $y = xz$ and $z = xy$. Since $x, y, z > 0$, these three equations are equivalent to $x^2 = y^2 = z^2 = xyz$, and therefore $x = y = z \in (0, 1)$. At such points,

$$d(x, x, x) = 2x^3 - 3x^2 + 1 = 2(x - 1)^2(x + 1/2) > 0.$$

Since $\min_{\partial\mathcal{P}} d = 0$, obviously $\min_{\mathcal{P}} d \le 0$. If we had $\min_{\mathcal{P}} d < 0$, then d would attain its minimum in the interior of $\mathcal{P}$, but as we have seen, d is (strictly) positive at any of its critical points, which is an obvious contradiction. Thus $\min_{\mathcal{P}} d = 0$, meaning that $\mathcal{P} \subset \mathcal{E}$, and we have seen that $d(x,y,z)$ is 0 if and only if $(x,y,z) = (1,r,r)$, $(r,1,r)$ or $(r,r,1)$ for some $r \in [0,1]$. This completes the proof. □

For such vectors of tail dependence coefficients inducing a positive definite tail correlation matrix, we have

$$\mathbf{V}^{-1}\mathbf{1} = \frac{1}{\gamma^2(1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23})}\begin{pmatrix}(1-r_{23})(1+r_{23}-r_{12}-r_{13})\\(1-r_{13})(1+r_{13}-r_{12}-r_{23})\\(1-r_{12})(1+r_{12}-r_{13}-r_{23})\end{pmatrix}.$$

By the set of inequalities (C.40), this vector has nonnegative entries, and the variance-optimal weight vector is then again a convex combination. The fact that the asymptotic variance $3/(\mathbf{1}^\top\mathbf{V}^{-1}\mathbf{1})$ of the variance-optimal pooled estimator is not less than γ^2 for $m=3$ is now obvious, because this pooled estimator is a convex combination of estimators having individual variance γ^2 and nonnegative correlations. We provide another, purely analytical proof based on the Lagrange multiplier method and the fact that

$$\frac{3}{\mathbf{1}^\top\mathbf{V}^{-1}\mathbf{1}} = \gamma^2\frac{3(1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}r_{13}r_{23})}{3-r_{12}^2-r_{13}^2-r_{23}^2+2(r_{12}r_{13}+r_{12}r_{23}+r_{13}r_{23}-r_{12}-r_{13}-r_{23})}.$$

Proposition C.2. *Work with the notation of Proposition C.1. Then for any* $(x,y,z) \in \mathcal{P} \setminus (\mathcal{P} \cap \partial\mathcal{E})$*,*

$$\frac{3(1-x^2-y^2-z^2+2xyz)}{3-x^2-y^2-z^2+2(xy+xz+yz-x-y-z)} \ge 1$$

with equality if and only if $x=y=z=0$.

Proof of Proposition C.2. The numerator in the ratio on the left-hand side of the announced inequality is positive on $\mathcal{S} = \mathcal{P} \setminus (\mathcal{P} \cap \partial\mathcal{E})$, and so is the ratio itself (because it is inversely proportional to the sum of the elements of a positive definite symmetric matrix). It is thus straightforward to note that proving the desired inequality is equivalent to showing that, on $\mathcal{S}$,

$$f(x,y,z) = -x^2-y^2-z^2-xy-xz-yz+x+y+z+3xyz \ge 0.$$

We equivalently show that $\min_{\mathcal{S}} f = 0$. As in the proof of Proposition C.1, we do so by minimizing first f on the boundary of $\mathcal{S}$, before looking for its potential extrema within the interior of $\mathcal{S}$.

If $x=0$, the constraints become $y,z \ge 0$ and $y+z \le 1$, and the function f becomes

$$f(0,y,z) = -y^2-z^2-yz+y+z = yz-(y+z)(y+z-1).$$

Then obviously $f(0,y,z) \ge 0$ and f attains this lower bound at $y=z=0$ (the choices $(y,z)=(0,1)$ or $(1,0)$ are not admissible since the corresponding points belong to $\partial\mathcal{E}$). Similarly in the cases $y=0$ and $z=0$.

If $x+y-z=1$, the function f becomes

$$\begin{aligned} f(x,y,x+y-1) &= -x^2-y^2-(x+y-1)^2+2x+2y-1 \\ &\quad -xy-x(x+y-1)-y(x+y-1)+3xy(x+y-1) \end{aligned}$$

$$
\begin{aligned}
&= -3x^2(1-y) + x(3y^2 - 8y + 5) - (3y^2 - 5y + 2) \\
&= -3(1-y)(x^2 + [y - 5/3]x - [y - 2/3]) \\
&= 3(1-y)(1-x)(x + y - 2/3).
\end{aligned}
$$

Note that $x + y - 2/3 = z + 1/3 > 0$, and the cases $x = 1$ or $y = 1$ are inadmissible since $(1, 0, 0)$ and $(0, 1, 0)$ belong to $\partial\mathcal{E}$. Then $f(x, y, x+y-1) > 0$, and similarly in the cases $y+z-x = 1$ and $z+x-y = 1$. This means that the minimum of f on the boundary of $\mathcal{S}$ is exactly 0 (and attained at $(0, 0, 0)$ only), and in particular that $\min_{\mathcal{S}} f \leq 0$.

We seek the potential extrema of f in the interior of $\mathcal{S}$, for which we calculate its critical points. The gradient of f is

$$
\nabla f(x, y, z) = \begin{pmatrix} -2x - y - z + 1 + 3yz \\ -2y - x - z + 1 + 3xz \\ -2z - x - y + 1 + 3xy \end{pmatrix}.
$$

Cancelling this gradient leads to the three equations

$$
\begin{cases} (1 - 3y)z = 1 - 2x - y, \\ (1 - 3x)z = 1 - 2y - x, \\ 2z + x + y - 1 - 3xy = 0. \end{cases}
$$

If $y = 1/3$, then the first equation yields $x = 1/3$ and the third one yields $z = 1/3$, and $(1/3, 1/3, 1/3)$ is indeed an admissible solution of the gradient equations. At this point $f(1/3, 1/3, 1/3) = 4/9 > 0$, so $(1/3, 1/3, 1/3)$ is not a minimum of f on $\mathcal{S}$. If instead $y \neq 1/3$, then the first equation yields $z = (1 - 2x - y)/(1 - 3y)$. Plugging this into the second equation results in

$$
(1 - 3x)(1 - 2x - y) = (1 - 3y)(1 - 2y - x) \Leftrightarrow 3y^2 - 2y = 3x^2 - 2x.
$$

This equation (with unknown y) has solutions $y = x$ and $y = (2 - 3x)/3$. The solution $y = x$ leads to $z = (1 - 2x - y)/(1 - 3y) = 1$, and then the third equation leads to $3x^2 - 2x - 1 = 0$, whose solutions are $x = 1$ and $x = -1/3$. The latter is obviously not admissible, and neither is the former because it produces the solution $(1, 1, 1)$, which does not belong to $\mathcal{S} = \mathcal{P} \setminus \partial\mathcal{E}$. The solution $y = (2 - 3x)/3$ leads to $z = (1 - 2x - y)/(1 - 3y) = -1/3$, which is not admissible. From this discussion it follows that f does not attain its minimum in the interior of $\mathcal{S}$, and so $\min_{\mathcal{S}} f = 0$, as announced, with the minimum attained at $(0, 0, 0)$ only. □

A similar discussion does not appear to be feasible in dimension $m > 3$, because the set of constraints on the r_{ij} becomes very complex; characterizations are given by [9], [10] and [24], although they do not seem suited to the kind of calculations needed here. In particular, whether the variance-optimal pooled estimator is still indeed a convex combination remains an open question for $m > 3$.

C.3. About the difference between the test statistic Λ_n of equal tail indices and the test statistic of [17]

The key difference between the proposed test statistic Λ_n and the proposal of [17] is in the estimation of the variance component. Both have the form

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

where $\check{\mathbf{V}}_{\boldsymbol{c}}$ is an estimator of $\mathbf{V}_{\boldsymbol{c}}$. We use the estimator

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

i.e. we estimate the covariance matrix under the most general model where the γ_j are possibly different. This is natural because the log-likelihood ratio formulation requires an expression of the estimated covariance matrix that is common to the null and alternative models if the resulting deviance is to have a simple closed form. By contrast, from the description made in Section 3 of [17], the version used therein appears to be

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

The two tests are asymptotically equivalent under the null hypothesis. However, we recommend to at least report their results alongside each other, because $\overline{\mathbf{V}}_{\boldsymbol{c}}^{(\mathrm{KFL})}$ is not consistent under the alternative hypothesis, so that our proposed test can be expected to be more powerful in certain cases.

C.4. About Remark 8

Recall the notation $\mu^{(\mathrm{Hill})} = \lambda/(1-\rho)$. If $\mu^{(\mathrm{Var})}$ denotes the asymptotic bias of $\widehat{\gamma}_n(\widetilde{\omega}_n^{(\mathrm{Var})})$ then

$$\frac{\mu^{(\mathrm{Hill})}}{\mu^{(\mathrm{Var})}} = \left(\sum_{j=1}^m \frac{1}{c_j}\right)^{1-\rho} \left(\sum_{j=1}^m \frac{1}{b_j}\right)^{\rho} \Bigg/ \sum_{j=1}^m \left(\frac{1}{c_j}\right)^{1-\rho} \left(\frac{1}{b_j}\right)^{\rho}.$$

If $\rho < 0$, we use the Hölder inequality $\sum_{j=1}^m x_j y_j \leq (\sum_{j=1}^m x_j^p)^{1/p} (\sum_{j=1}^m y_j^q)^{1/q}$, valid for any nonnegative real numbers $x_1, y_1, \ldots, x_m, y_m$ and $p, q \geq 1$ such that $1/p + 1/q = 1$. Apply this inequality for $x_j = b_j^{\rho/(1-\rho)}$, $y_j = c_j^{-1} b_j^{-\rho/(1-\rho)}$, $p = -(1-\rho)/\rho$ and $q = 1-\rho$ to get

$$\sum_{j=1}^m \frac{1}{c_j} \leq \left(\sum_{j=1}^m \frac{1}{b_j}\right)^{-\rho/(1-\rho)} \left(\sum_{j=1}^m \left(\frac{1}{c_j}\right)^{1-\rho} \left(\frac{1}{b_j}\right)^{\rho}\right)^{1/(1-\rho)} \Leftrightarrow \frac{\mu^{(\mathrm{Hill})}}{\mu^{(\mathrm{Var})}} \leq 1.$$

Equality between the asymptotic bias terms holds if and only if, with the above notation, the vectors $(x_1^p, \ldots, x_m^p)$ and $(y_1^q, \ldots, y_m^q)$ are linearly dependent, which is immediately seen to be equivalent to $b_j/c_j = K$, a constant independent of j, and therefore $b_j/c_j = b_1/c_1 = 1$.

C.5. About the asymptotic variance in Theorem 6

We remark that the asymptotic variance γ^2 obtained in Theorem 2 of [3], with the naive weights $\omega_j = 1/m$ for $j \in \{1, \ldots, m\}$, is not correct. In fact, the asymptotic variance is $v\gamma^2$, where in general $v > 1$, see Theorem 6. This higher variance is not surprising because, in the case of unbalanced k_j, machines with the lowest k_j tend to provide less information than those with the largest k_j, and therefore a loss of information should be expected in comparison with the case where all the k_j are equal. This insight can be checked by considering, for example, a simple situation where X is purely Pareto distributed with tail index γ, the number m of machines is even, and $k_j = 1$ for j odd and $k_j = 2$ for j even. In this situation, each $\widehat{\gamma}_j(k_j)$ is in fact simply an exponential random variable with mean γ and variance γ^2 for j odd, and a mean of two independent such random variables when j is even. Consequently

$$\operatorname{Var}\left(\sum_{j=1}^{m}\sqrt{k}(\widehat{\gamma}_n(1/m,\ldots,1/m)-\gamma)\right) = \frac{3m/2}{m^2}\left(\sum_{l=1}^{m/2}\gamma^2+\sum_{l=1}^{m/2}\frac{\gamma^2}{2}\right) = \frac{9\gamma^2}{8}.$$

This matches our result since $\sum_{j=1}^m k_j = 3m/2$ and $\sum_{j=1}^m \omega_j^2/k_j = 3/4m$, so that $v = 9/8$ indeed.

We illustrate further this phenomenon by providing a numerical illustration of the weak convergence of the distributed tail index estimator, in the case $m \to \infty$, to the asymptotic distribution given in Theorem 6. We do so using the following procedure:

- Let m vary in $\{200; 2{,}000; 20{,}000; 200{,}000\}$, and set $N = 50{,}000$. Let $n_m = (\lfloor \log(m/2) \rfloor)^2$.
- Simulate m i.i.d. samples of n_m data points from a Burr distribution with tail index 1 and second-order parameter -1, *i.e.* the datasets are generated with the distribution function

  $$F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho},\ x > 0,\ \text{with } \gamma = -\rho = 1.$$

  Within the first (resp. last) $m/2$ samples, compute the Hill estimate $\widehat{\gamma}_j(k_j)$ with $k_j = 1$ (resp. $k_j = 6$). Then, compute the naive distributed estimate based on these m subsamples and record the total effective sample size $k = (m/2) \times 1 + (m/2) \times 6 = 7m/2$. Repeat these steps N times.
- Draw the histogram of these N distributed estimates. Calculate the mean $\overline{\gamma}$ of the estimates and superimpose to the histogram:
  - The normal density curve with mean $\overline{\gamma}$ and variance $1/k$,
  - The normal density curve with mean $\overline{\gamma}$ and variance $49/(24k)$.

  The asymptotic variance in the first of these normal density curves is the one announced in Theorem 2 of [3]. The asymptotic variance in the second of these curves is the one provided by our Theorem 6 because, with $\omega_j = 1/m$,

  $$k\sum_{j=1}^{m}\frac{\omega_j^2}{k_j} = \left(\frac{m}{2}\times 1+\frac{m}{2}\times 6\right)\times\frac{1}{m^2}\left(\frac{m}{2}\times 1+\frac{m}{2}\times\frac{1}{6}\right) = \frac{49}{24}.$$

Results are given in Figure C.1. It is clear from this figure that Theorem 6 provides the correct asymptotic variance for the naive distributed estimator, in a setup where the assumptions of Theorem 6 are satisfied, and that the asymptotic variance announced in Theorem 2 in [3] is in fact not correct. Note that the histograms and density fits are not centered at the true value $\gamma = 1$. This is expected and due to the bias incurred by using the Hill estimator, which is seen to decrease as m increases.

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

We assume that the m−dimensional i.i.d. random vectors $\boldsymbol{X}_i$ follow one of the four models listed below. Two of these four models are based on Archimedean copulae, which we briefly recall. Further details

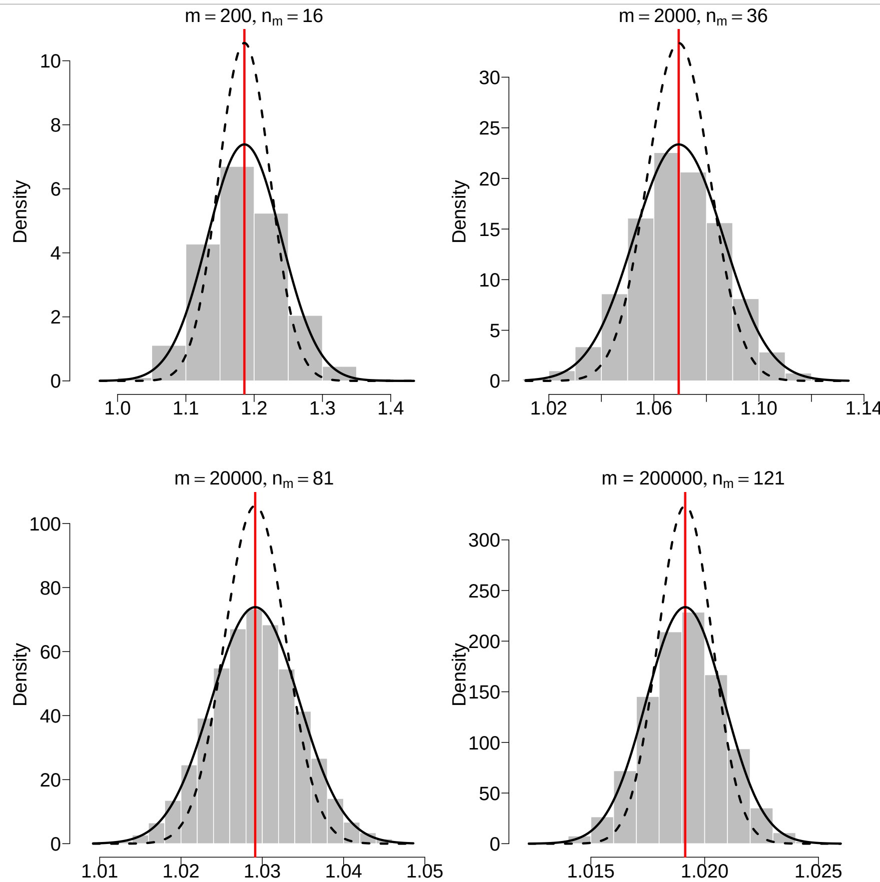

Figure C.1: Empirical distribution of the naive distributed Hill estimator in the simulation setup described in Section C.5. Top left: $m = 200$, top right: $m = 2{,}000$, bottom left: $m = 20{,}000$, bottom right: $m = 200{,}000$. The red vertical line represents the mean of the distributed Hill estimates, 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. $49/(24k)$).

can be found in [15]. Let $\varphi : (0,1] \longrightarrow [0,\infty)$ be a convex and strictly decreasing function with $\varphi(1) = 0$ and $\varphi(t) \uparrow \infty$ as $t \downarrow 0$. The Archimedean copula in dimension m with generator φ is the m-dimensional distribution function C with uniform marginals defined by

$$C(\boldsymbol{u}) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_m)),\ \boldsymbol{u} = (u_1, \ldots, u_m) \in [0,1]^m.$$

The Archimedean families we consider are, first, the Clayton family, defined through the generator $\varphi(u) = \eta^{-1}(u^{-\eta} - 1)$ for $\eta > 0$. Here the components of $\boldsymbol{u}$ become independent for $\eta \to 0$, and completely dependent for $\eta \to \infty$. We also consider the Gumbel family, defined through the generator $\varphi(u) = (-\log(u))^{\eta}$ for $\eta \geq 1$, with $\eta = 1$ representing the case of independent variables and $\eta \to \infty$ the case of perfectly dependent variables. Our experiments are based on the below models for $\boldsymbol{X} = (X_1, \ldots, X_m)$, with $2 \leq m \leq 5$ and in each case $n = \sum_{j=1}^{m} n_j = 1{,}000$.

(a) [Gaussian-Student model] Let $m = 2$, and assume that $\boldsymbol{U}$ follows a 2−dimensional Gaussian copula with correlation parameter $\rho_{1,2} = 0.8$. Take $X_j = |F_\nu^{-1}(U_j)|$ where F_ν is the Student distribution function with $\nu = 1$ degree of freedom. Then $\boldsymbol{X}$ has (absolute value) Student marginal distributions with tail index 1 and a Gaussian copula dependence structure. Here we set $n_1 = 50$ and $n_2 = 950$.

(b) [Multivariate Student model] Let $m = 3$, and assume that $\boldsymbol{X} = (|Y_1|, |Y_2|, |Y_3|)^\top$ where $\boldsymbol{Y} = (Y_1, Y_2, Y_3)^\top$ follows a 3−dimensional zero-mean multivariate Student distribution with $\nu = 1$ degree of freedom and a scale matrix $\boldsymbol{\rho}$ defined elementwise as ($\rho_{1,1} = \rho_{2,2} = \rho_{3,3} = 1$, $\rho_{1,2} = 0.8$, $\rho_{1,3} = 0.6$, $\rho_{2,3} = 0.4$). Here we set $n_1 = 50$, $n_2 = 450$ and $n_3 = 500$.

(c) [Clayton-Fréchet model] Let $m = 4$, and assume that $\boldsymbol{U}$ follows a 4−dimensional Clayton copula with dependence parameter $\eta = 2$. Take $X_j = -1/\log(U_j)$. Then $\boldsymbol{X}$ has Fréchet marginal distributions with tail index 1 and a Clayton copula dependence structure. Here we set $n_1 = 150$, $n_2 = 200$, $n_3 = 250$ and $n_4 = 400$.

(d) [Gumbel-Fréchet model] Let $m = 5$, and assume that $\boldsymbol{U}$ follows a 5−dimensional Gumbel copula with dependence parameter $\eta = 2$. Take $X_j = -1/\log(U_j)$. Then $\boldsymbol{X}$ has Fréchet marginal distributions with tail index 1 and a Gumbel copula dependence structure. Here we set $n_1 = 150$, $n_2 = 150$, $n_3 = 200$, $n_4 = 200$ and $n_5 = 300$.

The univariate margins of the above multivariate distributions have the same tail index $\gamma = 1$. Furthermore, the elements of $\boldsymbol{X}$ are asymptotically independent in models (a) and (c), in the sense that all pairwise tail copulae are equal to zero, and asymptotically dependent in models (b) and (d).

We simulate $N = 1{,}000$ samples of i.i.d. observations from the models (a)-(d). For each simulated dataset, we estimate the tail index γ using our proposed pooled Hill estimators, and then the extreme quantile at level $1 - p = 1 - 1/n = 0.999$ using pooled Weissman estimators. The pooling weights considered are naive weights (the standard mean across subsamples, denoted by Pool-NAIVE), variance-optimal (Pool-AVAR) and AMSE-optimal weights (Pool-AMSE-NP) defined together right before Corollary 1, and AMSE-optimal weights obtained by pooling second-order estimates (Pool-AMSE, see the comment below Corollary 2). Regarding pooled Weissman estimators, we consider the geometrically pooled estimator $\widehat{q}_n^\star(1 - p|\boldsymbol{\omega})$ with $\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}_n$ being the aforementioned naive, variance-optimal and AMSE-optimal weights. We also consider the simple arithmetic mean of Weissman estimators, namely $\overline{q}_n^\star(1 - p|1/m, \ldots, 1/m) = \frac{1}{m}\sum_{j=1}^m \widehat{q}_j^\star(1 - p|k_j)$ (denoted by Pool-NAIVE-A). These estimators are compared to the benchmark Hill and Weissman estimators on the pooled sample, as appropriate. We repeat these estimation steps for 50 equally spaced values of the effective sample fraction $k/n \equiv (k_j/n_j) \times 100\% \in (1\%, 50\%)$ in each subsample (assumed to be the same for each j). We compute empirical Mean Squared Errors (MSEs) for each tail index estimator, and empirical relative MSEs of each log-quantile estimator. We also compute a Monte Carlo approximation of the actual coverage probability of the asymptotic confidence interval with 95% nominal level relative to each estimator. The asymptotic confidence intervals of the proposed pooled estimators are presented and studied in Corollaries 3 and 4. For the Hill estimator, see Theorem 3.2.5, p.74 in [7] whereby we assume that the asymptotic distribution is normal with mean 0 and variance γ^2/k. Note that this is a misspecification of the actual asymptotic variance when asymptotic dependence between subsamples is present: to the best of our knowledge there is no rigorous theoretical result available about the asymptotic normality of the Hill estimator in this context. For the Weissman estimator, we reformulate Theorem 4.3.8, p.138 in [7] as

$$\frac{\sqrt{k}}{\log(k/(np))} \log \frac{\widehat{q}_n^{\star,(\mathrm{Hill})}(1 - p|k)}{q(1 - p)} \xrightarrow{d} \mathcal{N}(0, \gamma^2)$$

from which we construct the asymptotic confidence interval for $q(1 - p)$ in the obvious way. [We assume here again that the asymptotic bias is 0, and likewise, using the above convergence results in a misspecification of the asymptotic variance when asymptotic dependence between subsamples is present.] Results are given in Figures D.2–D.5.

To assess the performance of the proposed likelihood ratio-based test statistics, we alter models (a)–(d) as follows:

(a') Generate $\boldsymbol{U}$ and X_2 as in model (a), and take $X_1 = |F^{-1}_{1/\gamma_1}(U_1)|$.

(b') Generate (Y_1, Y_2, Y_3) and X_2, X_3 as in model (b), and take $X_1 = |Y_1|^{\gamma_1}$.

(c') Generate $\boldsymbol{U}$ and X_2, X_3, X_4 as in model (c), and take $X_1 = (-\log(U_1))^{-\gamma_1}$.

(d') Generate $\boldsymbol{U}$ and X_2, X_3, X_4, X_5 as in model (d), and take $X_1 = (-\log(U_1))^{-\gamma_1}$.

In each of these models we let γ_1, the tail index of the first marginal distribution, vary between 0.2 and 5, and we carry out the tail homogeneity test based on the test statistic Λ_n at the nominal type I error rate 5%. We represent in Figure D.6 the rejection rate of the test as a function of γ_1. In Figure D.7, we represent these same rejection rates, only with balanced sample sizes in each model, that is, $(n_1, n_2) = (500, 500)$ 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'). We omit the numerical calculation of the rejection rate of the tail homoskedasticity test based on the test statistic $L_n(p)$ at level $1 - p = 0.999$, for which the conclusions are qualitatively the same.

Finally, to get an idea about the performance of the estimators and their related confidence intervals when the total sample size n is small, we repeated our experiments with subsample sizes equal to 40% of their values in models (a)–(d), resulting in a lower total sample size $n = \sum_{j=1}^{m} n_j = 400$. Results are provided in Figures D.8–D.11.

D.1.2. Distributed inference of extreme values

In this section we assume that the data are independent within and across samples, *i.e.* we work in the distributed inference context. We first consider the dimensions $m = 5, 10, 20$ and the following models:

(e) [Burr model] Let $m = 5$ and the X_j be independent having a Burr distribution, that is, with common distribution function $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$, $x > 0$, where $\gamma = 1$ and $\rho = -1$. Here we set $n_1 = 150$, $n_2 = 150$, $n_3 = 200$, $n_4 = 200$ and $n_5 = 300$.

(f) [Student model] Let $m = 10$ and the $X_j = |Y_j|$ be independent with the Y_j having a Student distribution with $\nu = 1$ degree of freedom. Here we set $n_j = 50$ for each $j \in \{1, 2, \ldots, 9\}$ and $n_{10} = 550$.

(g) [Fréchet model] Let $m = 20$ and the X_j be independent having a common Fréchet distribution with tail index 1. Here we set $n_j = 50$ for each $j \in \{1, 2, \ldots, 20\}$.

We compare the same estimators as in Section D.1.1, with the same setup, although the pooling weights are calculated assuming independence between subsamples. In other words, we use the expression of variance-optimal weights and AMSE-optimal weights with pooled second-order parameter estimates in Section 3.2, and the expression of AMSE-optimal weights in Section 2.2 with the constraint that the $\widehat{R}_{j,\ell}$ are all 0 when $j \neq \ell$. Results are provided in Figures D.12–D.14.

We further examine the advantage of using AMSE-optimal weights in the following illustrative case in dimension 2:

(h) [Burr model, unbalanced sample sizes, equal effective sample sizes] Let $m = 2$ and X_1, X_2 be independent having a Burr distribution with parameters $\gamma = 1$ and $\rho = -1$. We choose $(n_1, n_2) \in \{(200, 800), (100, 900), (50, 950)\}$, and $k_1 = k_2$.

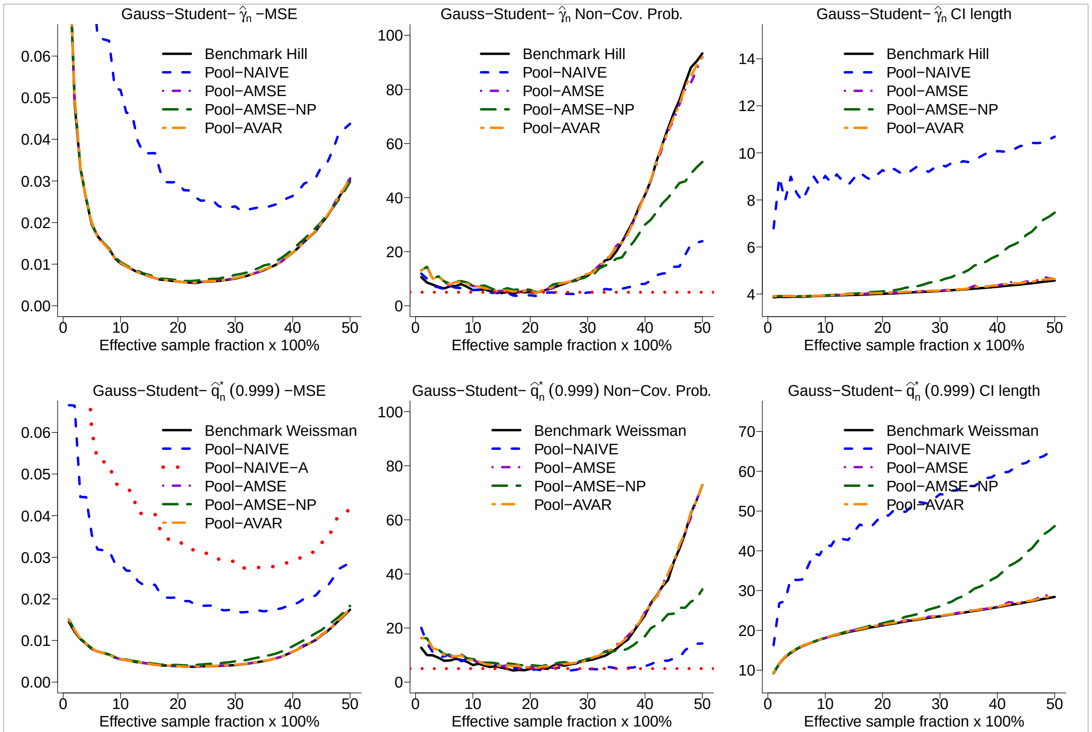

Figure D.2: Simulation results, general pooling setting, Gaussian-Student model (a). Top row: tail index estimation; bottom row: extreme quantile estimation at level $1-p=0.999$. Left panels: MSE of the point estimators; middle panels: non-coverage probability of the asymptotic confidence intervals, 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 $\log q(1-p)$.

Again $N = 1{,}000$ samples of i.i.d. copies from this model are simulated and the same competing tail index estimators are compared with the same points of comparison (we omit the details about extreme quantile estimators, where the conclusion is identical). Results are presented in Figure D.15.

We finally assess the benefit of using, in the context of extreme quantile estimation, the assumption $(\mathcal{H})$ of tail homoskedasticity when it is valid, *i.e.* we compare the geometrically pooled extreme quantile estimator with variance-optimal weights to the subsample Weissman estimator in which the tail index estimator is taken to be the variance-optimal pooled version, that is, the estimator $\widehat{q}_1^\star(1-p|k_1,\widehat{\omega}_n)$ where $\widehat{\omega}_n$ denotes variance-optimal weights. We do so in the following models:

(Q-a) [Burr model, dimension $m=2$, $\rho=-1$] Let $m=2$ and X_1, X_2 be independent with common Burr distribution having parameters $\gamma=1$ and $\rho=-1$. Here we take $n_1 = 900, 500, 100$ and $n_2 = n - n_1 = 100, 500, 900$.

(Q-b) [Burr model, dimension $m=2$, $\rho=-1/2$] As in the previous model, with $\rho=-1/2$.

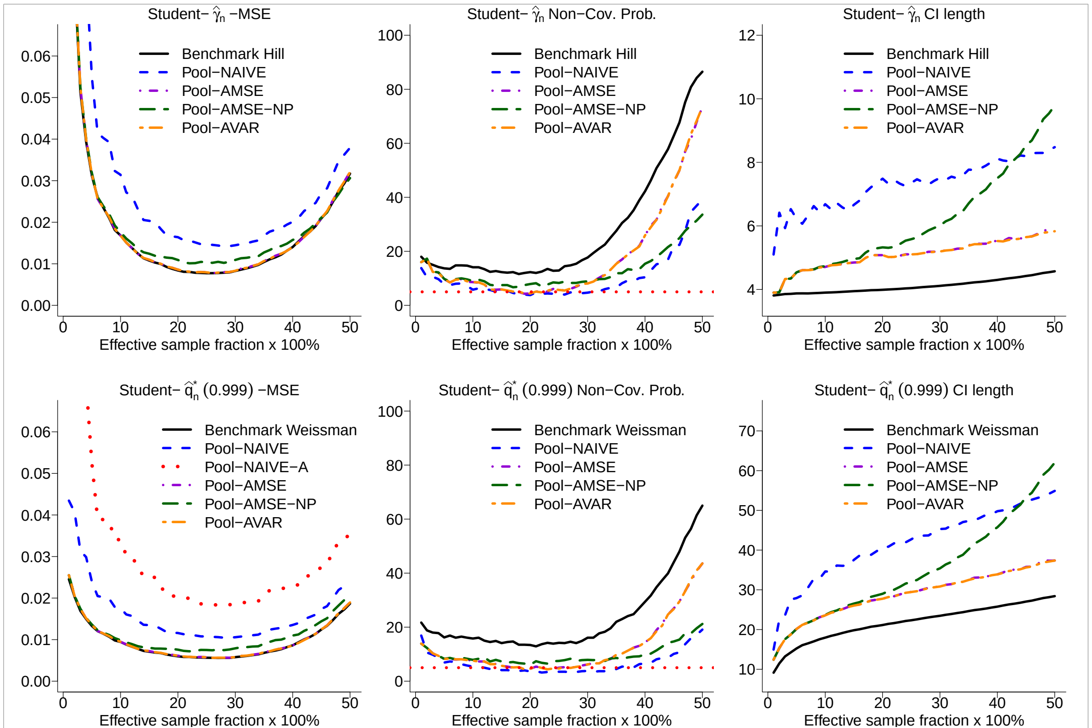

Figure D.3: As in Figure D.2, in the Multivariate Student model (b).

(Q-c) [Burr model, dimension $m = 5$] As in model (e), but we take $n_1 = 800, 500, 100$ and $n_2 = n_3 = n_4 = n_5 = (n - n_1)/4 = 50, 125, 225$.

Results are represented in Figure D.16. In this series of graphs, the brown curve labeled Pool-AVAR-Shape represents the subsample Weissman estimator $\widehat{q}_1^{\star}(1 - p|k_1, \widehat{\omega}_n)$.

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

We examine the performance of the proposed pooled tail index estimators in conjunction with filtering in location-scale models, in the following two examples:

(AR-1-a) [Gaussian-Student innovations] Let $m = 2$, and assume that $\boldsymbol{U}$ follows a 2−dimensional Gaussian copula with correlation parameter $\rho_{1,2} = 0.8$. Take $\varepsilon_j = F_\nu^{-1}(U_j)$ where F_ν is the Student distribution function with $\nu = 1$ degree of freedom. The observations $(X_{t,1}, X_{t,2})$ are taken from the pair of AR(1) processes defined recursively as $(X_{t+1,1}, X_{t+1,2}) = \phi(X_{t,1}, X_{t,2}) + \sqrt{1-\phi^2}(\varepsilon_{t+1,1}, \varepsilon_{t+1,2})$, where the $(\varepsilon_{t,1}, \varepsilon_{t,2})$ are i.i.d. replications of $(\varepsilon_1, \varepsilon_2)$ and $\phi = 0.5$. In this model we consider the two situations $(n_1, n_2) = (500, 500)$ where sample sizes are balanced and $(n_1, n_2) = (250, 750)$.

(AR-1-b) [Multivariate Student innovations] Let $m = 3$, and assume that $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)^\top$ follows a 3−dimensional zero-mean multivariate Student distribution with $\nu = 1$ degree of freedom and a scale matrix $\boldsymbol{\rho}$ defined elementwise as $(\rho_{1,1} = \rho_{2,2} = \rho_{3,3} = 1,\ \rho_{1,2} = 0.8,\ \rho_{1,3} = 0.6,\ \rho_{2,3} = 0.4)$. The observations are taken from the trivariate AR(1) process $(X_{t+1,1}, X_{t+1,2}, X_{t+1,3}) = \phi(X_{t,1}, X_{t,2}, X_{t,3}) + \sqrt{1-\phi^2}(\varepsilon_{t+1,1}, \varepsilon_{t+1,2}, \varepsilon_{t+1,3})$, where the $(\varepsilon_{t,1}, \varepsilon_{t,2}, \varepsilon_{t,3})$ are i.i.d. replications of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and

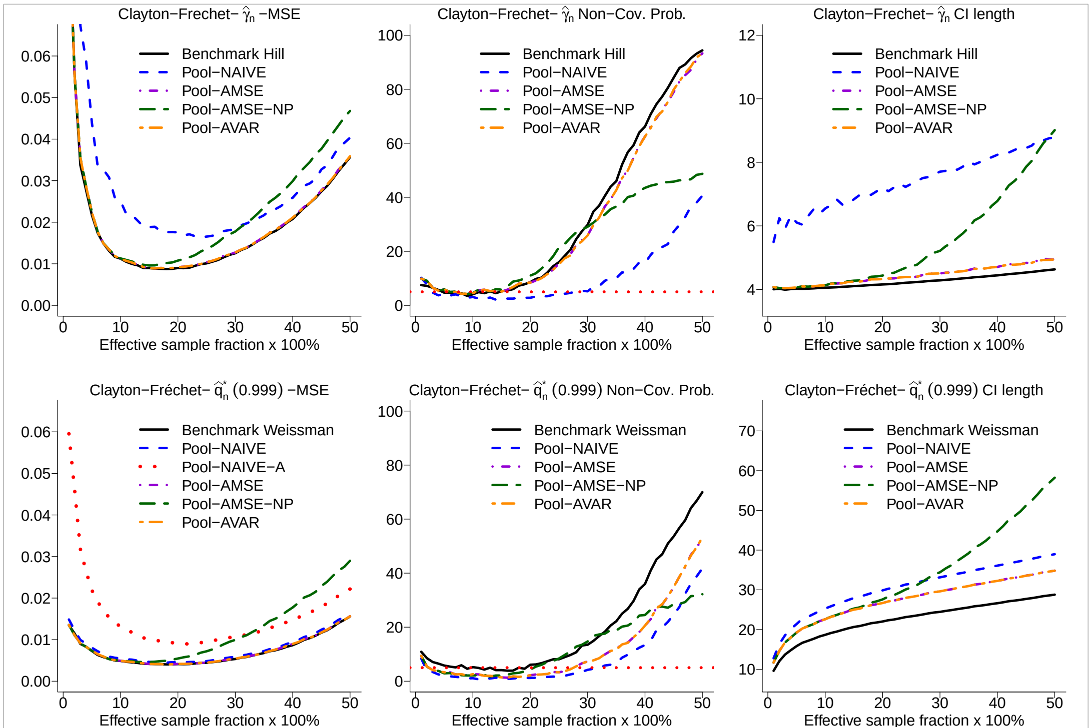

Figure D.4: As in Figure D.2, in the Clayton-Fréchet model (c).

$\phi = 0.5$. In this model we consider the two situations $(n_1, n_2, n_3) = (333, 333, 334)$ where sample sizes are balanced and $(n_1, n_2, n_3) = (200, 400, 400)$.

In both cases we first filter each time series $(X_{t,j})$ using the correct AR(1) model and a Student maximum likelihood estimator of ϕ, computed using the function `ugarchfit` of the R package `rugarch`. This yields subsamples of residuals $\widehat{\varepsilon}_{t,j}$, to which we apply the tail index estimators of Section D.1.1. Since standard tail index estimation methods have higher asymptotic variance in linear autoregressive models [see *e.g.* 8], it is interesting here to evaluate the orders of magnitude of the variance, bias, and MSE of the resulting pooled estimates constructed on filtered data, as a way to see whether they roughly behave as if they were using independent and identically distributed data in the first place. Results are displayed in Figures D.17 and D.18.

Finally, and specifically to assess the price of misspecification, we keep this exact same methodology in the following two extra scenarios:

(AR-2-a) [Gaussian-Student innovations] As in model (AR-1-a), but $(X_{t,1}, X_{t,2})$ are taken from the pair of AR(2) processes $(X_{t+1,1}, X_{t+1,2}) = \phi_1(X_{t,1}, X_{t,2}) + \phi_2(X_{t-1,1}, X_{t-1,2}) + (\varepsilon_{t+1,1}, \varepsilon_{t+1,2})$, where the $(\varepsilon_{t,1}, \varepsilon_{t,2})$ are i.i.d. replications of $(\varepsilon_1, \varepsilon_2)$ and $(\phi_1, \phi_2) = (0.6, -0.28)$.

(AR-2-b) [Multivariate Student innovations] As in model (AR-1-b), but $(X_{t,1}, X_{t,2}, X_{t,3})$ are taken from the trivariate AR(2) process $(X_{t+1,1}, X_{t+1,2}, X_{t+1,3}) = \phi_1(X_{t,1}, X_{t,2}, X_{t,3}) + \phi_2(X_{t-1,1}, X_{t-1,2}, X_{t-1,3}) + (\varepsilon_{t+1,1}, \varepsilon_{t+1,2}, \varepsilon_{t+1,3})$, where the $(\varepsilon_{t,1}, \varepsilon_{t,2}, \varepsilon_{t,3})$ are i.i.d. replications of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and $(\phi_1, \phi_2) = (0.6, -0.28)$.

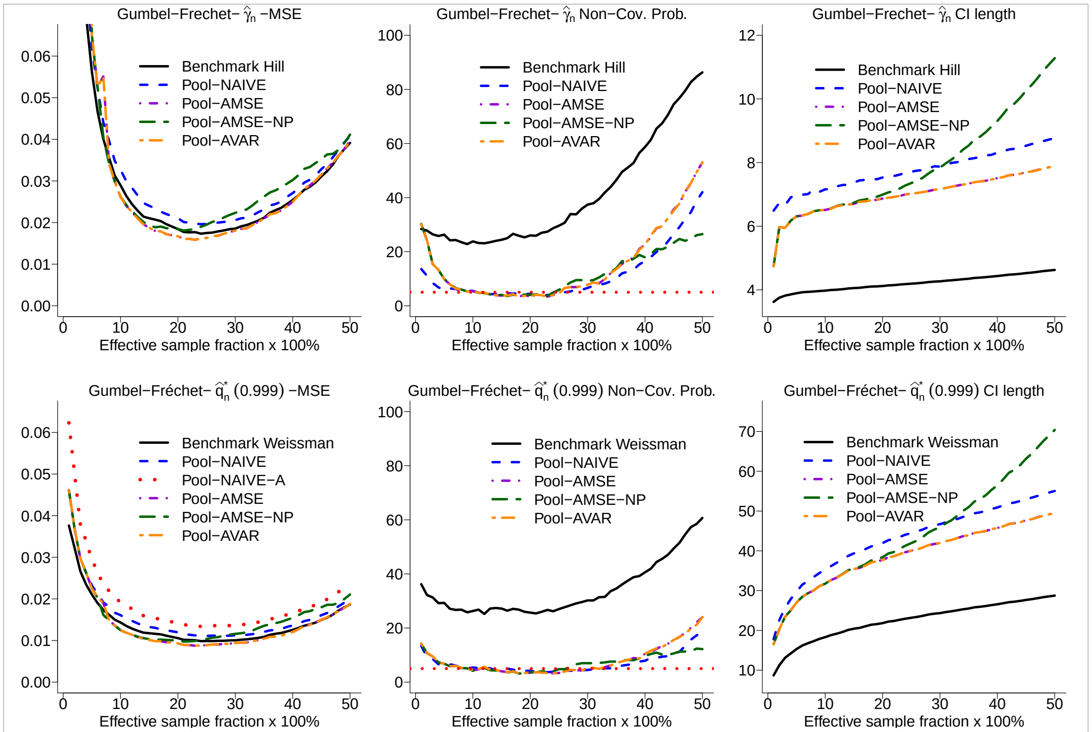

Figure D.5: As in Figure D.2, in the Gumbel-Fréchet model (d).

Both of the last two models have an AR(2) structure, but are estimated as AR(1), meaning that the dynamics of the process are misspecified. In particular, the estimator of ϕ_1 thus obtained and used at the filtering step in order to produce residuals is not consistent. Results are displayed in Figures D.19 and D.20.

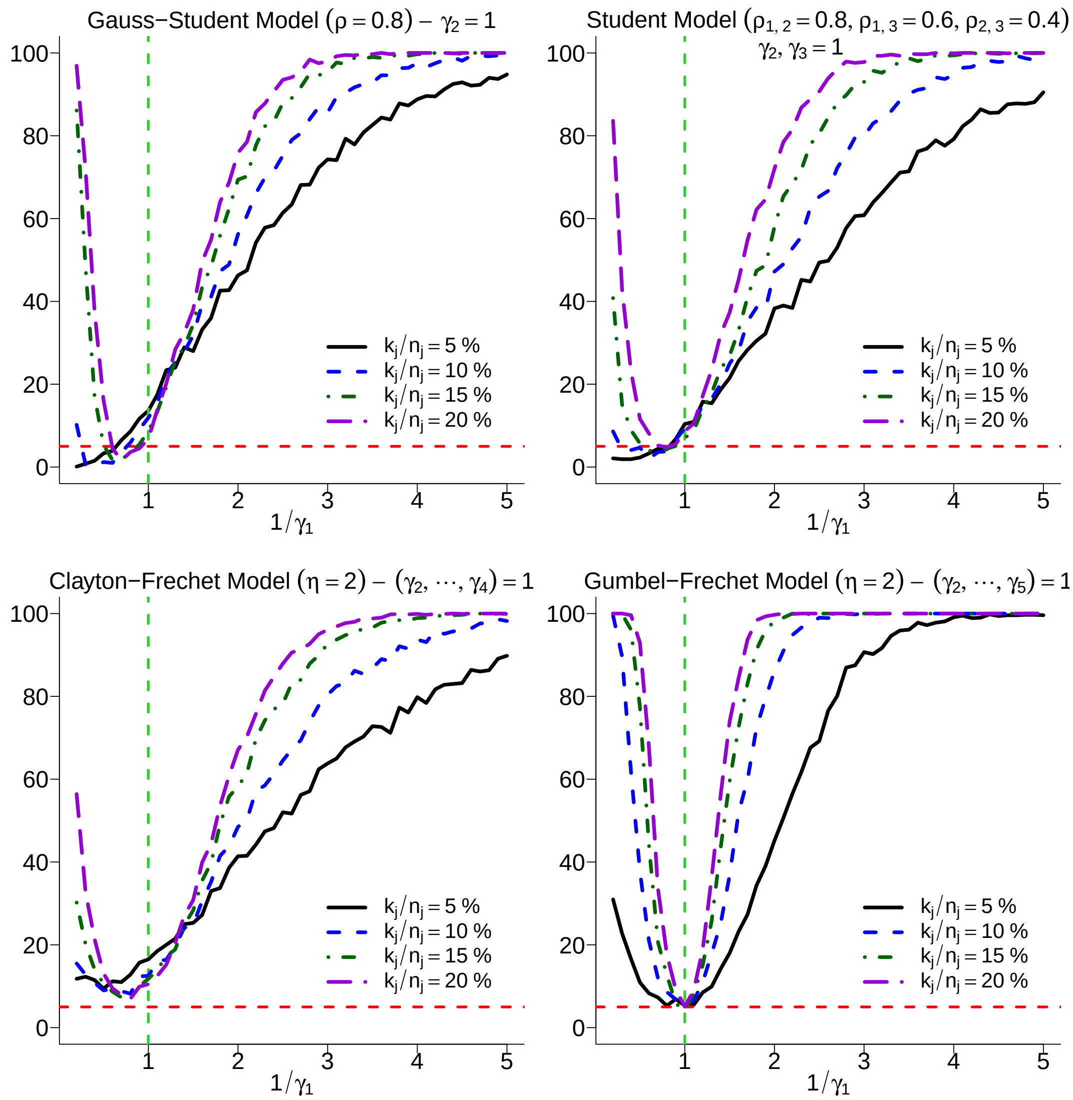

Figure D.6: Simulation results, general pooling setting, rejection rate of the test of tail homogeneity based on Λ_n with nominal type I error equal to 5%. Top left panel: Gaussian-Student model (a'), top right panel: Multivariate Student model (b'), bottom left panel: Clayton-Fréchet model (c'), bottom 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.

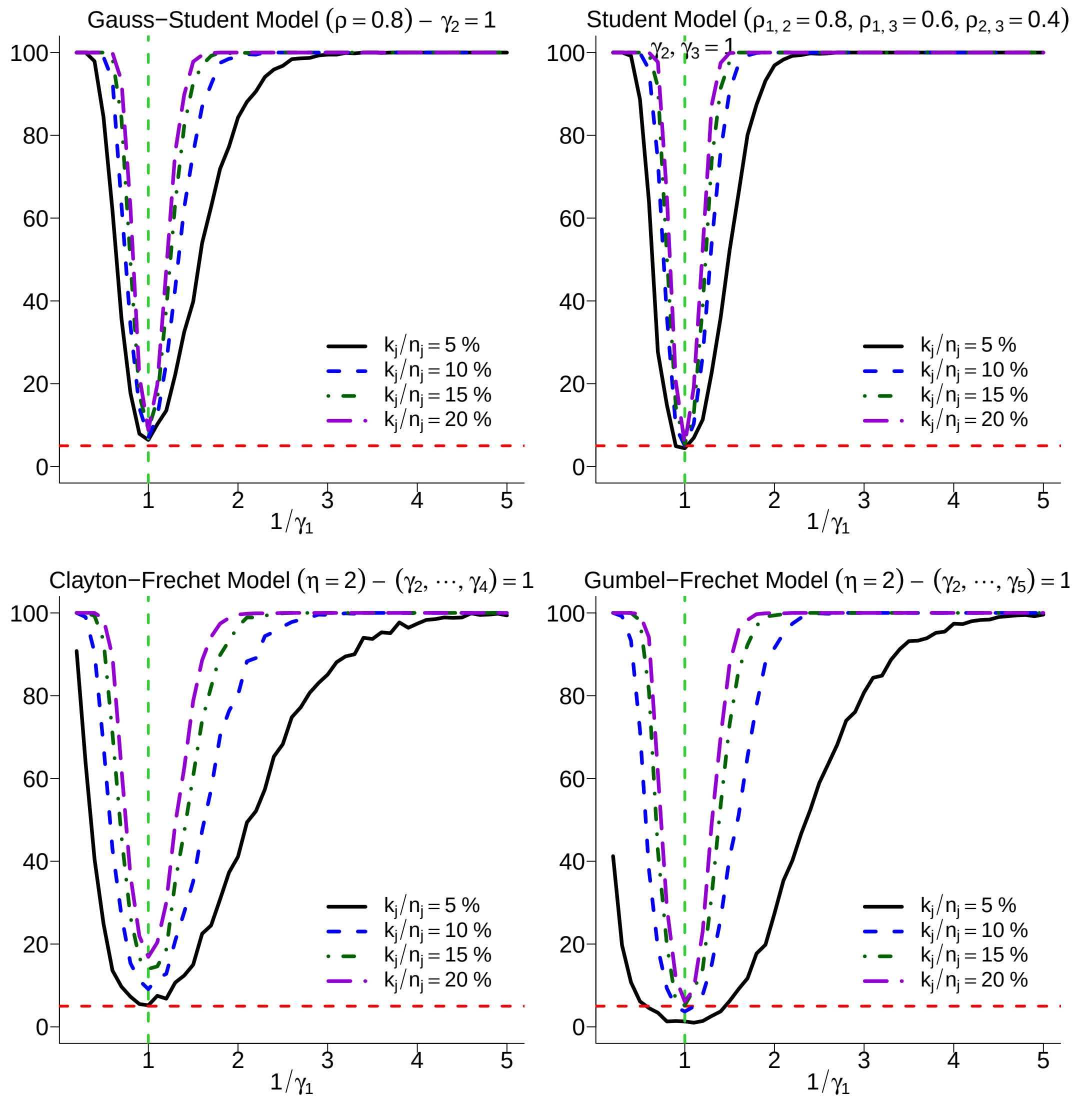

Figure D.7: As in Figure D.6, with equal subsample sizes in each model, namely $(n_1,n_2) = (500,500)$ in model (a'), $(n_1,n_2,n_3) = (333,333,334)$ in model (b'), $(n_1,n_2,n_3,n_4) = (250,250,250,250)$ in model (c') and $(n_1,n_2,n_3,n_4,n_5) = (200,200,200,200,200)$ in model (d').

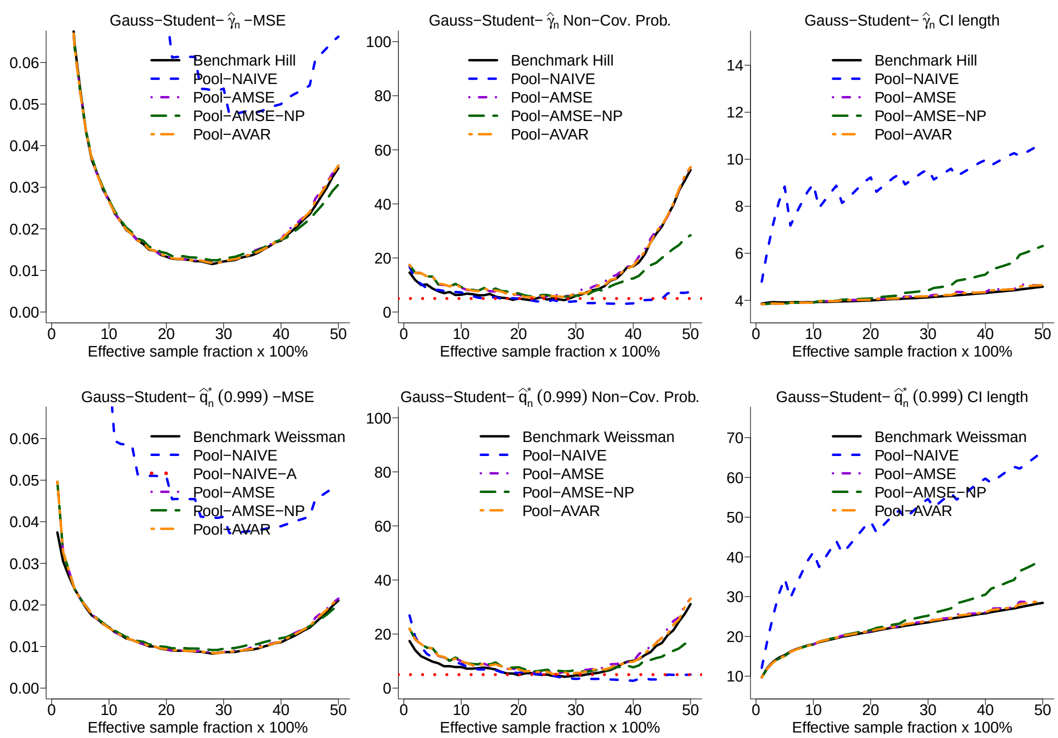

Figure D.8: As in Figure D.2, in the Gaussian-Student model (a), with $(n_1, n_2) = (20, 380)$.

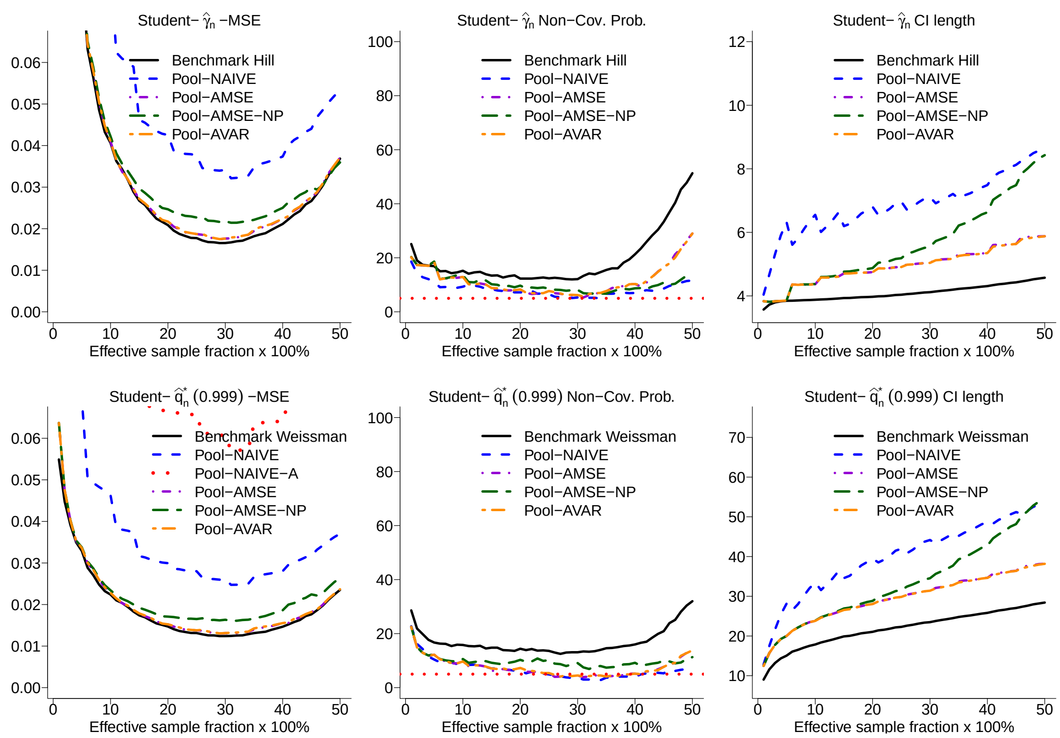

Figure D.9: As in Figure D.2, in the Multivariate Student model (b), with $(n_1, n_2, n_3) = (20, 180, 200)$.

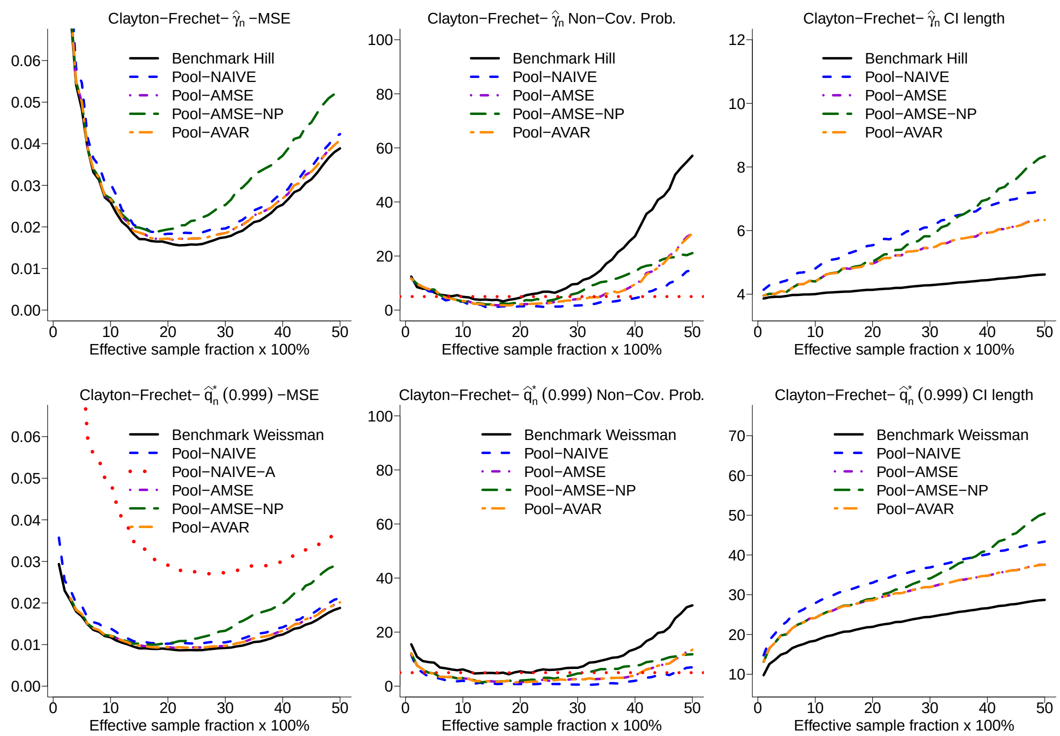

Figure D.10: As in Figure D.2, in the Clayton-Fréchet model (c), with $(n_1, n_2, n_3, n_4) = (60, 80, 100, 160)$.

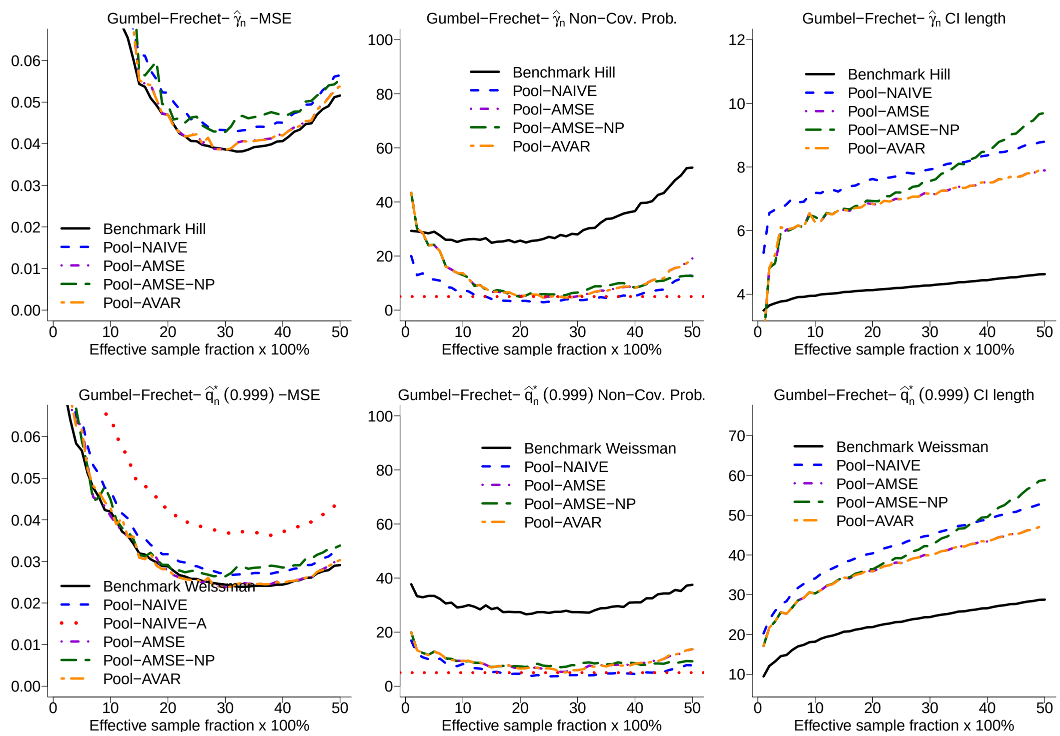

Figure D.11: As in Figure D.2, in the Gumbel-Fréchet model (d), with $(n_1, n_2, n_3, n_4, n_5) = (60, 60, 80, 80, 120)$.

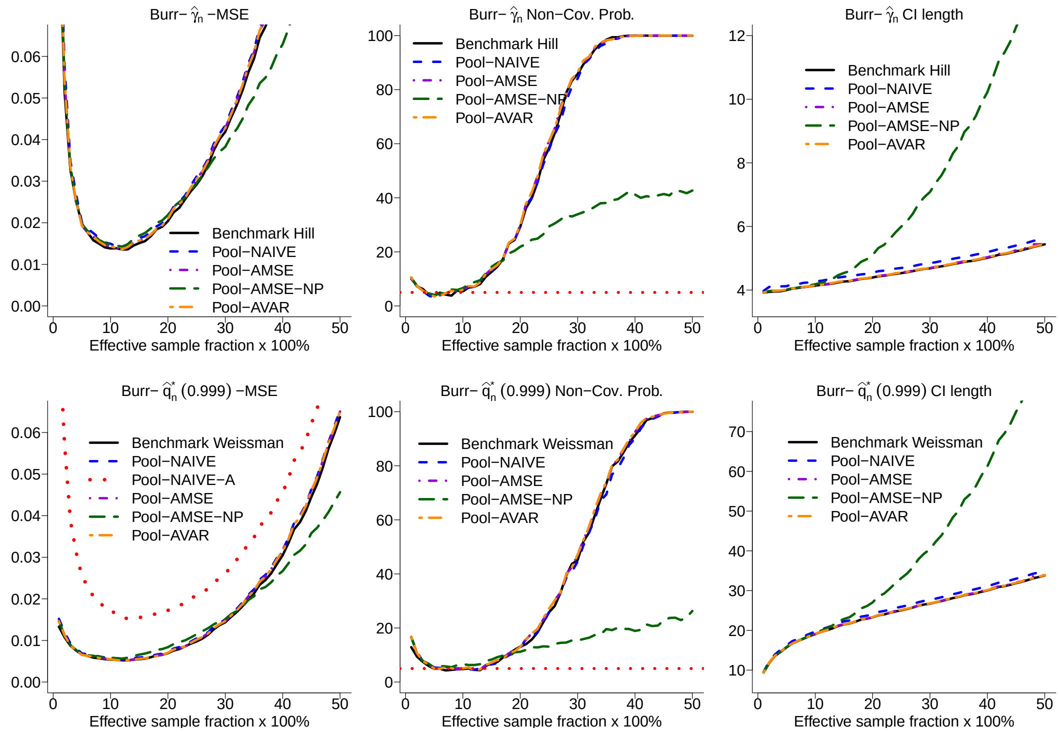

Figure D.12: Simulation results, distributed inference context, Burr model (e). Top row: tail index estimation; bottom row: extreme quantile estimation at level $1 - p = 0.999$. Left panels: MSE of the point estimators; middle panels: non-coverage probability of the asymptotic confidence intervals, where 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 $\log q(1 - p)$.

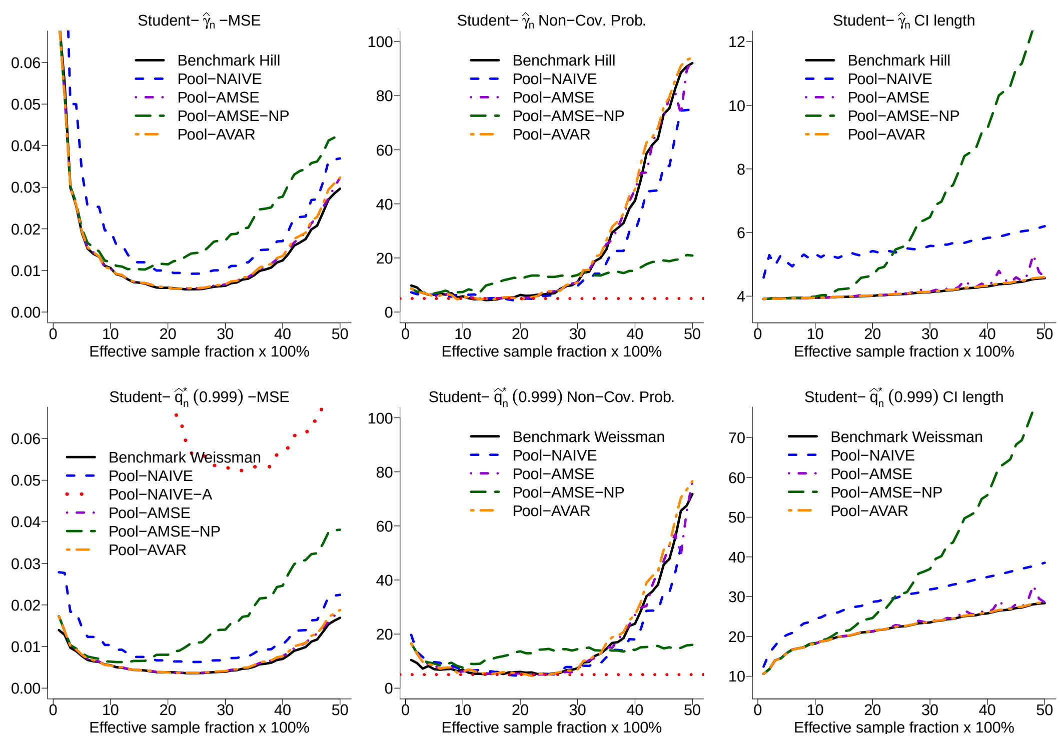

Figure D.13: As in Figure D.12, in the Student model (f).

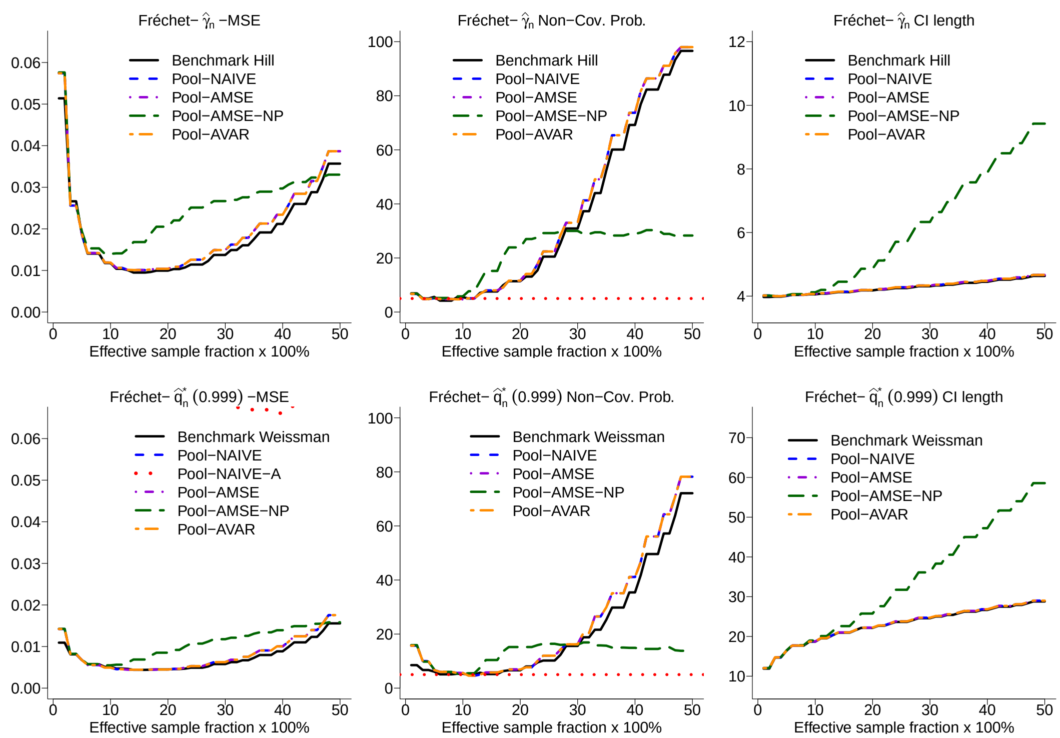

Figure D.14: As in Figure D.12, in the Fréchet model (g).

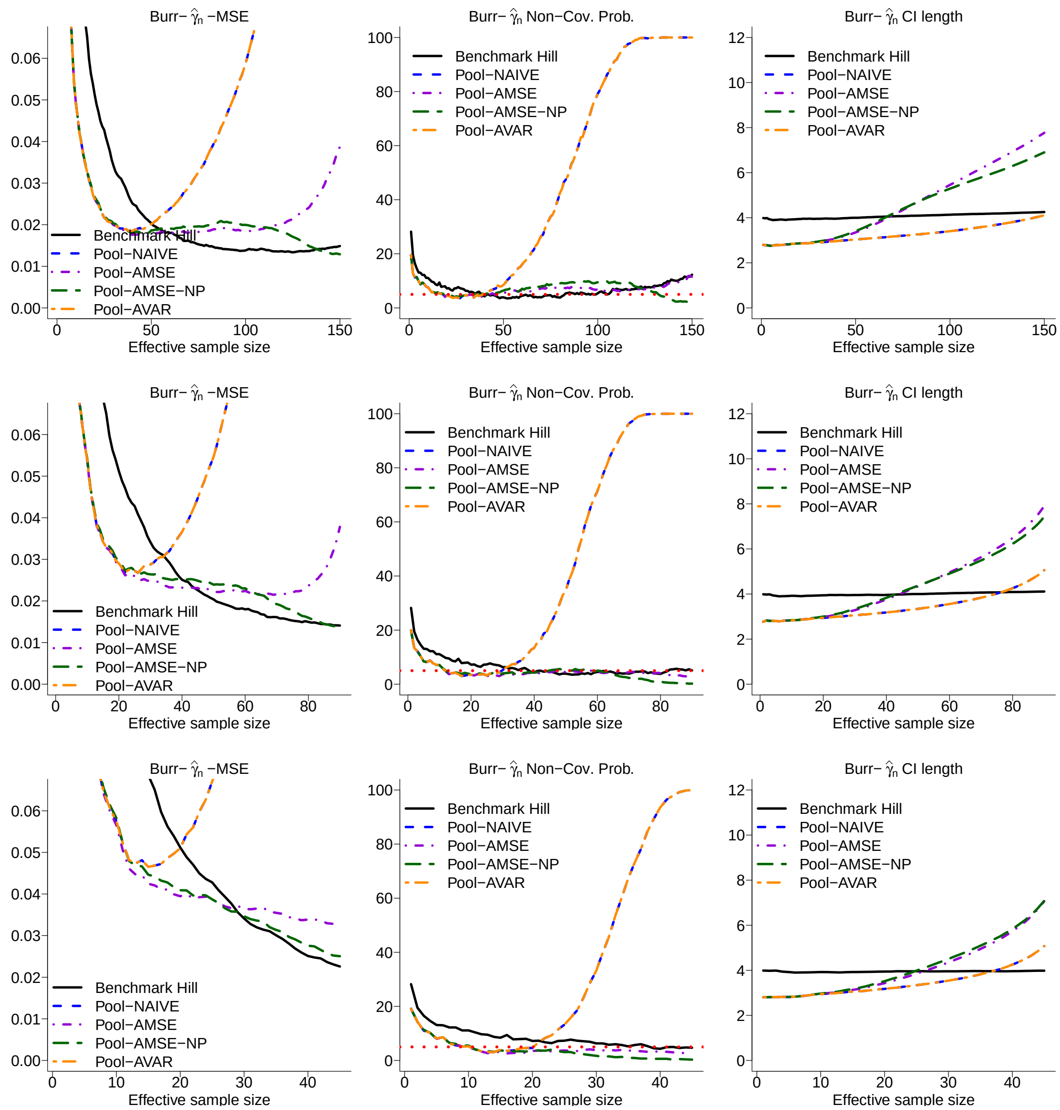

Figure D.15: Simulation results, distributed inference context, Burr model (h). Left panels: MSE of the tail index point estimators, middle panels: non-coverage probability of the asymptotic confidence intervals, right panels: average length of the 95% confidence intervals multiplied by $\sqrt{k}$. In 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 $k_1 = k_2$.

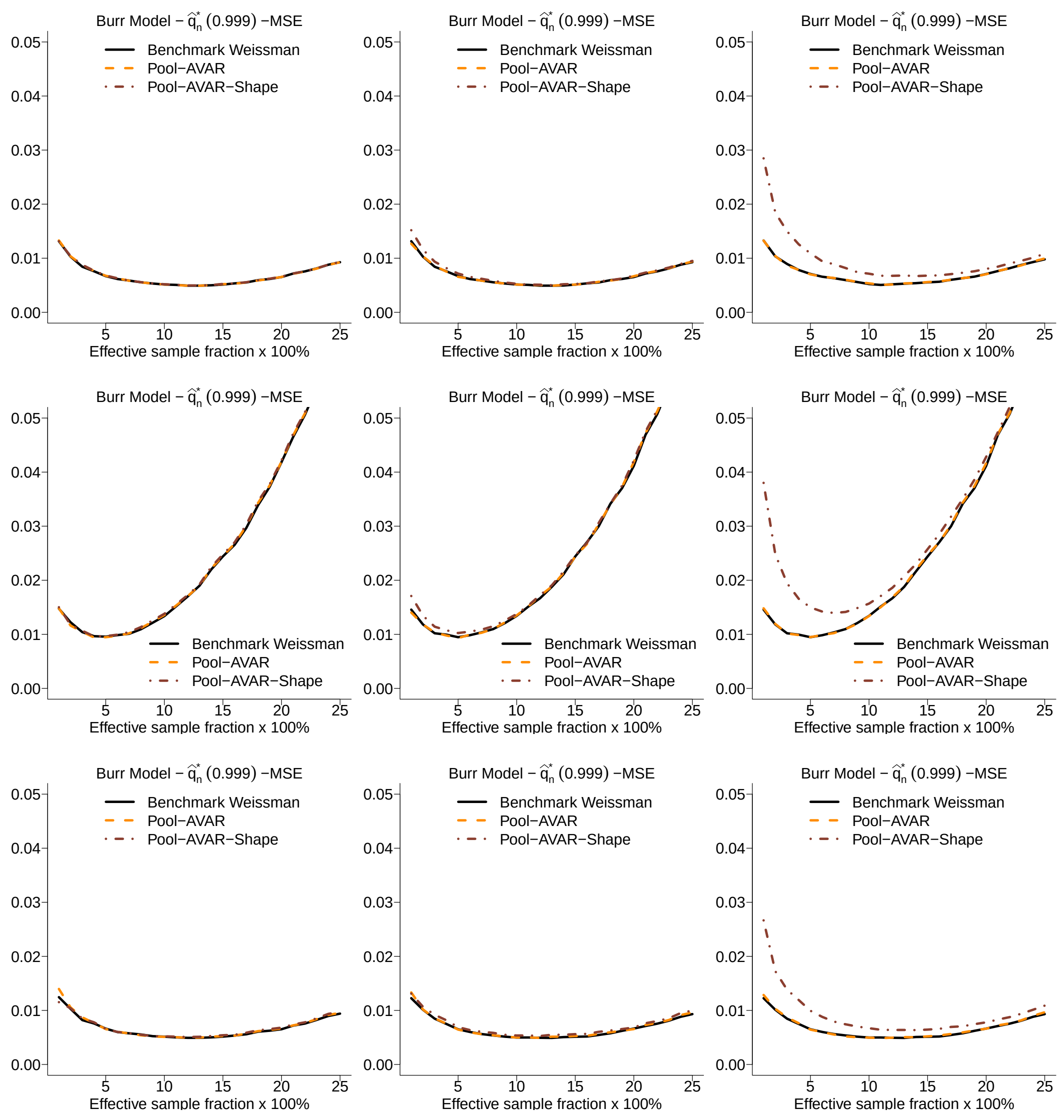

Figure D.16: Simulation results, distributed inference context, extreme quantile estimation at level $1-p=0.999$ in the Burr models (Q-a), (Q-b), (Q-c). Top row, model (Q-a), from left to right: $n_1=900$, $n_1=500$, $n_1=100$; middle row, model (Q-b), from left to right: $n_1=900$, $n_1=500$, $n_1=100$; 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.

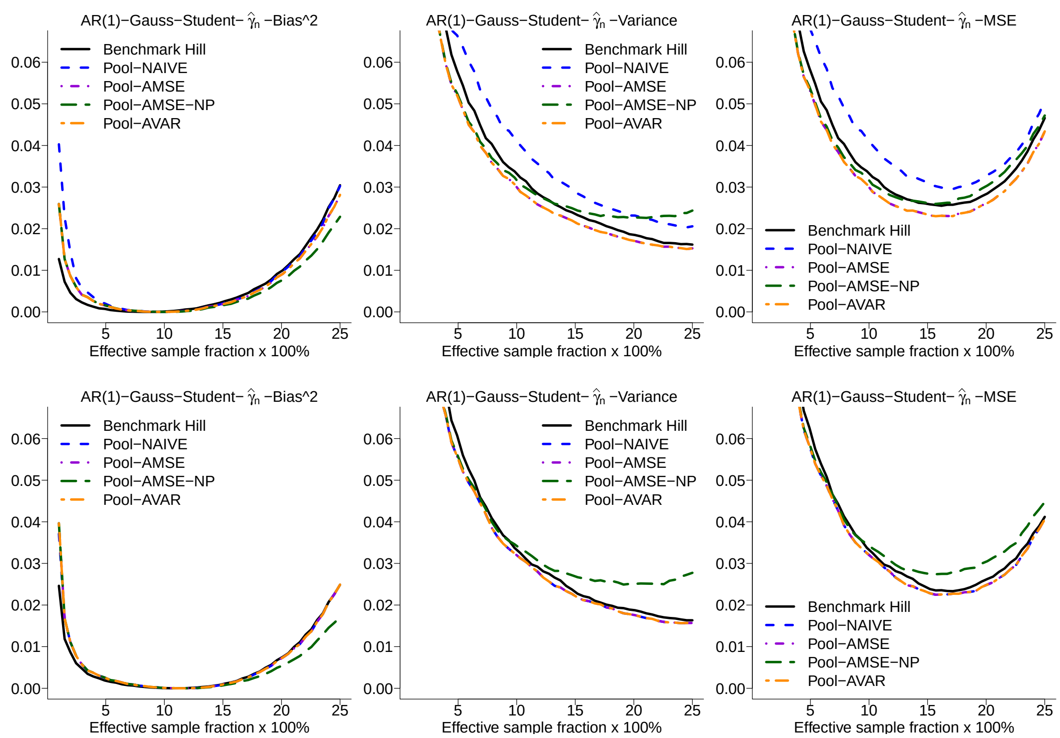

Figure D.17: Simulation results, pooled estimators using residuals, Gaussian-Student innovations model (AR-1-a). All panels relate to tail index estimation. Left panels: squared bias, middle panels: variance, right panels: MSE. All results are represented as functions of the sample fraction k_j/n_j (identical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

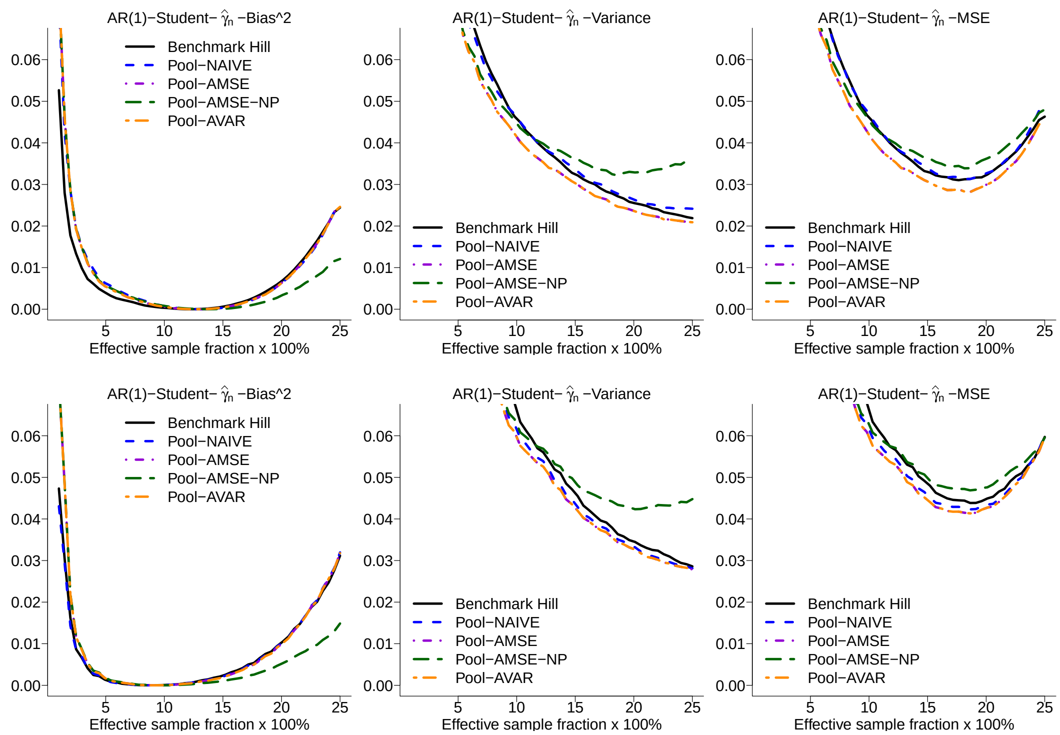

Figure D.18: Simulation results, pooled estimators using residuals, Multivariate-Student innovations model (AR-1-b). All panels relate to tail index estimation. Left panels: squared bias, middle panels: variance, right panels: MSE. All results are represented as functions of the sample fraction k_j/n_j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

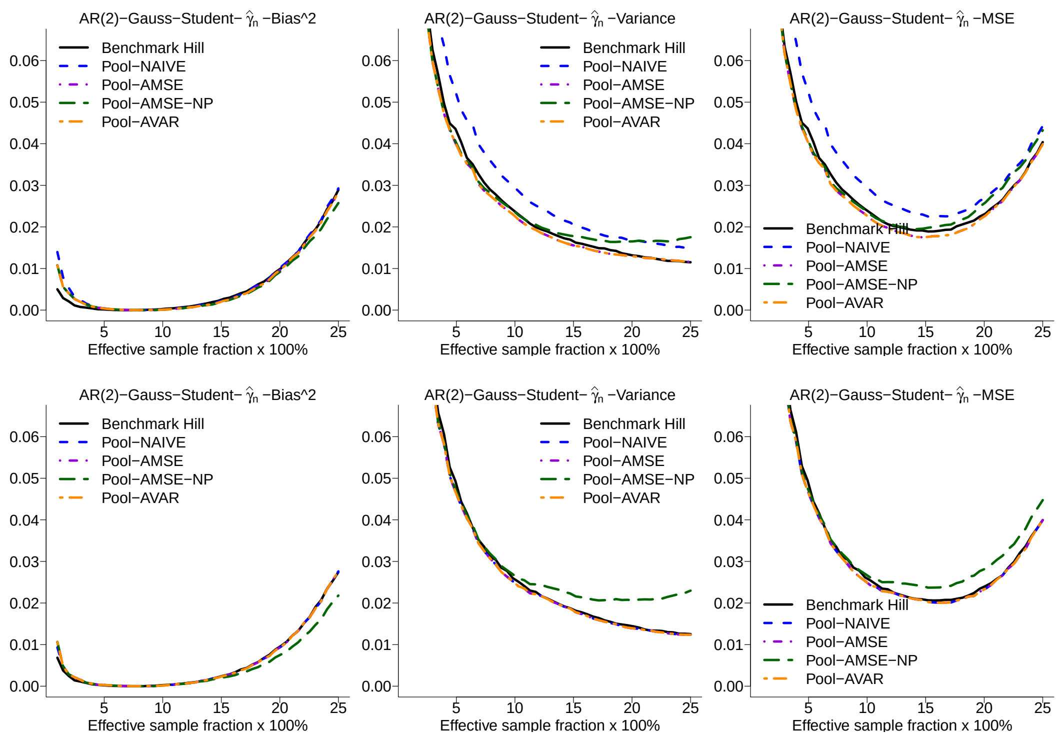

Figure D.19: As in Figure D.17, in the Gaussian-Student innovations model (AR-2-a) with misspecified AR(1) dynamics.

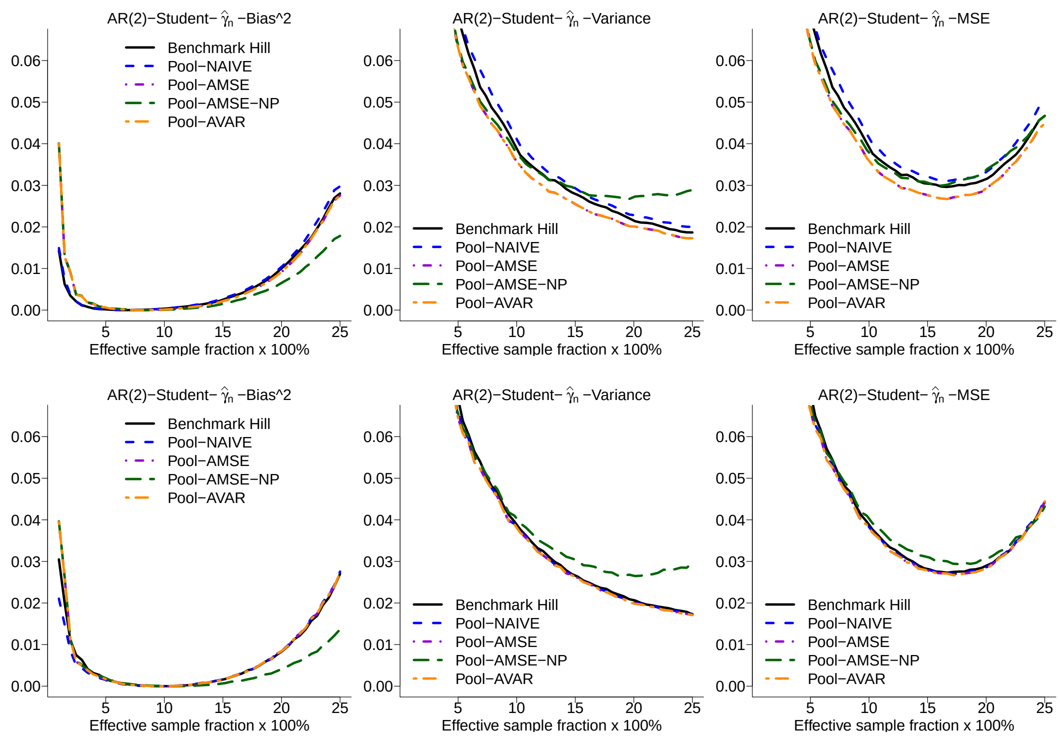

Figure D.20: As in Figure D.18, in the Multivariate-Student innovations model (AR-2-b) with misspecified AR(1) dynamics.

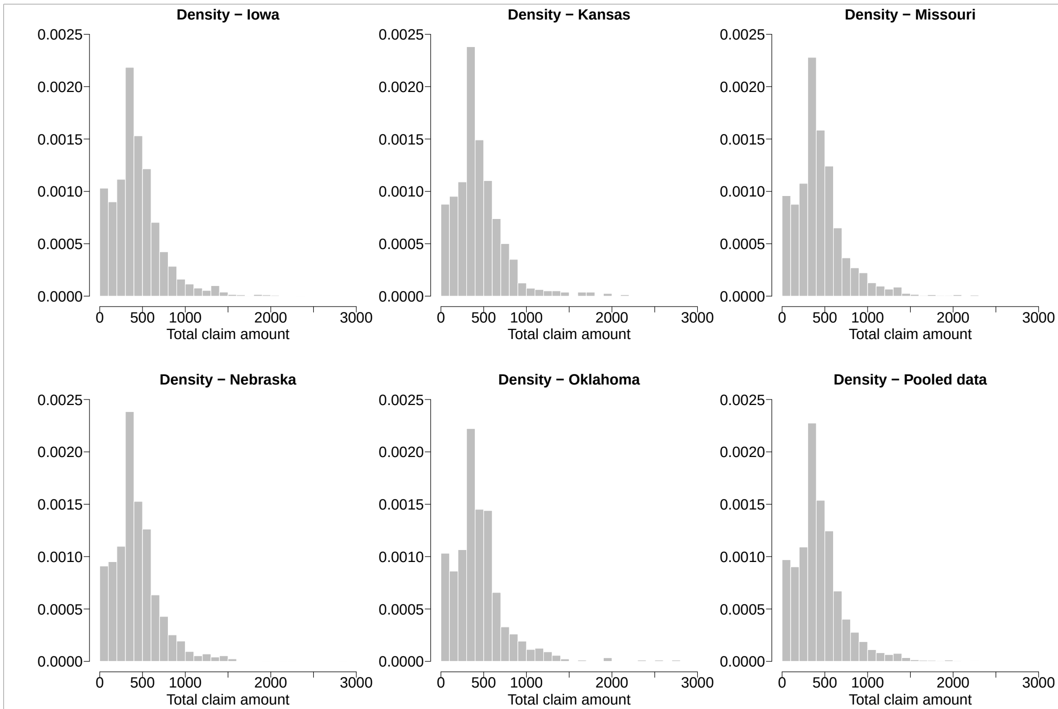

Figure D.21: Car insurance data: Histograms of total claim amounts.

D.2. Data analysis

We give here extra results related to our two real data analyses.

D.2.1. Distributed inference for car insurance data

Figure D.21 shows the frequency of total claim amounts for each state (the five first panels, from left to right and top to bottom) and for the data set of pooled claims (bottom right panel). Figure D.22 represents the individual Hill estimates of the tail index and Weissman estimates of the extreme quantile at level 0.9999 in each state.

D.2.2. Pooling for regional inference on extreme rainfall

Figure D.23 gives histograms of the data we use in the 8 stations. Figure D.24 provides plots of the individual Hill estimates at each station, along with 95% confidence intervals. Table D.1 gives a summary of the exploratory extreme value analysis at each station.

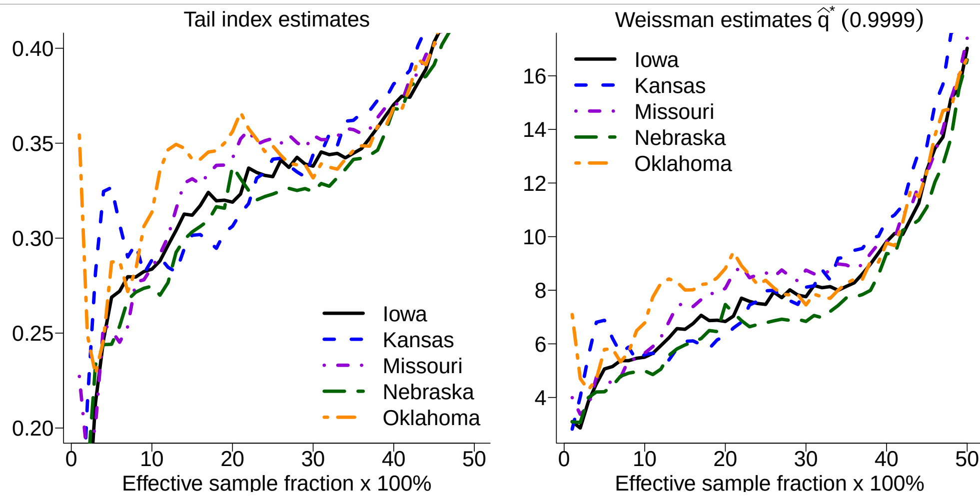

Figure D.22: Car insurance data: Hill estimates $\widehat{\gamma}_j(k_j)$ (left) and Weissman estimates $\widehat{q}_j^{\star}(0.9999|k_j)$ (right) for each state, as functions of the sample fraction k_j/n_j, assumed to be identical in each sub-sample.

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.

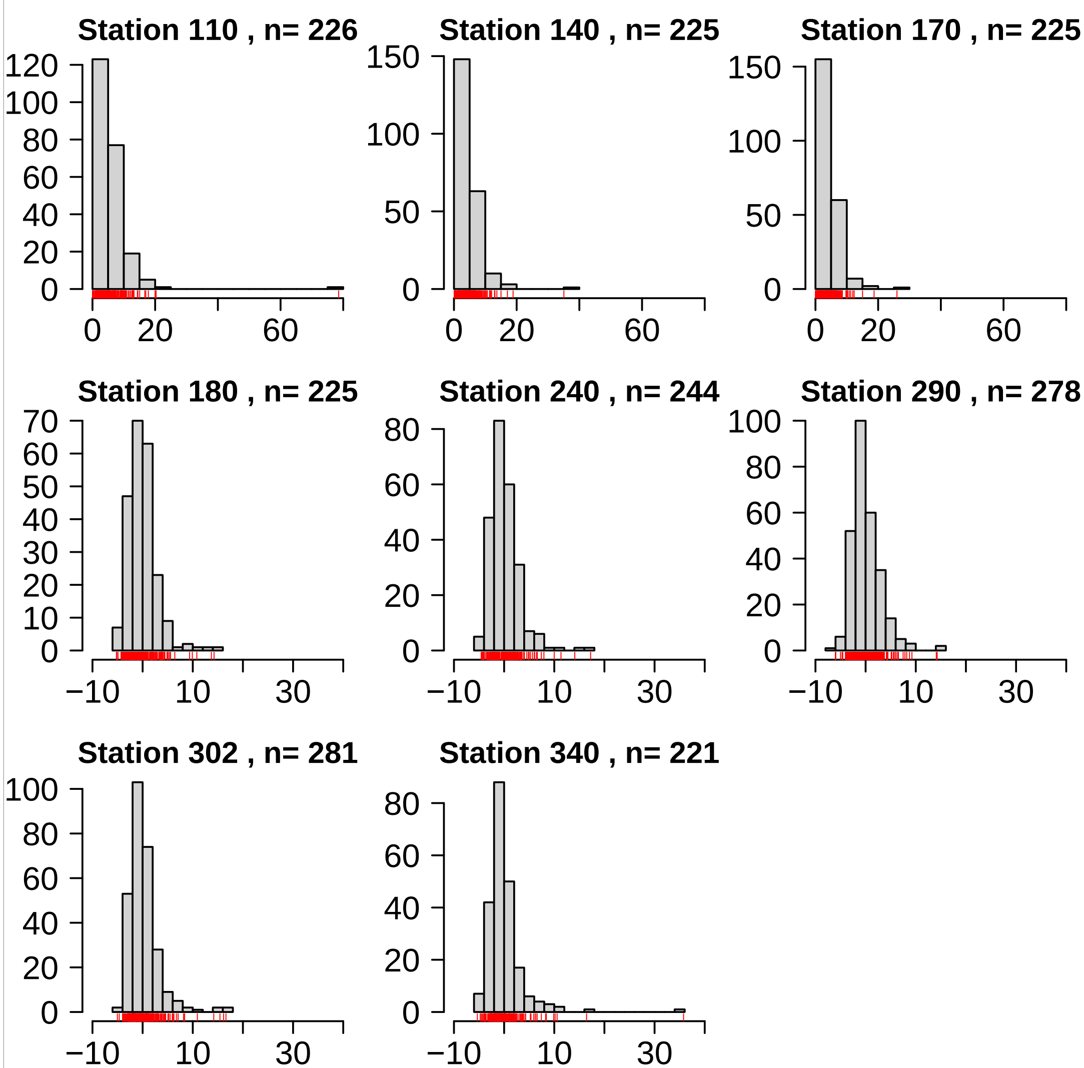

Figure D.23: Florida rainfall data: Histograms of the raw data for the three stations in the top panel (red cluster), and of the residuals obtained from the fitted SARMA models of the five remaining stations (green cluster).

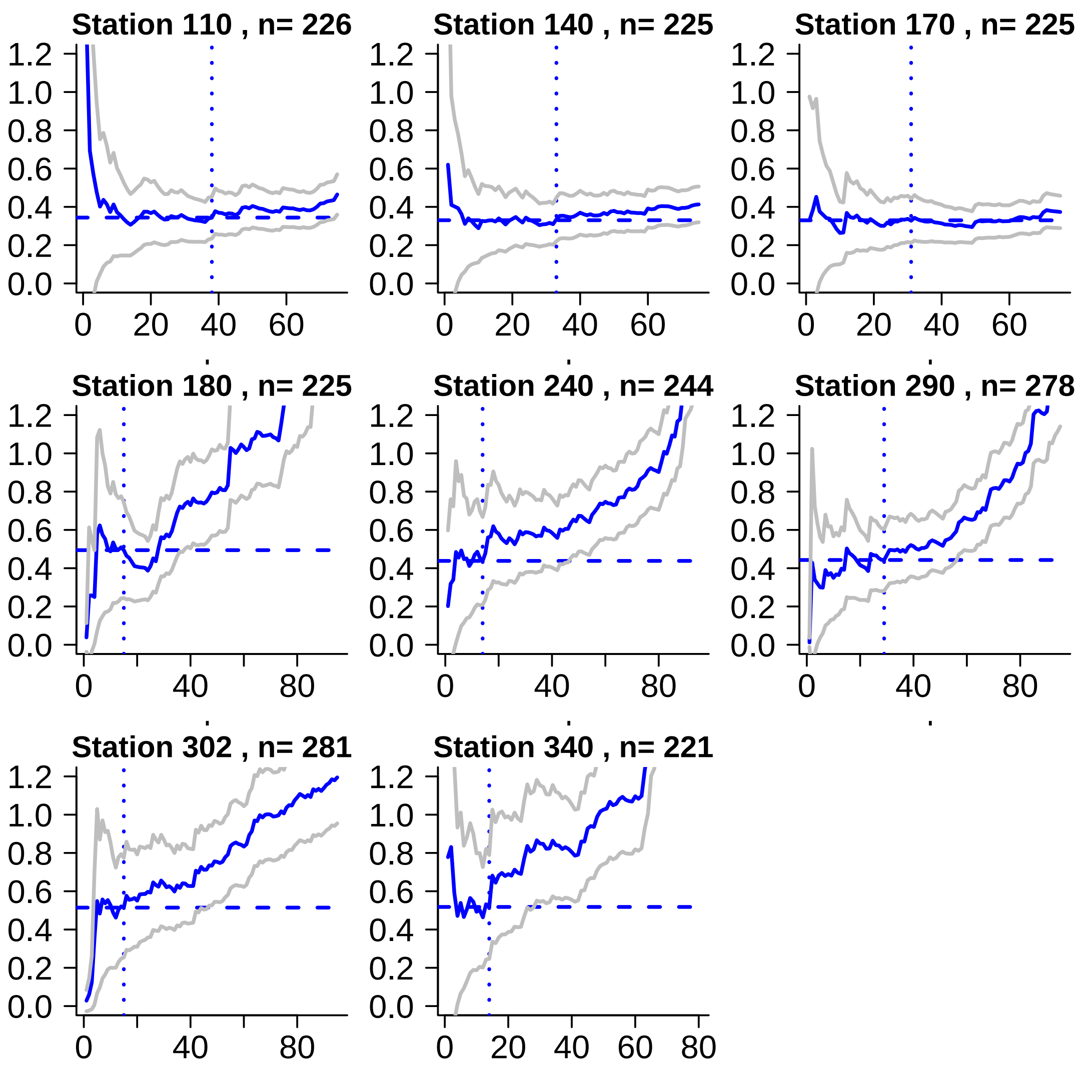

Figure D.24: Florida rainfall data: Individual tail index estimators. For each station j, we represent the Hill estimate $\widehat{\gamma}_j(k_j)$ (solid blue) and its associated asymptotic 95% confidence interval (solid gray), as functions of the effective sample size $k = k_j$. The corresponding Hill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

References

[1] Billingsley, P. (1995). *Probability and Measure (Third Edition)*. John Wiley and Sons, New York.

[2] Caeiro, F., Gomes, M. I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3** 113–136.

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

[4] Daouia, A., Girard, S. and Stupfler, G. (2020). Tail expectile process and risk assessment. *Bernoulli* **26** 531–556.

[5] David, P. and Gu, W. (2019). A Riemannian structure for correlation matrices. *Operators and Matrices* **13** 607–627.

[6] David, H. A. and Nagaraja, H. N. (2003). *Order Statistics (Third edition)*. Wiley Series in Probability and Statistics.

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

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

[9] Embrechts, P., Hofert, M. and Wang, R. (2016). Bernoulli and tail-dependence compatibility. *Annals of Applied Probability* **26** 1636–1658.

[10] Fiebig, U. R., Strokorb, K. and Schlather, M. (2017). The realization problem for tail correlation functions. *Extremes* **20** 121–168.

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

[12] Girard, S., Stupfler, G. and Usseglio-Carleve, A. (2022). On automatic bias reduction for extreme expectile estimation. *Statistics & Computing* **32** 64.

[13] Gomes, M. I., Brilhante, M. F. and Pestana, D. (2016). New reduced-bias estimators of a positive extreme value index. *Communications in Statistics-Simulation and Computation* **45** 833–862.

[14] Greub, W. and Rheinboldt, W. (1959). On a generalization of an inequality of L. V. Kantorovich. *Proceedings of the American Mathematical Society* **10** 407–415.

[15] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

[16] Journée, M., Bach, F., Absil, P. A. and Sepulchre, R. (2010). Low-Rank Optimization on the Cone of Positive Semidefinite Matrices. *SIAM Journal on Optimization* **20** 2327–2351.

[17] Kinsvater, P., Fried, R. and Lilienthal, J. (2016). Regional extreme value index estimation and a test of tail homogeneity. *Environmetrics* **27** 103–115.

[18] Laurent, M. and Poljak, S. (1995). On a positive semidefinite relaxation of the cut polytope. *Linear Algebra and its Applications* **223–224** 439–461.

[19] Lax, P. D. (2007). *Linear Algebra and its Applications*. John Wiley & Sons.

[20] Nikoloulopoulos, A. K., Joe, H. and Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes* **12** 129–148.

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

[22] Rice, J. R. (1966). A Theory of Condition. *SIAM Journal on Numerical Analysis* **3** 287–310.

[23] Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics* **33** 307–335.

[24] Strokorb, K. (2013). Characterization and construction of max-stable processes, PhD thesis, Georg-August-Universität Göttingen.

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

[26] Tricomi, F. G. and Erdélyi, A. (1951). The asymptotic expansion of a ratio of gamma functions. *Pacific Journal of Mathematics* **1** 133–142.