Wang, Chenhui; Cai, Juan Juan; Lin, Yicong; Schaumburg, Julia

Working Paper

Clustering extreme value indices in large panels

Tinbergen Institute Discussion Paper, No. TI 2025-029/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Wang, Chenhui; Cai, Juan Juan; Lin, Yicong; Schaumburg, Julia (2025) : Clustering extreme value indices in large panels, Tinbergen Institute Discussion Paper, No. TI 2025-029/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/316218

TI 2025-029/III
Tinbergen Institute Discussion Paper

Clustering Extreme Value Indices in Large Panels

Chenhui Wang[1]
Juan Juan Cai[2]
Yicong Lin[3]
Julia Schaumburg[4]

1 Vrije Universiteit Amsterdam

2 Vrije Universiteit Amsterdam, Tinbergen Institute

3 Vrije Universiteit Amsterdam, Tinbergen Institute

4 Vrije Universiteit Amsterdam, Tinbergen Institute

Clustering extreme value indices in large panels

Chenhui Wang[1], Juan Juan Cai*[1,2], Yicong Lin[1,2] and Julia Schaumburg[1,2]

[1]Vrije Universiteit Amsterdam, [2]Tinbergen Institute

April, 2025

Abstract

We analyze a large panel of units grouped by shared extreme value indices (EVIs) and aim to identify these unknown groups. To achieve this, we order the Hill estimates of individual EVIs and segment them by minimizing the total squared distance between each estimate and its corresponding group average. We show that our method consistently recovers group memberships, and we establish the asymptotic normality of the proposed group estimator. The group estimator attains a faster convergence rate than the individual Hill estimator, leading to improved estimation accuracy. Simulation results reveal that our method achieves high empirical segmentation accuracy, and the resulting group EVI estimates substantially reduce mean absolute errors compared to individual estimates. We apply the proposed method to analyze a rainfall dataset collected from 4,735 stations across Europe, covering the winter seasons from January 1, 1950, to December 31, 2020, and find statistically significant evidence of an increase in the highest and a decrease in the lowest group EVI estimates, suggesting growing variability and intensification of extreme rainfall events across Europe.

***Keywords*:** extreme value index, clustering, group Hill estimator, large panels, rainfall

*Corresponding author: Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands. E-mail address: j.cai@vu.nl.
This work was supported by the Dutch National Science Foundation (NWO) [VI.VIDI.191.169 to J.S.]

1 Introduction

Estimation of the extreme value index (EVI) is fundamental to tail inference and has been a central focus in extreme value analysis. Traditionally, analyses focus on random variables with identical distribution and therefore a common EVI; see, e.g., Beirlant et al. (2004) and de Haan and Ferreira (2006) for textbook treatments. However, the assumption of tail index homogeneity across all individuals is often too restrictive. Recent advances have been made to accommodate heterogeneous EVIs. For instance, de Haan and Zhou (2021) allow for smooth variations in the indices of neighboring independent variables and introduce a local nonparametric estimator. Einmahl and He (2022, 2023) further permit fully heterogeneous indices without imposing a smoothness condition and study the EVI of the average distribution for independent but non-identically distributed variables.

In this paper, we consider a large panel setting in which both the number of cross-sectional units and time series observations diverge, under an unobserved group structure where members within each group share a common EVI. A simple method for identifying the groups based on their EVIs is proposed. Furthermore, we provide a group EVI estimator and show that it is asymptotically more efficient than individual-specific estimators. Like de Haan and Zhou (2021) and Einmahl and He (2022, 2023), we focus on heavy-tailed data, i.e., positive EVIs. Yet, we neither assume a smooth variation across EVIs nor account for full EVI heterogeneity.

Our proposed clustering framework draws inspiration from the extensive literature on estimating structural breaks in time series models (see, among many others, Bai, 1997a,b; Bai and Perron, 1998; Perron, 2006; Qu and Perron, 2007; Aue and Horváth, 2013; Li et al., 2024), which, nevertheless, has received little attention in extreme value theory. The idea is as follows. Assuming each unit within a group shares a common EVI, the corresponding individual estimates should center around this shared value. If the group indices are well separated, ordering all individual estimates from smallest to largest should reveal a gradual increase in levels, with

larger shifts marking the boundaries between different groups. The EVI estimates between consecutive level changes then form a group. Thus, our proposed procedure consists of three key steps. First, we estimate the EVIs for cross-sectional units using the Hill estimator and arrange them in order. Second, we segment the ordered estimates by their magnitude, determining the segmentation locations by minimizing the total squared deviation of individual estimates from their respective group averages. Finally, we compute the group EVI estimators using within-group averages in a similar manner to the distributed Hill estimator by Chen et al. (2022), which is derived from a divide-and-conquer algorithm.

Although our approach is reminiscent of regression models with structural breaks in the time series literature, there are some key differences. Most importantly, whereas the data modeled in structural break studies are directly observable, EVIs are not and have to be estimated first. As a result, the estimation errors of the EVIs, which are degenerate, must be carefully addressed in theory. This invalidates several key steps in existing proofs and calls for a different asymptotic treatment. We address these issues using fundamental results from extreme value theory and establish that our segmentation procedure is able to consistently identify the unknown groups under mild conditions. Specifically, we prove that the estimated shift locations converge in probability to the true locations. This is a stronger result than what is typically found in the structural break literature, where the break fraction estimator, defined as the estimated change point divided by the total number of observations, is consistent, but the estimated break locations themselves are generally not (Bai and Perron, 1998).

Building on the consistency results, we further establish the asymptotic normality of our estimator for the group-specific EVIs. As expected, the group estimators exhibit faster convergence than their individual-specific counterparts since they utilize information from both the cross-sectional and time dimensions. Notably, our theory allows for the independent selection of threshold sequences for segmentation and group estimator construction. Such flexibility proves

highly beneficial in our simulation study, where we observe that the mean absolute errors of the group estimators can be reduced by up to approximately 90% compared to the individual Hill estimators. Finally, we illustrate our proposed framework using rainfall data collected from 4,735 stations across Europe. We find statistically significant evidence of increased variability and intensification of extreme rainfall events across the region over time.

In addition to the structural break literature, our work also relates to several articles in extreme value theory that warrant discussion. de Carvalho et al. (2023) propose a similarity-based clustering algorithm to group time series based on the EVIs and scedasis functions; however, their approach lacks theoretical foundations. More closely related to our work, Dupuis et al. (2023) consider a panel generalized extreme value (GEV) regression model with a fixed number of individuals (i.e., $N < \infty$), where GEV parameters are modeled with covariates through link functions. They define groups based on the homogeneity of slope coefficients for explanatory variables. Two key aspects distinguish our method. First, since their approach relies on covariate information, the grouping results depend on the selection of covariates. Therefore, different model specifications can lead to substantially different group memberships, potentially undermining their practical interpretability. In contrast, our framework allows the EVIs to manifest their group identities. Second, we work in a large N setting ($N \to \infty$), which requires some uniform results that are not necessary in a finite N setting. Finally, Hou et al. (2024) define groups based on multiple tests for equality of EVIs between individuals. This approach could result in an individual belonging to two different groups, making it unclear what a group truly represents. Our approach, on the other hand, ensures unique memberships.

The remainder of the paper is organized as follows. Section 2 introduces the setup and the proposed clustering framework and establishes the asymptotic theory. In Section 3, we present a simulation study, where we also discuss a fast segmentation method. Section 4 illustrates our methods using a rainfall dataset. Section 5 concludes. The main proofs are presented in

Appendix A. Additional proofs, along with simulation and empirical results, are provided in the Online Appendix.

2 The clustering framework and asymptotic theory

Consider a row-wise i.i.d. triangular array $\{Z_{i,t},\, t = 1, \ldots, T_i,\, i = 1, \ldots, N\}$, where N represents the total number of cross-sectional units and T_i is the total number of time series observations for the ith unit. We adopt a framework where $N \to \infty$ and $T_i \to \infty$ for all $i = 1, \ldots, N$. As mentioned, we focus on heavy-tailed data.

Assumption 1 (Independent and heavy-tailed data)**.** *For $i = 1, \ldots, N$, $\{Z_{i,t},\, t = 1, \ldots, T_i\}$ is independently and identically distributed (i.i.d.) with a continuous marginal distribution F_i, and there exists a $\gamma_i > 0$ such that*

$$\lim_{s\to\infty} \frac{1 - F_i(sx)}{1 - F_i(s)} = x^{-1/\gamma_i}, \qquad x > 0. \tag{2.1}$$

Note that the condition in (2.1), known as the first-order regular variation condition, can also be equivalently expressed in terms of tail quantile functions U_i as follows:

$$\lim_{s\to\infty} \frac{U_i(sx)}{U_i(s)} = x^{\gamma_i}, \qquad x > 0, \tag{2.2}$$

where $U_i(x) = F_i^{-1}(1 - x^{-1})$ for $x > 1$, and F_i^{-1} denotes the left-continuous inverse of F_i. The set of extreme value indices (EVIs) $\{\gamma_i,\, i = 1, \ldots, N\}$ characterizes the extreme behavior of the cross-sectional units. Even small changes in γ_i can have a substantial impact on tail probabilities due to its exponential influence. We assume that $\{\gamma_i,\, i = 1, \ldots, N\}$ can be grouped into $G \in \mathbb{Z}^+$ distinct groups. That is, for each i, there exists some γ_{g_j} such that $\gamma_i = \gamma_{g_j}$ for some $j = 1, \ldots, G$. Without loss of generality, we assume that $\gamma_{g_1} < \gamma_{g_2} < \cdots < \gamma_{g_G}$.

Apart from the group structure in the EVIs, we do not impose any additional constraints on the distributions of the panel data. Within groups, the convergence speed of (2.2) may vary (see the second-order condition in Assumption 4). No smoothness conditions are required for the distributions of neighboring panel units, and the data can be on different scales.

2.1 Clustering EVIs

Our clustering framework consists of three key steps: ordering, segmenting, and estimating group EVIs. Throughout, we denote the integer part of $a \in \mathbb{R}^+$ by $[a]$. Let $\mathbb{1}\{\cdot\}$ be an indicator function. For a vector $\boldsymbol{x} \in \mathbb{R}^n$, its p-norm is denoted by $\|\boldsymbol{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$.

Step 1 (Ordering). Since the true EVIs are not observable, we begin by considering the Hill (1975) estimator of the individual extreme value index. The Hill estimator is given by

$$\hat{\gamma}_i^H(k_i) := \frac{1}{k_i} \sum_{j=0}^{k_i-1} \log Z_{i,(T_i-j)} - \log Z_{i,(T_i-k_i)}, \qquad i = 1, \ldots, N, \tag{2.3}$$

where $Z_{i,(1,T_i)} \leq Z_{i,(2,T_i)} \leq \ldots \leq Z_{i,(T_i,T_i)}$ denote the order statistics of $\{Z_{i,t},\, t = 1, \ldots, T_i\}$, and $k_i = k_i(T_i) < T_i$ is an intermediate sequence of integers. When the context is clear, we simplify the notation by writing $\hat{\gamma}_i^H$ instead of $\hat{\gamma}_i^H(k_i)$.

In what follows, let the Hill estimates be pre-ordered, i.e., $\hat{\gamma}_1^H \leq \hat{\gamma}_2^H \leq \ldots \leq \hat{\gamma}_N^H$. We assume that the Hill estimator recovers the underlying ordering $\gamma_1 \leq \gamma_2 \leq \ldots \leq \gamma_N$. This assumption is not restrictive if T_i is reasonably large. More discussion is provided in Section 2.2. In the ordering step, any alternatives to the Hill estimator can be used. For example, bias-corrected estimators (Ivette Gomes et al., 2008), or more computationally intensive threshold selection methods (Danielsson et al., 2001; Bader et al., 2018; Drees et al., 2020) may be applied. The key objective is to recover the underlying ordering accurately. In the subsequent steps of segmentation and group EVI estimation, we rely solely on the Hill estimator in (2.3) for its

computational efficiency and well-established asymptotic properties.

Step 2 (Segmenting). Recall that there are G distinct values of γ_i's, and let $m = G - 1$ be the number of segmentation points. Now, we discuss how to identify group memberships provided G is known. A data-driven method for determining G will be explored in Section 3.3. We denote the true group segmentation with $\boldsymbol{\ell}_0 = (l_{10}, \ldots, l_{m0})^\top$ such that

$$\begin{aligned}\gamma_1 = \gamma_2 = \ldots = \gamma_{l_{10}} < \gamma_{l_{10}+1} = \gamma_{l_{10}+2} = \ldots = \gamma_{l_{20}} \\ < \gamma_{l_{20}+1} = \ldots < \ldots = \ldots \gamma_{l_{m0}} < \gamma_{l_{m0}+1} = \ldots = \gamma_N.\end{aligned} \tag{2.4}$$

Clearly, $\gamma_{l_{j0}} = \gamma_{g_j}$ for $j = 1, \ldots, G$, where $l_{G0} = N$. In practice, $\boldsymbol{\ell}_0$ is unknown and has to be estimated. We use $\left\{\hat{\gamma}_i^H,\, 1 \le i \le N\right\}$ to estimate $\boldsymbol{\ell}_0$. This task can be viewed as identifying break locations in change-in-mean models. To illustrate this, for $i = 1, \ldots, N$, it is clear that $\hat{\gamma}_i^H$ can be equivalently represented as

$$\hat{\gamma}_i^H = \gamma_{g_1} + \Delta_1 \mathbb{1}\{i > l_{10}\} + \Delta_2 \mathbb{1}\{i > l_{20}\} + \ldots + \Delta_m \mathbb{1}\{i > l_{m0}\} + \hat{\varepsilon}_i, \tag{2.5}$$

where $\Delta_j = \gamma_{g_{j+1}} - \gamma_{g_j}$, $j = 1, \ldots, m$. Note that $\hat{\varepsilon}_i = \hat{\gamma}_i^H - \gamma_i$ represents estimation errors from the individual Hill estimator and depends on the data.

We are now ready to estimate $\boldsymbol{\ell}_0$, along with $(\gamma_{g_1}, \ldots, \gamma_{g_G})^\top$. Specifically, we rewrite Model (2.5) in compact form, replacing the true segmentation $\boldsymbol{\ell}_0$ with a candidate break location vector $\boldsymbol{\ell} = (l_1, \ldots, l_m)^\top \in \boldsymbol{\mathcal{L}} \subset \{1, 2, \ldots, T\}^m$. Define $\hat{\boldsymbol{\Gamma}} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_N)^\top$, $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_N)^\top$, and let $\boldsymbol{X}_{\boldsymbol{\ell}} = \left(x_{ij},\, i = 1, \ldots, N,\, j = 1, \ldots, G\right)$ be an $N \times G$ matrix with the entries $x_{ij} = \mathbb{1}\{i > l_{j-1}\}$, where $l_0 = 0$. We arrive at

$$\hat{\boldsymbol{\Gamma}} = \boldsymbol{X}_{\boldsymbol{\ell}} \boldsymbol{\beta}_{\boldsymbol{\ell}} + \hat{\boldsymbol{\varepsilon}}, \tag{2.6}$$

where $\boldsymbol{\beta}_{\boldsymbol{\ell}}$ is a G-dimensional coefficient vector. Note that the true $\boldsymbol{\beta}_{\boldsymbol{\ell}_0}$ is given by $\boldsymbol{\beta}_{\boldsymbol{\ell}_0} = (\gamma_{g_1}, \Delta_1, \cdots, \Delta_m)^\top$. We estimate the parameters by minimizing the sum of squared residuals (SSR) as follows:

$$\hat{\boldsymbol{\ell}} = \underset{\boldsymbol{\ell} \in \boldsymbol{\mathcal{L}}}{\operatorname{argmin}} \operatorname{SSR}(\boldsymbol{\ell}) = \underset{\boldsymbol{\ell} \in \boldsymbol{\mathcal{L}}}{\operatorname{argmin}} \left\| \hat{\boldsymbol{\Gamma}} - \boldsymbol{X}_{\boldsymbol{\ell}} \hat{\boldsymbol{\beta}}_{\boldsymbol{\ell}} \right\|_2^2, \qquad \hat{\boldsymbol{\beta}}_{\boldsymbol{\ell}} = \left(\boldsymbol{X}_{\boldsymbol{\ell}}^\top \boldsymbol{X}_{\boldsymbol{\ell}} \right)^{-1} \boldsymbol{X}_{\boldsymbol{\ell}}^\top \hat{\boldsymbol{\Gamma}}. \tag{2.7}$$

Minimizing $\operatorname{SSR}(\cdot)$ jointly for all elements in $\boldsymbol{\ell}$ requires searching through all possible combinations of break locations, making the construction of $\hat{\boldsymbol{\ell}}$ computationally intensive when G is large, with an operational complexity of order $O(N^m)$. To address this, we employ an efficient algorithm proposed by Bai and Perron (2003) based on dynamic programming, which reduces the complexity to order $O(N^2)$.

Given $\hat{\boldsymbol{\ell}}$, we then obtain the coefficient estimator:

$$\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\ell}}} = \left(\hat{\beta}_{\hat{\boldsymbol{\ell}},1}, \hat{\beta}_{\hat{\boldsymbol{\ell}},2}, , \ldots, \hat{\beta}_{\hat{\boldsymbol{\ell}},G} \right)^\top = \left(\boldsymbol{X}_{\hat{\boldsymbol{\ell}}}^\top \boldsymbol{X}_{\hat{\boldsymbol{\ell}}} \right)^{-1} \boldsymbol{X}_{\hat{\boldsymbol{\ell}}}^\top \hat{\boldsymbol{\Gamma}}.$$

It is worth highlighting that both $\hat{\boldsymbol{\ell}} = \hat{\boldsymbol{\ell}}(\boldsymbol{k})$ and $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\ell}}} = \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\ell}}}(\boldsymbol{k})$, where $\boldsymbol{k} = (k_1, \ldots, k_N)$, depend on the intermediate threshold sequence $\{k_i,\, i = 1, \ldots, N\}$ through $\hat{\boldsymbol{\Gamma}}$.

Step 3 (Estimating group EVIs). Given the estimated break locations $\hat{\boldsymbol{\ell}}$, we proceed by estimating the extreme value index γ_{g_j} for each group. At this stage, intermediate thresholds may be selected that are different from those used in the segmentation step. Let $\tilde{\boldsymbol{k}} = (\tilde{k}_1, \ldots, \tilde{k}_N)$, where each $\tilde{k}_i$ is not necessarily equal to k_i used in (2.3). We propose the following group Hill estimator based on a simple average:

$$\hat{\gamma}_{g_j}\left(\hat{\boldsymbol{\ell}}, \tilde{\boldsymbol{k}}\right) = \frac{1}{\hat{l}_j - \hat{l}_{j-1}} \sum_{i=\hat{l}_{j-1}+1}^{\hat{l}_j} \hat{\gamma}_i^H(\tilde{k}_i), \qquad j = 1, \ldots, G, \tag{2.8}$$

where $\hat{\gamma}_i^H(\tilde{k}_i)$ is defined in (2.3). In practice, variants of averages may also be considered. Weighted averages, which assign greater weight to individuals in the middle of a group, may be preferable, especially when T_i is small. In such cases, individual Hill estimates are less accurate, which can affect the accuracy of pre-ordering.

Although our group Hill estimator is conceptually similar to the distributed Hill estimator introduced in Chen et al. (2022), their estimator is derived from a divide-and-conquer algorithm in which the groups are predetermined. In contrast, our framework aims to recover the *unknown* group identities. Furthermore, we accommodate heterogeneous distributions by allowing different intermediate thresholds for each individual, making our estimator more general than the one considered in their work.

2.2 Asymptotic theory

In this section, we establish the asymptotic properties of $\hat{\boldsymbol{\ell}}$ and the asymptotic normality of the group estimator $\hat{\gamma}_{g_j}\big(\hat{\boldsymbol{\ell}}, \tilde{\boldsymbol{k}}\big)$. The proofs can be found in Appendix A. We begin by establishing a crucial uniform bound for the partial sum processes of $\{\hat{\epsilon}_i\}$, which will be referenced throughout the proofs concerning the asymptotic properties of $\hat{\boldsymbol{\ell}}$.

As noted earlier, our approach offers a flexible strategy that allows for the independent selection of intermediate thresholds at each stage. The choice of thresholds is a critical and challenging aspect in the extreme value theory literature, as it involves a trade-off between bias and variance in the estimator. To ensure the validity of our approach, we first impose some mild assumptions on the intermediate sequences $\{k_i,\, i = 1, \ldots, N\}$ used in the segmentation step.

Assumption 2 (Conditions on k_i)**.** *For each* $i = 1, \ldots, N$, $k_i/T_i \to 0$ *as* $T_i \to \infty$ *and* $N \max_{1 \le i \le N} \exp(-k_i) \to 0$ *as* $N \to \infty$.

Assumption 2 is easy to satisfy and allows for a wide range of choices for $\boldsymbol{k}$. The condition $k_i/T_i \to 0$ is a standard requirement in extreme value statistics, ensuring that the threshold

$Z_{i,(T_i-k_i)}$ for each panel lies in the tail of the underlying distribution. Define $k_{\min} = \min_{i=1,\ldots,N} k_i$. The condition on N and $k_{\min}$ accounts for the heterogeneous tail convergence rates across different units. It is satisfied, for example, by $k_{\min} = O(N^{\alpha})$ or $k_{\min} = O\big((\log N)^{1+\alpha}\big)$, for any $\alpha > 0$.

The following key lemma establishes the uniform asymptotic property on the partial sum of $\hat{\varepsilon}_i = \hat{\gamma}_i^H(k_i) - \gamma_i$, highlighting a fundamental difference from conventional structural break literature.

Lemma 1. *Under Assumptions 1 and 2, we have as $N \to \infty$, for any $1 \le l_1 < l_2 \le N$,*

$$\sup_{l_1 \le l \le l_2} \frac{1}{(l - l_1 + 1)} \left| \sum_{i=l_1}^{l} \hat{\varepsilon}_i \right| = o_P(1). \tag{2.9}$$

The proof of Lemma 1 relies on an application of the well-known Hájek-Rényi inequality and several key inequalities for regularly varying functions from extreme value theory. Note that the scaled partial sum in (2.9) converges to zero in probability uniformly, even when the lower bound $l_1 \ge 1$ does not diverge. This distinction sets our framework apart from the structural break literature, where, for instance, the quantity $\sup_{l_1 \le l \le l_2} \frac{1}{(l-l_1+1)} \left| \sum_{i=l_1}^{l} \hat{\varepsilon}_i \right|$ is typically $O_P(1)$, but not $o_P(1)$ if l_1 is fixed (see, e.g., Bai, 1997b, Lemma A.3). This difference arises because the $\hat{\varepsilon}_i$'s, for $i = 1, \ldots, N$, shrink to zero in probability, resulting in a stronger signal compared to $O_P(1)$ errors typically encountered in regression models with structural breaks.

Assumption 3 (Group size)**.** *There exist $0 < \lambda_{10} < \ldots < \lambda_{m0} < 1$ such that $l_{j0} = [N\lambda_{j0}]$ for $j = 1, \ldots, m$.*

Assumption 3 ensures that the difference between two consecutive segmentation locations is sufficiently large to identify the group memberships. This assumption is commonly adopted in the structural break literature, see, for example, Bai and Perron (1998, Assumption A.5) and Qu and Perron (2007, Assumption A8).

We are now ready to provide a justification for the consistency of our segmentation procedure. To this end, let $\boldsymbol{\lambda}_0 = (\lambda_{10}, \ldots, \lambda_{m0})$ and $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m) = \hat{\boldsymbol{\ell}}/N$ be an estimator of $\boldsymbol{\lambda}_0$. The next theorem establishes the consistency of $\hat{\boldsymbol{\lambda}}$ and (even) $\hat{\boldsymbol{\ell}}$.

Theorem 1 (Consistency and rate of convergence)**.** *Suppose Assumptions 1, 2 and 3 hold. Then, we have* $\left\|\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0\right\|_1 \xrightarrow{p} 0$ *and* $\left\|\hat{\boldsymbol{\ell}} - \boldsymbol{\ell}_0\right\|_1 \xrightarrow{p} 0$ *as* $N \to \infty$.

Theorem 1 establishes that the proportion of group memberships for each group, relative to the total number of individuals, can be consistently estimated. Building on the consistency of $\hat{\boldsymbol{\lambda}}$, it further shows that the segmentation location $\hat{\boldsymbol{\ell}}$ is consistent. This result is stronger than that in the structural break literature, such as Bai (1997b, Proposition 1) and Bai and Perron (1998, Proposition 2), where $\hat{\boldsymbol{\lambda}}$ is N-consistent, but $\hat{\boldsymbol{\ell}}$ is typically not consistent for $\boldsymbol{\ell}_0$. The key intuition here is that the "error" term $\hat{\epsilon}_i = o_P(1)$ for $i = 1, \ldots, N$, as discussed earlier, allowing us to derive a sharper uniform bound for the partial sum of $\{\hat{\epsilon}_i\}$, as provided in Lemma 1.

Admittedly, this result relies on the assumption that the underlying block structure (2.4) is recovered by the pre-ordered individual Hill estimates (2.3). Although this requirement may not hold for small T_i, our simulation results indicate that our procedure maintains high empirical segmentation accuracy and efficient group estimates, even with some degree of incorrect pre-ordering. Aside from this, the conditions required in Theorem 1 are rather mild. Assumptions 1 and 2 merely ensure the consistency of the individual Hill estimators. Notably, the result reveals that with a correct ordering, the identification of clusters offers considerable flexibility in the choice of k_i. This is a key advantage of our framework, which views segmentation as identifying break locations.

We now establish the asymptotic normality of the group Hill estimators, requiring a second-order condition to characterize the convergence rate of the first-order condition in Eq. (2.2).

Assumption 4 (Second-order regular variation condition)**.** *For each* $i = 1, \ldots, N$*, there exist a*

$\rho_i < 0$ *and a positive or negative function* A_i *with* $\lim_{s\to\infty} A_i(s) = 0$ *such that for all* $x > 0$,

$$\lim_{s\to\infty} \frac{\frac{U_i(sx)}{U_i(s)} - x^{\gamma_i}}{A_i(s)} = x^{\gamma_i} \frac{x^{\rho_i} - 1}{\rho_i}. \tag{2.10}$$

A similar assumption to Assumption 4 can be found in, for example, Chen et al. (2022, Condition 4). However, it is important to note that we allow the second-order auxiliary functions A_i to vary across individuals, whereas Chen et al. (2022) impose homogeneity, i.e., $A_i \equiv A$ for all i. This added flexibility complicates the proofs for asymptotic normality, as we require uniform convergence results (such as Lemma S.1.1), which are not needed in Chen et al. (2022).

The following assumption imposes restrictions on the choice of $\tilde{k}_i$ for $i = 1, \ldots, N$, which is essential for constructing our group Hill estimator in (2.8) and for deriving its limiting distribution.

Assumption 5 (Conditions on $\tilde{k}_i$)**.** *For each* $i = 1, \ldots, N$, $\tilde{k}_i/T_i \to 0$ *as* $T_i \to \infty$, *and* $N \max_{1\le i\le N} \exp\left(-\tilde{k}_i/\log(\tilde{k}_i)\right) \to 0$ *as* $N \to \infty$. *Moreover,*

$$\max_{1\le i\le N} \left| A_i\left(\frac{T_i}{\tilde{k}_i}\right)\right| \to 0, \tag{2.11}$$

$$\left(\sum_{i=l_{(j-1)0}+1}^{l_{j0}} \tilde{k}_i^{-1}\right)^{-1/2} \sum_{i=l_{(j-1)0}+1}^{l_{j0}} A_i\left(\frac{T_i}{\tilde{k}_i}\right) = O(1), \tag{2.12}$$

$$\left(\sum_{i=l_{(j-1)0}+1}^{l_{j0}} \tilde{k}_i^{-1}\right)^{-1/2} \sum_{i=l_{(j-1)0}+1}^{l_{j0}} \frac{A_i\left(\frac{T_i}{\tilde{k}_i}\right)}{1-\rho_i} \to B_j, \tag{2.13}$$

as $N \to \infty$, *where* $|B_j| < \infty$ *for* $j = 1, \ldots, G$.

Assumption 5 is required to control the asymptotic bias of the group Hill estimator. The condition $N \max_{1\le i\le N} \exp\left(-\tilde{k}_i/\log(\tilde{k}_i)\right) \to 0$ is slightly stronger than the condition $N \max_{1\le i\le N} \exp(-k_i) \to 0$ required in Assumption 2. The conditions (2.11)-(2.13) serve as the analogs for clustering of the standard assumptions for the asymptotic normality of the individual Hill estimator (de Haan and Ferreira, 2006, Theorem 3.2.5), which, for instance, necessitates

that the limit of $\sqrt{\tilde{k}_i} A_i\left(\frac{T_i}{\tilde{k}_i}\right)$ exists and is finite.

Theorem 2 (Asymptotic normality of group Hill estimator)**.** *Assume that the conditions in Theorem 1, Assumptions 4 and 5 hold. As* $N \to \infty$*,*

$$\frac{l_{j0} - l_{(j-1)0}}{\sqrt{\sum_{i=l_{(j-1)0}+1}^{l_{j0}} \tilde{k}_i^{-1}}} \left(\hat{\gamma}_{g_j}\left(\hat{\boldsymbol{\ell}}, \tilde{\boldsymbol{k}}\right) - \gamma_{g_j}\right) \xrightarrow{d} \mathcal{N}\left(B_j, \gamma_{g_j}^2\right), \qquad j = 1, \ldots G. \tag{2.14}$$

In addition, $\hat{\gamma}_{g_j}\left(\hat{\boldsymbol{\ell}}, \tilde{\boldsymbol{k}}\right)$ *are independent for different* $j = 1, \ldots, G$*.*

Theorem 2 provides a limiting approximation for our group Hill estimator. Compared to the asymptotic result of the distributed Hill estimator in Chen et al. (2022, Theorem 4), our estimation framework differs in two key aspects: first, it accounts for the estimation uncertainty of $\hat{\boldsymbol{\ell}}$; second, it allows for heterogeneous second-order parameters ρ_i.

Moreover, from Theorem 2, we also see the benefits of independently selecting threshold sequences for segmentation and group EVI estimation. Specifically, choosing $\tilde{k}_i$ much smaller than k_i can reduce the bias of $\hat{\gamma}_{g_j}$ while still achieving a faster convergence rate than the individual Hill estimator by exploiting cross-sectional information. To understand the theoretical gain of our group Hill estimator, we consider a special case where group 1 is fully homogeneous. Specifically, for $i = 1, \ldots, l_{10}$, we assume $F_i = F_1$, $T_i = T_1$, and $\tilde{k}_i = \tilde{k}_1$. By Theorem 2, the asymptotic bias and variance of our group estimator are given by $\frac{A_1\left(\frac{T_1}{\tilde{k}_1}\right)}{1-\rho_1}$ and $\frac{\gamma_{g_1}^2}{\tilde{k}_1 l_{10}}$, respectively. As a result, the optimal asymptotic mean squared error (MSE) is $O\left((T_1 l_{10})^{-\frac{2\rho_1}{2\rho_1-1}}\right)$, with the corresponding choice of $\tilde{k}_1$ satisfying $\tilde{k}_1 = O\left(T_1^{\frac{2\rho_1}{2\rho_1-1}} l_{10}^{\frac{1}{2\rho_1-1}}\right)$. For the individual Hill estimator $\hat{\gamma}_1^H(k_1)$, the optimal asymptotic MSE is $O\left(T_1^{-\frac{2\rho_1}{2\rho_1-1}}\right)$ (see de Haan and Ferreira, 2006, Theorem 3.2.5), with the optimal choice of k_1 satisfying $k_1 = O\left(T_1^{\frac{2\rho_1}{2\rho_1-1}}\right)$. Obviously, the optimal MSE of our estimator is of a lower order than that of the individual Hill estimator. Moreover, it suggests that one needs to choose a smaller $\tilde{k}_i$ for the group estimator, compared to the optimal choice of the individual Hill estimator.

3 Simulation study

We examine the empirical accuracy of our segmentation method and then evaluate the performance of the resulting group Hill estimator, assuming the number of groups G is known. Subsequently, a data-driven procedure is investigated to select G.

We consider data generating processes based on two families of heavy-tailed distributions:

(i) Burr distribution, with CDF given by $F(x) = 1 - (1 + x^c)^{-1}$, for $x > 0$, where $c > 0$. This corresponds to an extreme value index $\gamma = 1/c$ and the second order index $\rho = -1$.

(ii) Student's t distribution with degrees of freedom $\nu > 0$, where $\gamma = 1/\nu$ and $\rho = -2/\nu$.

For each group, half of the panel units are generated from the Burr distribution and the other half from the Student's t distribution, with a common extreme value index achieved by setting $\nu = c$. We set the time dimension $T_i = T \in \{1000, 3000\}$ for all panel units. The number of groups is $G \in \{3, 5\}$, with each group containing a balanced number of units $S_G \in \{100, 300\}$, yielding a total of $N = G \cdot S_G$ cross-sectional units. For simplicity, the EVIs are evenly spaced within the interval $[0.2, 1.5]$. For example, when $G = 3$, we set $(\gamma_{g_1}, \gamma_{g_2}, \gamma_{g_3}) = (0.5, 0.85, 1.5)$. All results are reported based on $M = 10^4$ Monte Carlo replicates.

Before presenting the simulation results, we first highlight a key implementation aspect, namely, the joint minimization problem in (2.7). Since our model fits within a structural break framework, existing computational methods can be readily applied to our segmentation procedure. We advocate using the fast dynamic programming approach proposed by Bai and Perron (2003), which globally minimizes the SSR in (2.7) and has an operational complexity of order $O(N^2)$ rather than $O(N^m)$. The sequential estimation procedure proposed by Bai (1997a) offers a computationally efficient alternative to the joint estimation.[1] However, joint minimization is crucial in our context for the following reason. In finite samples, due to the

[1]Specifically, one begins by assuming a single segmentation point and obtains an estimate by minimizing the SSR in (2.7) for $m = 1$. Once the first location is estimated, the sample is split into two sub-samples, and a one-location model is applied to each. The second location is then selected as the one that results in the largest reduction in SSR. This process continues iteratively until the predetermined number of segmentation locations is reached.

Table 1: Group-wise empirical accuracy (in percent).

			$T = 1000$					$T = 3000$				
G	S_G	R_k	g_1	g_2	g_3	g_4	g_5	g_1	g_2	g_3	g_4	g_5
3	100	9%	100	99.92	98.39			100	100	100		
		12%	99.98	99.97	99.33			100	100	100		
	300	9%	100	99.94	98.21			100	100	100		
		12%	99.99	99.99	99.19			100	100	100		
5	100	9%	97.05	98.53	92.38	79.24	74.66	99.02	99.92	99.74	97.17	94.97
		12%	79.03	98.26	96.49	85.35	80.80	83.62	99.59	99.90	98.46	96.42
	300	9%	97.42	98.58	92.41	79.40	75.01	99.62	99.96	99.78	97.08	94.89
		12%	81.31	98.58	96.53	85.36	80.88	89.51	99.86	99.95	98.42	96.27

estimation uncertainty, the ordered individual Hill estimates $\hat{\gamma}_1^H \leq \hat{\gamma}_2^H \leq \cdots \leq \hat{\gamma}_N^H$ often resemble models with trend breaks (Perron and Zhu, 2005; Beutner et al., 2023) rather than level shifts. In the context of trend breaks, Yang (2017) shows that consistent estimation of break locations can only be guaranteed if the number of breaks is correctly specified. In other words, a sequential estimation procedure proposed by Bai (1997a) may yield less accurate estimates of $\boldsymbol{\ell}_0$ in finite samples.

3.1 Empirical accuracy of segmentation

Recall that the asymptotic properties derived in Section 2.2 rely on the assumption that the Hill estimates preserve the true group ordering, that is, the estimates for group 1 are always smaller than those for group 2, for example. However, this is not guaranteed in finite samples. In this subsection, we assess the segmentation accuracy based on the Hill estimates, allowing for potential misordering.

As also discussed in Section 2.2, the condition on k_i is fairly general. For finite samples, we recommend choosing $k_i = [R_k\, T_i]$, with $R_k \in \{9\%, 12\%\}$, a rather common choice in extreme value studies. We assess grouping performance using group-wise empirical accuracy, defined as $M^{-1} \sum_{m=1}^{M} S_G^{-1} C_j^{(m)}$, where $C_j^{(m)}$ is the number of correctly classified individuals in group $j = 1, \ldots, G$ in the mth repetition. Table 1 presents the results.

(i) We observe generally satisfactory performance across settings. When $G = 3$, the empirical accuracy exceeds 98% for all settings. As the number of groups increases to five, the problem becomes more challenging. Nevertheless, for sufficiently large sample size ($T = 3000$), the accuracy remains above 95% in most cases.

(ii) When $(G, T) = (5, 1000)$, the segmentation accuracy becomes more sensitive to the choice of R_k (or, equivalently, k_i). Notably, group 1 exhibits higher accuracy when a smaller k_i (corresponding to $R_k = 9\%$) is used, whereas groups 4 and 5 benefit from a larger k_i (corresponding to $R_k = 12\%$) . This pattern aligns with theoretical expectations: the optimal choice of k_i for the Hill estimator depends on the second-order parameter ρ_i (see the discussion below Theorem 2). In group 1, where $\rho_i \in (-1, -0.4)$, the smaller absolute value of ρ_i suggests that a smaller k_i is preferable. In contrast, groups 4 and 5, which have larger $|\rho_i|$, require larger k_i for more accurate estimation.

Based on the observations above, empirical segmentation accuracy could potentially be improved by using different k_i values for different individuals. For example, one might consider data-driven approach of threshold selection. We leave this possibility for future research.

3.2 Finite sample performance of group Hill estimator

We assess the improvement of our group EVI estimator given by (2.8) over the individual Hill estimator defined in (2.3). We focus on the setting $T_i = 1000$ and $G \in \{3, 5\}$ and fix $k_i = [12\%T_i] = 120$. Motivated by Theorem 2 and the subsequent discussion, a smaller choice of $\tilde{k}_i$ is preferable when constructing the group Hill estimator. We choose $\tilde{k}_i = [\widetilde{R}_k T_i]$, where $\widetilde{R}_k \in \{2\%, 3\%, 4\%\}$. To ensure a fair comparison for individual Hill estimator $\hat{\gamma}_i^H(k_i^H)$, we set $k_i^H = [R_H T_i]$, with $R_H \in \{5\%, 7\%\}$. Finally, we use mean absolute errors (MAE) to evaluate

Table 2: Empirical MAE (in percent) of group and individual Hill estimators for $T = 1000$.

		$\widetilde{R}_k$			R_H				$\widetilde{R}_k$			R_H	
(G, S_G)	Group	2%	3%	4%	5%	7%	(G, S_G)	Group	2%	3%	4%	5%	7%
	1	3.64	4.54	5.36	7.05	9.70		1	3.64	4.54	5.37	7.04	9.69
$(3, 100)$	2	1.66	1.55	1.67	17.09	8.32	$(3, 300)$	2	1.12	1.24	1.51	17.08	8.32
	3	2.84	2.48	2.42	8.40	14.68		3	1.80	1.79	1.99	8.39	14.67
	1	2.39	3.00	3.56	7.05	8.32		1	2.57	3.23	3.84	7.04	8.32
	2	3.38	2.86	2.47	6.39	6.09		2	2.75	2.15	1.72	6.40	6.10
$(5, 100)$	3	2.19	2.25	2.46	9.71	8.40	$(5, 300)$	3	1.86	2.12	2.45	9.70	8.39
	4	4.11	4.23	4.48	13.40	11.51		4	3.85	4.15	4.46	13.38	11.50
	5	4.90	5.05	5.41	17.07	14.68		5	4.42	4.82	5.27	17.10	14.69

performance:

$$\text{MAE}_j^{\text{grp.}} = \frac{1}{M}\sum_{m=1}^{M}\left|\hat{\gamma}_{g_j}^m - \gamma_{g_j}\right|, \qquad \text{MAE}_j^{\text{ind.}} = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{S_G}\sum_{i=1}^{S_G}\left|\hat{\gamma}_i^{H,m} - \gamma_{g_j}\right|, \quad j = 1, \ldots, G,$$

where $\hat{\gamma}_{g_j}^m = \hat{\gamma}_{g_j}^m\big(\hat{\boldsymbol{\ell}}, \tilde{\boldsymbol{k}}\big)$ represents the group Hill estimates in the mth Monte Carlo simulation, and similarly for the individual Hill estimates, $\hat{\gamma}_i^{H,m} = \hat{\gamma}_i^{H,m}(k_i^H)$. Table 2 present the results for $T = 1000$, with all values expressed as percentages. Similar observations hold for $T = 3000$, see Online Appendix S.2.

(iii) For all simulation settings, the MAE of our proposed group estimator $\hat{\gamma}_{g_j}$ consistently outperforms the individual Hill estimator, with MAE ratios (group divided by individual) ranging from approximately 10% to 70%. The boxplots in Figure 1 clearly illustrate that our group estimator significantly improves empirical efficiency in terms of MAE, with the group estimates consistently exhibiting lower MAEs compared to the individual estimates.

(iv) When $G = 3$, Table 2 indicates that the MAE decreases as the number of group units S_G increases from 100 to 300. This aligns with the theoretical rate of convergence in Theorem 2, which becomes $\sqrt{S_G\tilde{k}_i}$ in the setting here. However, this pattern does not hold when $G = 5$. One contributing factor is the potential inaccuracy in the ordering and segmentation stages, which can adversely affect the performance of the group Hill estimator.

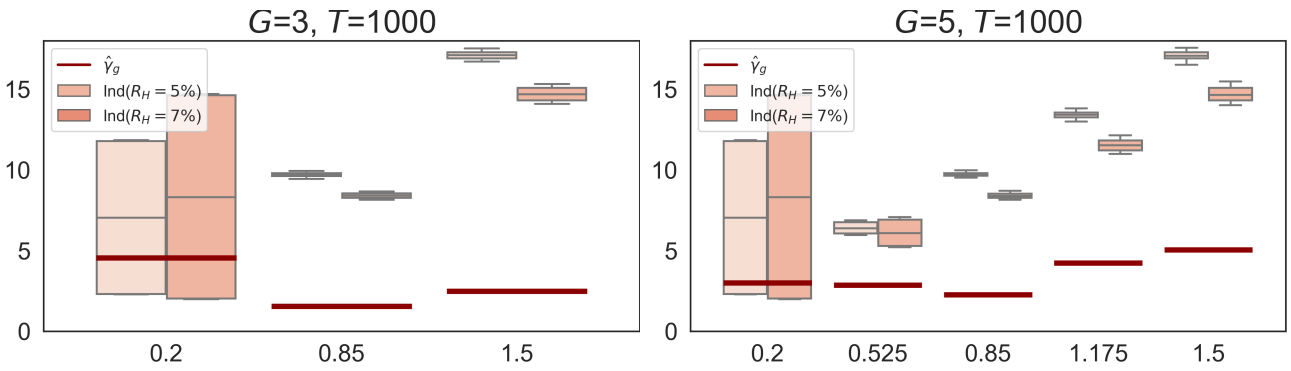


Figure 1: Group-wise MAEs (in percent) for the group Hill estimates constructed in (2.8) (solid red lines) and boxplots of individual Hill estimates (2.3). The left figure corresponds to $G = 3$ groups, while the right one represents $G = 5$ groups. Here, $(T, S_G, \widetilde{R}_k) = (1000, 100, 3\%)$.

3.3 Determining the number of groups

Thus far, we have assumed that the number of groups G is known. When G is unknown, we adapt the elbow method, a widely used technique in machine learning. This method relies on the objective function value, SSR, in (2.7). Specifically, we focus on the reduction in SSR when an additional group is added, which decreases in absolute value and eventually approaches zero. The optimal number of groups, denoted as $\hat{G}$, is determined based on the fraction of SSR reductions from adding an additional segment and a predefined threshold $\mathcal{T}$:

$$\hat{G} = \underset{m\in\{1,\ldots,m_{\max}\}}{\operatorname{argmin}} \left\{ \frac{\mathrm{SSR}(\hat{\boldsymbol{\ell}}_m) - \mathrm{SSR}(\hat{\boldsymbol{\ell}}_{m+1})}{\mathrm{SSR}(\hat{\boldsymbol{\ell}}_m) - \mathrm{SSR}_0} \leq \mathcal{T} \right\} + 1, \qquad m_{\max} \in \mathbb{Z}^+, \tag{3.1}$$

where SSR_0 represents the SSR without any segmentation, $\mathcal{T} \in \{0.015, 0.02, 0.025, 0.03, 0.035\}$ and $m_{\max} = 7$. The DGP remains the same as before. We continue to use $k = [12\% \, T]$ as the first-stage threshold. The results are reported in Table 3, where we reduce the number of Monte Carlo replications to $M = 2000$ to reduce computational cost.

The table shows the strong performance of the elbow method in accurately selecting the number of groups, achieving at least 80% accuracy and often 100% for a threshold $\mathcal{T} \in [0.02, 0.03]$. However, this selection method lacks a formal theoretical foundation. Developing a theoretically justified approach remains an open question for future research.

Table 3: Empirical accuracy (in percent) of the elbow method for determining G.

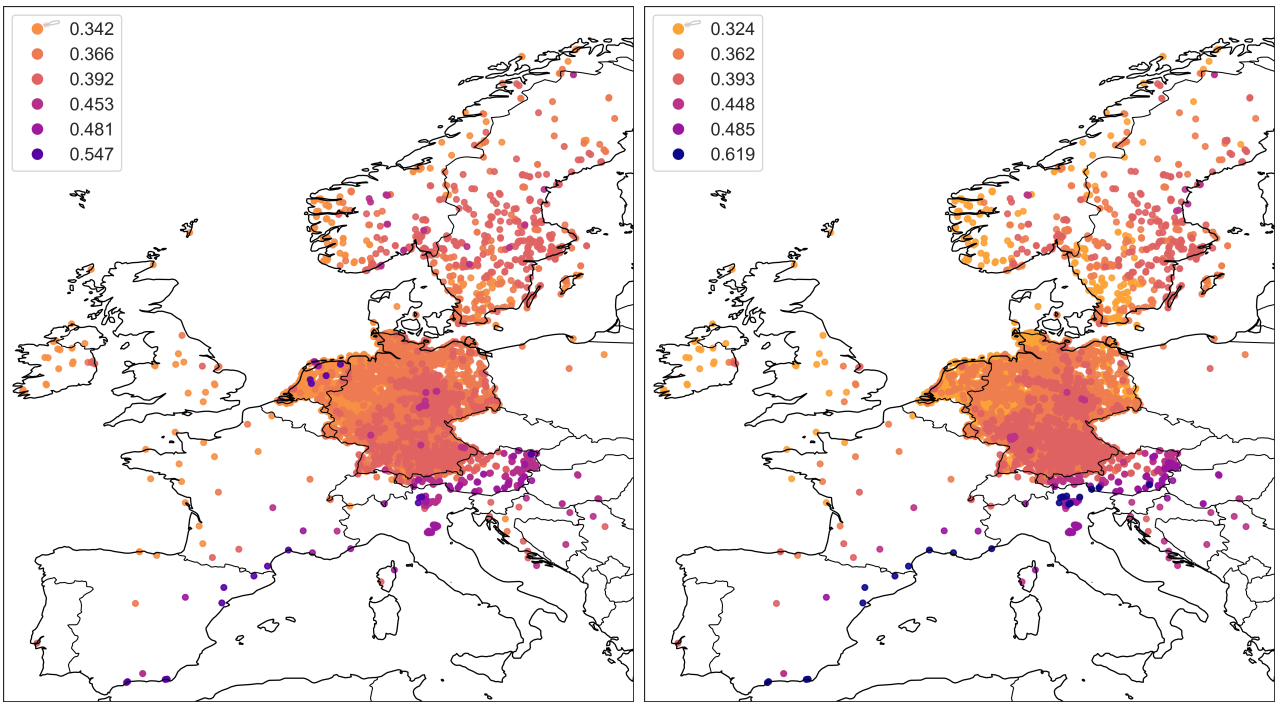
			$\mathcal{T}$				
G	T	S_G	0.015	0.02	0.025	0.03	0.035
3	1000	100	21.25	89.10	99.80	100	100
		300	8.90	97.60	100	100	100
	3000	100	100	100	100	100	100
		300	100	100	100	100	100
5	1000	100	99.45	100	100	99.60	71.50
		300	100	100	100	100	96.55
	3000	100	100	100	100	100	100
		300	100	100	100	100	100

4 Empirical analysis of rainfall data

We illustrate our proposed clustering framework using daily precipitation data in Europe, collected from the European Climate Assessment & Dataset, see Klein Tank et al. (2002).[2] In particular, we are interested in the change of the tail behavior of precipitation volume over time to assess whether observed shifts toward heavier extreme rainfall align with the IPCC's recent findings of increasing intensity and frequency of extreme precipitation events across Europe due to climate change (Bednar-Friedl et al., 2022).

We analyze daily precipitation records, measured in 0.1 mm, from 1950 to 2020, and we split the dataset into two distinct periods accordingly: 1950-1985 (Period 1) and 1986-2020 (Period 2). An initial dataset comprising 17,128 meteorological stations was collected, after which the scope of analysis was narrowed to stations that met the following criteria. First, stations are required to maintain consecutive records spanning both periods. Second, they are located in Europe, further defined by a longitude range of -10° to 21° and a latitude range of 35° to 72°. Given the climatic importance of winter precipitation (Kundzewicz et al., 2006) and to avoid addressing seasonal patterns, we focus exclusively on winter observations (December, January, and February). To ensure a sufficient number of extreme observations, stations are further

[2] https://www.ecad.eu/dailydata/predefinedseries.php, blended ECA Daily precipitation amount RR, accessed on January 27, 2025.



(a) Period 1 (1950–1985) **(b)** Period 2 (1986–2020)

Figure 2: Maps depicting the weather stations in our sample for both periods. The dots represent grouped estimates ($G = 6$) for the EVI of precipitation volume, with darker shades indicating higher values.

required to have recorded a minimum of 3,000 winter days in each period (approximately 35 years). This filtering step reduces the dataset to 4,769 stations. For each period, the moment estimator for extreme value indices (Dekkers et al., 1989) is applied using 12% of each individual station's sample size as a threshold to detect negative EVIs. The final dataset includes only the intersection of stations meeting all criteria, resulting in a total of 4,735 stations with positive EVIs for further analysis.

We estimate the EVIs of the precipitation volume for all weather stations separately in the two time periods using the Hill estimator. Building on the recommendations from the simulation results, the first-stage threshold k_i is set as $[12\% \, T_i]$, and the group EVI estimates are obtained with $\tilde{k}_i = [3\% \, T_i]$. Moreover, to determine the number of groups within each period, we set $m_{\max} = 14$ in (3.1), allowing for a maximum of 15 groups in each period. The elbow method with a threshold of 0.025 yields optimal $G = 6$ for both periods. As a robustness check, we also include the results for $G = 4$ and $G = 5$ in the Online Appendix S.3.

Table 4: Transition matrix for $G = 6$ between period 1 (rows) and period 2 (columns).

$p_1 \setminus p_2$	1	2	3	4	5	6
1	766	468	68	0	0	0
2	379	1067	506	0	0	0
3	41	386	693	44	0	0
4	2	4	30	100	55	4
5	5	0	0	15	72	11
6	0	4	0	2	0	12

Figure 2 shows the resulting estimates on a map of the selected area. We observe that for the first period, EVI estimates range from 0.342 (group 1) to 0.547 (group 6), while in the second period, the range expands from 0.324 to 0.619. Spatially, clusters of high EVI values are consistently identified in the Alpine region (Northern Italy, Austria) and along the Mediterranean coast in France and Spain during both periods. Notably, in the first period, some high EVI estimates in the Netherlands may reflect the influence of the exceptional storm in 1953. For Germany, the initial period displays relatively homogeneous EVI values predominantly within groups 2 and 3. In contrast, the second period reveals clearer spatial differentiation between the eastern and western parts and the rest of the country, with eastern and western regions experiencing reduced rainfall extremes. A comparable trend is observed in Scandinavia: while period 1 shows relatively low spatial heterogeneity, period 2 exhibits increased variability, characterized by drier conditions along coastal regions.

The transition matrix between period 1 and period 2 is shown in Table 4. It illustrates the persistence in rainfall intensity group assignments, indicated by large diagonal entries. However, notable shifts also occur, with many stations transitioning upward to groups characterized by more intense rainfall (e.g., from group 2 to group 3 and from group 4 to group 5). Conversely, some downward transitions are also present (e.g., from group 2 to group 1), but comparably less frequent. These shifts not only indicate an overall increase in extreme rainfall intensities across Europe, but also provide evidence for spatial shifts in rainfall patterns.

We further conduct a sequence of t-tests for $H_0 : \gamma_{g_j}^{p_1} = \gamma_{g_j}^{p_2}$ against $H_1 : \gamma_{g_j}^{p_1} \neq \gamma_{g_j}^{p_2}$ based

Table 5: Test statistics (4.1) with significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

group 1	group 2	group 3	group 4	group 5	group 6
13.68***	3.59***	0.53	0.98	0.64	3.97***

on Theorem 2, where $\gamma_{g_j}^{p_1}$ represents the j_{th} group's EVI in period 1 and $\gamma_{g_j}^{p_2}$ in period 2, with $j = 1, \ldots, G$. Assuming the observations are independent across the two periods, under the null hypothesis, we have

$$\frac{\hat{\gamma}_{g_j}^{p_1} - \hat{\gamma}_{g_j}^{p_2}}{\sqrt{\frac{\left(\hat{\gamma}_{g_j}^{p_1}\right)^2 \left(\sum_{i=\tilde{l}_{j-1}^{p_1}+1}^{\tilde{l}_j^{p_1}} \tilde{k}_i^{-1}\right)}{\left(\tilde{l}_j^{p_1} - \tilde{l}_{j-1}^{p_1}\right)^2} + \frac{\left(\hat{\gamma}_{g_j}^{p_2}\right)^2 \left(\sum_{i=\tilde{l}_{j-1}^{p_2}+1}^{\tilde{l}_j^{p_2}} \tilde{k}_i^{-1}\right)}{\left(\tilde{l}_j^{p_2} - \tilde{l}_{j-1}^{p_2}\right)^2}}} \xrightarrow{d} \mathcal{N}(0,1), \tag{4.1}$$

where $\hat{l}_j^{p_1}$ is the estimate of l_{j0} in period 1, and similarly, $\hat{l}_j^{p_2}$ is the estimate for period 2. At the 1% significance level, our analysis rejects the null hypothesis for groups 1, 2, and 6, revealing a distinct pattern in the EVI estimates: a notable increase in group 6 and a significant decrease in groups 1 and 2; see Table 5. These results further substantiate our evidence for an underlying trend toward increasing variability and intensification of extreme rainfall events across Europe.

5 Conclusion

We analyzed a large panel (i.e., $(N, T_i) \to \infty$) of individuals sharing a common set of extreme value indices (EVIs) and proposed an easy-to-implement clustering framework to identify unknown group memberships. Specifically, we ranked individual Hill estimates and segmented the groups by minimizing the quadratic distance between individual estimates and their respective group averages. Our approach is inspired by the estimation of break locations in the structural break literature; however, establishing our asymptotic framework requires multiple key results from extreme value theory. We theoretically proved that our method consistently estimates segmentation locations and thereby recovers the underlying group identities, which is a stronger

result than those typically found in the structural break literature. After successfully forming the groups, we introduced a group estimator and proved its asymptotic normality. Our approach offers considerable flexibility by enabling independent selection of thresholds for both segmentation and group EVI estimation. This flexibility enables our group estimator to achieve faster convergence rates without increasing bias. The simulation results reveal that the group estimator substantially outperforms individual Hill estimators in terms of accuracy and efficiency.

We applied our method to analyze a rainfall dataset collected from 4,735 stations across Europe. Our findings suggest an increase in heterogeneity, regional differentiation, and variability in extreme precipitation patterns, consistent with assessments by the IPCC and the European Environmental Agency (European Environmental Agency, 2021). This analysis highlights the importance of accounting for regional heterogeneity when modeling extreme precipitation risks and provides a framework for tracking changes in the context of ongoing climate change.

Looking ahead, developing a theoretically justified procedure for selecting the number of groups in our framework remains an interesting avenue for future research.

References

Aue, A. and L. Horváth (2013). Structural breaks in time series. *Journal of Time Series Analysis 34*(1), 1–16.

Bader, B., J. Yan, and X. Zhang (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Annals of Applied Statistics 12*(1), 310–329.

Bai, J. (1997a). Estimating multiple breaks one at a time. *Econometric Theory 13*(3), 315–352.

Bai, J. (1997b). Estimation of a change point in multiple regression models. *Review of Economics and Statistics 79*, 551–563.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

Bednar-Friedl, B., G. Biesbroek, J. Carnicer, et al. (2022). Europe: from chapters and cross-chapter papers. In *Climate Change 2022: Impacts, Adaptation and Vulnerability*, pp. 1817–1927. Cambridge University Press.

Beirlant, J., Y. Goegebeur, J. Segers, and J. L. Teugels (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.

Beutner, E., Y. Lin, and S. Smeekes (2023). GLS estimation and confidence sets for the date of a single break in models with trends. *Econometric Reviews 42*(2), 195–219.

Chen, L., D. Li, and C. Zhou (2022). Distributed inference for the extreme value index. *Biometrika 109*(1), 257–264.

Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis 76*(2), 226–248.

de Carvalho, M., R. Huser, and R. Rubio (2023). Similarity-based clustering for patterns of extreme values. *Stat 12*(1), e560.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer: New York.

de Haan, L. and C. Zhou (2021). Trends in extreme value indices. *Journal of the American Statistical Association 116*(535), 1265–1279.

Dekkers, A. L., J. H. Einmahl, and L. De Haan (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 1833–1855.

Drees, H., A. Janßen, S. I. Resnick, and T. Wang (2020). On a minimum distance procedure for threshold selection in tail analysis. *SIAM Journal on Mathematics of Data Science 2*(1), 75–102.

Dupuis, D. J., S. Engelke, and L. Trapin (2023). Modeling panels of extremes. *Annals of Applied Statistics 17*(1), 498–517.

Einmahl, J. H. and Y. He (2022). Extreme value estimation for heterogeneous data. *Journal of Business & Economic Statistics 41*(1), 255–269.

Einmahl, J. H. and Y. He (2023). Extreme value inference for heterogeneous power law data. *Annals of Statistics 51*(3), 1331–1356.

European Environmental Agency (2021). Europe's changing climate hazards - an index-based interactive EEA report. *EEA Web Report*.

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

Hou, Y., X. Leng, L. Peng, and Y. Zhou (2024). Panel quantile regression for extreme risk. *Journal of Econometrics 240*(1), 105674.

Ivette Gomes, M., L. De Haan, and L. H. Rodrigues (2008). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society Series B: Statistical Methodology 70*(1), 31–52.

Klein Tank, A. M., J. Wijngaard, G. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, et al. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *International Journal of Climatology: A Journal of the Royal Meteorological Society 22*(12), 1441–1453.

Kundzewicz, Z. W., M. Radziejewski, and I. Pinskwar (2006). Precipitation extremes in the changing climate of europe. *Climate Research 31*(1), 51–58.

Li, J., L. Chen, W. Wang, and W. B. Wu (2024). ℓ^2 inference for change points in high-dimensional time series via a Two-Way MOSUM. *Annals of Statistics 52*(2), 602–627.

Perron, P. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics 1*(2), 278–352.

Perron, P. and X. Zhu (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics 129*, 65–119.

Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica 75*(2), 459–502.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press: Cambridge.

Yang, J. (2017). Consistency of trend break point estimator with underspecified break number. *Econometrics 5*(1).

A Proofs of main results

We provide the proofs for Theorem 1 and Theorem 2. Additional proofs, including that of Lemma 1, are collected in Online Appendix S.1.

Proof of Theorem 1. We first establish the consistency of $\hat{\boldsymbol{\lambda}}$, upon which we then derive its rate of convergence. For convenience, we write $\operatorname{argmin}_{\boldsymbol{\ell}} = \operatorname{argmin}_{\boldsymbol{\ell}\in\boldsymbol{\mathcal{L}}}$ without ambiguity. Recall that $l_0 = 0$ and $l_G = N$. For any candidate segmenting location $\boldsymbol{\ell} = (l_1, \ldots, l_m)^\top$, define

$$\hat{\gamma}(l_{j-1}, l_j) = \frac{1}{l_j - l_{j-1}} \sum_{i=l_{j-1}+1}^{l_j} \hat{\gamma}_i^H, \qquad j = 1, \ldots, G. \tag{A.1}$$

A straightforward algebraic calculation shows that $\hat{\gamma}(l_{j-1}, l_j) = \hat{\gamma}_{g_j}(\boldsymbol{\ell})$ and the sum of squared residuals can be expressed as

$$\begin{aligned}
\mathrm{SSR}(\boldsymbol{\ell}) =& \frac{1}{N} \sum_{j=1}^{G} \sum_{i=l_{j-1}+1}^{l_j} \big(\hat{\gamma}(l_{j-1}, l_j) - \hat{\gamma}_i^H\big)^2 \\
=& \frac{1}{N} \sum_{j=1}^{G} \left((l_j - l_{j-1}) \hat{\gamma}(l_{j-1}, l_j)^2 - 2\hat{\gamma}(l_{j-1}, l_j) \sum_{i=l_{j-1}+1}^{l_j} \hat{\gamma}_i^H + \sum_{i=l_{j-1}+1}^{l_j} \big(\hat{\gamma}_i^H\big)^2 \right) \\
=& -\frac{1}{N} \sum_{j=1}^{G} (l_j - l_{j-1}) \hat{\gamma}(l_{j-1}, l_j)^2 + \frac{1}{N} \sum_{i=1}^{N} \big(\hat{\gamma}_i^H\big)^2 .
\end{aligned}$$

This implies that $\hat{\boldsymbol{\ell}} = \operatorname{argmin}_{\boldsymbol{\ell}} \mathrm{SSR}(\boldsymbol{\ell}) = \arg\max_{\boldsymbol{\ell}} N^{-1} \sum_{j=1}^{G} (l_j - l_{j-1}) \hat{\gamma}(l_{j-1}, l_j)^2$. Since $\hat{\gamma}_i^H = \gamma_i + \hat{\varepsilon}_i$, we can write $N^{-1} \sum_{j=1}^{G} (l_j - l_{j-1}) \hat{\gamma}(l_{j-1}, l_j)^2 = S_{1N}(\boldsymbol{\ell}) + S_{2N}(\boldsymbol{\ell}) + S_{3N}(\boldsymbol{\ell})$, where

$$\begin{aligned}
S_{1N}(\boldsymbol{\ell}) &= \sum_{j=1}^{G} \frac{1}{N(l_j - l_{j-1})} \left(\sum_{i=l_{j-1}+1}^{l_j} \gamma_i \right)^2, \quad S_{2N}(\boldsymbol{\ell}) = \sum_{j=1}^{G} \frac{2}{N(l_j - l_{j-1})} \sum_{i=l_{j-1}+1}^{l_j} \gamma_i \sum_{i=l_{j-1}+1}^{l_j} \hat{\varepsilon}_i, \\
S_{3N}(\boldsymbol{\ell}) &= \sum_{j=1}^{G} \frac{1}{N(l_j - l_{j-1})} \left(\sum_{i=l_{j-1}+1}^{l_j} \hat{\varepsilon}_i \right)^2 .
\end{aligned}$$

Part I. Consistency of $\hat{\boldsymbol{\lambda}}$: Recall $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\ell}}/N$. Viewing $\hat{\boldsymbol{\lambda}}$ as an M-estimator, that is

$$\hat{\boldsymbol{\lambda}} = \operatorname*{argmax}_{\boldsymbol{\lambda}} \Big(S_{1N}(N\boldsymbol{\lambda}) + S_{2N}(N\boldsymbol{\lambda}) + S_{3N}(N\boldsymbol{\lambda}) \Big), \tag{A.2}$$

we shall apply Theorem 5.7 from van der Vaart (2000) to obtain its consistency. Note that, without loss of generality, we omit taking integer parts in (A.2), assuming that $N\boldsymbol{\lambda}$ is already

a vector of integers. We show that the deterministic term $S_{1N}(N\boldsymbol{\lambda})$ dominates the objective function (A.2) asymptotically. Let $\lambda_0 = \lambda_{00} = 0$ and $\lambda_G = \lambda_{G0} = 1$ and define a piecewise constant function g given by

$$g(x) = \sum_{j=1}^{G} \gamma_{g_j} \mathbb{1}\{\lambda_{(j-1)0} < x \leq \lambda_{j0}\}, \qquad x \in (0,1). \tag{A.3}$$

Using (A.3), one has

$$\begin{aligned}\sum_{i=l_{j-1}+1}^{l_j} \gamma_i &= \sum_{i=l_{j-1}+1}^{l_j} \sum_{k=1}^{G} \gamma_{g_k} \mathbb{1}\{N\lambda_{(k-1)0} < i \leq N\lambda_{k0}\} \\ &= \sum_{k=1}^{G} \gamma_{g_k} \int_{l_{j-1}+1}^{l_j} \mathbb{1}\{N\lambda_{(k-1)0} < x \leq N\lambda_{k0}\}\,\mathrm{d}x = N\int_{\lambda_{j-1}}^{\lambda_j} g(x)\,\mathrm{d}x.\end{aligned}$$

Then, S_{1N} can be equivalently written as

$$S_{1N}(N\boldsymbol{\lambda}) = \sum_{j=1}^{G} \frac{1}{\lambda_j - \lambda_{j-1}} \left(\int_{\lambda_{j-1}}^{\lambda_j} g(x)\,\mathrm{d}x\right)^2 =: S(\boldsymbol{\lambda}).$$

By applying Cauchy-Schwarz inequality, we obtain

$$S(\boldsymbol{\lambda}) \leq \sum_{j=1}^{G} \frac{1}{\lambda_j - \lambda_{j-1}} \int_{\lambda_{j-1}}^{\lambda_j} 1\,\mathrm{d}x \int_{\lambda_{j-1}}^{\lambda_j} g^2(x)\,\mathrm{d}x = \int_0^1 g^2(x)\,\mathrm{d}x = S(\boldsymbol{\lambda}_0),$$

where the equality holds if and only if $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$. Thus it holds that for any $\epsilon > 0$,

$$\sup_{\boldsymbol{\lambda}:\, \|\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0\|_1 > \epsilon} S_{1N}(N\boldsymbol{\lambda}) < S(\boldsymbol{\lambda}_0).$$

In viewing of Theorem 5.7 from van der Vaart (2000), it suffices to prove that $\sup_{\boldsymbol{\lambda}} \left|S_{2N}(N\boldsymbol{\lambda}) + S_{3N}(N\boldsymbol{\lambda})\right| = o_P(1)$. However, this is immediate using Lemma 1. Specifically, due to the ascending order of γ_{g_i}, we have

$$\sup_{\boldsymbol{\lambda}} |S_{2N}(N\boldsymbol{\lambda})| = \sup_{\boldsymbol{\ell}} |S_{2N}(\boldsymbol{\ell})| \leq 2\gamma_{g_G} \sup_{\boldsymbol{\ell}} \sum_{j=1}^{G} \frac{1}{N} \left| \sum_{i=l_{j-1}+1}^{l_j} \hat{\varepsilon}_i \right| = o_P(1).$$

The $o_P(1)$ above holds because $\sup_{l_{j-1}+1 \leq l \leq l_j} N^{-1} \left|\sum_{i=l_{j-1}+1}^{l_j} \hat{\varepsilon}_i\right| = o_P(1)$ by Lemma 1 and $G < \infty$ fixed. Similarly, Lemma 1 implies that $\sup_{\boldsymbol{\lambda}} \left|S_{3N}(N\boldsymbol{\lambda})\right| = o_P(1)$.

Part II. Consistency of $\hat{\boldsymbol{\ell}}$: Given the consistency result on $\hat{\boldsymbol{\lambda}}$, it is sufficient to show that for some $\epsilon > 0$ and $M \geq 1$,

$$P\left(M \leq \left\|\hat{\boldsymbol{\ell}} - \boldsymbol{\ell}_0\right\|_1 \leq \epsilon N\right) \to 0, \tag{A.4}$$

as $T_i \to \infty$ (for $i = 1, \ldots, N$) and $N \to \infty$. Denote $V_{M,N\epsilon} = \{\boldsymbol{\ell} : M \leq \|\boldsymbol{\ell} - \boldsymbol{\ell}_0\|_1 \leq N\epsilon\}$, for

$$\epsilon = \frac{1}{4} \min_{1 \leq j \leq G} \left(\lambda_{j0} - \lambda_{(j-1)0}\right) \min\left(1, \frac{\min_{2 \leq j \leq G}(\gamma_{g_j} - \gamma_{g_{j-1}})}{\max_{2 \leq j \leq G}(\gamma_{g_j} - \gamma_{g_{j-1}})}\right). \tag{A.5}$$

We have

$$\begin{aligned}
&P\left(\hat{\boldsymbol{\ell}} \in V_{M,N\epsilon}\right) \leq P\left(\inf_{\boldsymbol{\ell} \in V_{M,N\varepsilon}} \mathrm{SSR}(\boldsymbol{\ell}) \leq \mathrm{SSR}(\boldsymbol{\ell}_0)\right) \\
=&P\left(\inf_{\boldsymbol{\ell} \in V_{M,N\varepsilon}} \left(-S_{1N}(\boldsymbol{\ell}) - S_{2N}(\boldsymbol{\ell}) - S_{3N}(\boldsymbol{\ell})\right) \leq -S_{1N}(\boldsymbol{\ell}_0) - S_{2N}(\boldsymbol{\ell}_0) - S_{3N}(\boldsymbol{\ell}_0)\right) \\
=&P\left(\inf_{\boldsymbol{\ell} \in V_{M,N\varepsilon}} \sum_{i=1}^{3} \left(S_{iN}(\boldsymbol{\ell}_0) - S_{iN}(\boldsymbol{\ell})\right) \leq 0\right) =: p_N.
\end{aligned}$$

We obtain $p_N \to 0$ by showing that,

$$\inf_{\boldsymbol{\ell} \in V_{M,N\varepsilon}} \frac{N\left(S_{1N}(\boldsymbol{\ell}_0) - S_{1N}(\boldsymbol{\ell})\right)}{\|\boldsymbol{\ell} - \boldsymbol{\ell}_0\|_1} \geq \frac{3}{5} \min_{1 \leq j \leq m} \left(\gamma_{g_{j+1}} - \gamma_{g_j}\right)^2, \tag{A.6}$$

$$\sup_{\boldsymbol{\ell} \in V_{M,N\varepsilon}} \frac{N\left|S_{iN}(\boldsymbol{\ell}_0) - S_{iN}(\boldsymbol{\ell})\right|}{\|\boldsymbol{\ell} - \boldsymbol{\ell}_0\|_1} = o_P(1), \qquad i = 2, 3. \tag{A.7}$$

We first establish (A.6). For convenience, let $\tilde{\boldsymbol{\ell}}(i) := \left(l_1, \ldots, l_i, l_{(i+1)0}, \ldots, l_{m0}\right)^\top$ for $i = 1, \ldots, m$, where the first i entries are taken from $\boldsymbol{\ell}$, and the remaining ones come from the true location parameter $\boldsymbol{\ell}_0$. Moreover, let $\tilde{\boldsymbol{\ell}}(0) = \boldsymbol{\ell}_0$. By construction, for any $\boldsymbol{\ell} \in V_{M,N\varepsilon}$, we have $l_j < l_{(j+1)0}$, ensuring that the entries in $\tilde{\boldsymbol{\ell}}(i)$ remain in ascending order. Then, by repeatedly adding and subtracting, we can write

$$S_{1N}(\boldsymbol{\ell}_0) - S_{1N}(\boldsymbol{\ell}) = \sum_{j=1}^{m} \left(S_{1N}\big(\tilde{\boldsymbol{\ell}}(j-1)\big) - S_{1N}\big(\tilde{\boldsymbol{\ell}}(j)\big)\right). \tag{A.8}$$

Then, Eq. (A.6) follows from the next inequality:

$$N\left(S_{1N}\big(\tilde{\boldsymbol{\ell}}(j-1)\big) - S_{1N}\big(\tilde{\boldsymbol{\ell}}(j)\big)\right) \geq \frac{3}{5}|l_j - l_{j0}| \left(\gamma_{g_{j+1}} - \gamma_{g_j}\right)^2, \qquad j = 1, \ldots, m. \tag{A.9}$$

We present the proof for $j = 2$, as the argument is simpler for $j = 1$ and extends analogously to

other values of j. Note that

$$N\Big(S_{1N}\big(\tilde{\boldsymbol{\ell}}(1)\big)-S_{1N}\big(\tilde{\boldsymbol{\ell}}(2)\big)\Big)=\left[\frac{1}{l_{20}-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right)^2+\frac{1}{l_{30}-l_{20}}\left(\sum_{i=l_{20}+1}^{l_{30}}\gamma_i\right)^2\right]$$
$$-\left[\frac{1}{l_2-l_1}\left(\sum_{i=l_1+1}^{l_2}\gamma_i\right)^2+\frac{1}{l_{30}-l_2}\left(\sum_{i=l_2+1}^{l_{30}}\gamma_i\right)^2\right]. \tag{A.10}$$

Now, we divide the proof into two main cases: ***i.*** $l_2>l_{20}$ and ***ii.*** $l_2<l_{20}$.

i. If $l_2>l_{20}$, using (A.10), we have

$$\begin{aligned}
&N\Big(S_{1N}\big(\tilde{\boldsymbol{\ell}}(1)\big)-S_{1N}\big(\tilde{\boldsymbol{\ell}}(2)\big)\Big)\\
=&\frac{1}{l_{20}-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right)^2+(l_{30}-l_{20})\,\gamma_{g_3}^2-\frac{1}{l_2-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i+(l_2-l_{20})\,\gamma_{g_3}\right)^2-(l_{30}-l_2)\,\gamma_{g_3}^2\\
=&(l_2-l_{20})\left(\frac{1}{(l_{20}-l_1)(l_2-l_1)}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right)^2-\frac{2\gamma_{g_3}}{l_2-l_1}\sum_{i=l_1+1}^{l_{20}}\gamma_i+\frac{l_{20}-l_1}{l_2-l_1}\gamma_{g_3}^2\right)\\
=&\frac{l_2-l_{20}}{(l_{20}-l_1)(l_2-l_1)}\left(\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right)-(l_{20}-l_1)\gamma_{g_3}\right)^2.
\end{aligned}\tag{A.11}$$

If $l_1\geq l_{10}$, the right-hand side of (A.11) simplifies to $\frac{l_2-l_{20}}{(l_{20}-l_1)(l_2-l_1)}(\gamma_{g_3}-\gamma_{g_2})^2$. Similarly, if $l_1<l_{10}$, we have

$$\begin{aligned}
&\frac{N\Big(S_{1N}\big(\tilde{\boldsymbol{\ell}}(1)\big)-S_{1N}\big(\tilde{\boldsymbol{\ell}}(2)\big)\Big)}{l_2-l_{20}}\\
=&\frac{1}{(l_{20}-l_1)(l_2-l_1)}\Big((l_{10}-l_1)\gamma_{g_1}+(l_{20}-l_{10})\gamma_{g_2}-(l_{20}-l_1)\gamma_{g_3}\Big)^2\\
=&\frac{1}{(l_{20}-l_1)(l_2-l_1)}\Big((l_{10}-l_1)(\gamma_{g_2}-\gamma_{g_1})+(l_{20}-l_1)(\gamma_{g_3}-\gamma_{g_2})\Big)^2\geq\frac{l_{20}-l_1}{l_2-l_1}(\gamma_{g_3}-\gamma_{g_2})^2.
\end{aligned}$$

Therefore, for any $\epsilon\leq(\lambda_{20}-\lambda_{10})/4$ as constructed in (A.5), it always holds that

$$\begin{aligned}
\frac{N\Big(S_{1N}\big(\tilde{\boldsymbol{\ell}}(1)\big)-S_{1N}\big(\tilde{\boldsymbol{\ell}}(2)\big)\Big)}{l_2-l_{20}}&\geq\frac{l_{20}-l_1}{l_2-l_1}(\gamma_{g_3}-\gamma_{g_2})^2\\
&=\frac{l_{20}-l_{10}+l_{10}-l_1}{l_{20}-l_{10}+l_2-l_{20}+l_{10}-l_1}(\gamma_{g_3}-\gamma_{g_2})^2\\
&\geq\frac{\lambda_{20}-\lambda_{10}-\epsilon}{\lambda_{20}-\lambda_{10}+\epsilon}(\gamma_{g_3}-\gamma_{g_2})^2\geq\frac{3}{5}(\gamma_{g_3}-\gamma_{g_2})^2,
\end{aligned}$$

Clearly, (A.9) holds for the case $l_2>l_{20}$.

ii. If $l_2 < l_{20}$, by (A.10) again, we can write

$$
\begin{aligned}
&N\Big(S_{1N}\big(\tilde{\boldsymbol{\ell}}(1)\big) - S_{1N}\big(\tilde{\boldsymbol{\ell}}(2)\big)\Big) \qquad\qquad (A.12)\\
&= \frac{1}{l_{20}-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right)^2 + (l_{30}-l_{20})\,\gamma_{g_3}^2 - \frac{1}{l_2-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i - (l_{20}-l_2)\,\gamma_{g_2}\right)^2\\
&\quad - \frac{1}{l_{30}-l_2}\big[\,(l_{20}-l_2)\,\gamma_{g_2} + (l_{30}-l_{20})\gamma_{g_3}\big]^2\\
&= \frac{1}{l_{20}-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right)^2 - \frac{1}{l_2-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right)^2 + \frac{2\,(l_{20}-l_2)\,\gamma_{g_2}}{l_2-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right) - \frac{(l_{20}-l_2)^2}{l_2-l_1}\gamma_{g_2}^2\\
&\quad - (l_{20}-l_2)\,\gamma_{g_3}^2 + 2\,(l_{20}-l_2)\,\gamma_{g_3}(\gamma_{g_3}-\gamma_{g_2}) - \frac{(l_{20}-l_2)^2}{l_{30}-l_2}(\gamma_{g_3}-\gamma_{g_2})^2\\
&= \frac{l_{20}-l_2}{l_2-l_1}\left[-\frac{1}{l_{20}-l_1}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right)^2 + 2\gamma_{g_2}\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right) - (l_{20}-l_2)\gamma_{g_2}^2\right]\\
&\quad - (l_{20}-l_2)\left[\gamma_{g_3}^2 - 2\gamma_{g_3}(\gamma_{g_3}-\gamma_{g_2}) + (\gamma_{g_3}-\gamma_{g_2})^2 - \frac{l_{30}-l_{20}}{l_{30}-l_2}(\gamma_{g_3}-\gamma_{g_2})^2\right]\\
&= -\frac{l_{20}-l_2}{(l_2-l_1)(l_{20}-l_1)}\left[\left(\sum_{i=l_1+1}^{l_{20}}\gamma_i\right) - (l_{20}-l_1)\gamma_{g_2}\right]^2 + \frac{(l_{20}-l_2)\,(l_{30}-l_{20})}{l_{30}-l_2}(\gamma_{g_3}-\gamma_{g_2})^2. \qquad (A.13)
\end{aligned}
$$

If $l_1 \geq l_{10}$, (A.13) leads to

$$
\frac{N\Big(S_{1N}\big(\tilde{\boldsymbol{\ell}}(1)\big) - S_{1N}\big(\tilde{\boldsymbol{\ell}}(2)\big)\Big)}{l_{20}-l_2} = \frac{l_{30}-l_{20}}{l_{30}-l_2}(\gamma_{g_3}-\gamma_{g_2})^2 = \frac{l_{30}-l_{20}}{l_{30}-l_{20}+l_{20}-l_2}(\gamma_{g_3}-\gamma_{g_2})^2 \geq \frac{4}{5}(\gamma_{g_3}-\gamma_{g_2})^2,
$$

provided that $\epsilon \leq (\lambda_{30}-\lambda_{20})/4$. If $l_1 < l_{10}$, from (A.13), we have

$$
\begin{aligned}
\frac{N\Big(S_{1N}\big(\tilde{\boldsymbol{\ell}}(1)\big) - S_{1N}\big(\tilde{\boldsymbol{\ell}}(2)\big)\Big)}{l_{20}-l_2} &= -\frac{(l_{10}-l_1)^2}{(l_2-l_1)(l_{20}-l_1)}(\gamma_{g_2}-\gamma_{g_1})^2 + \frac{l_{30}-l_{20}}{l_{30}-l_2}(\gamma_{g_3}-\gamma_{g_2})^2\\
&\geq -\left(\frac{\epsilon}{\lambda_{20}-\lambda_{10}+\epsilon}\right)^2(\gamma_{g_2}-\gamma_{g_1})^2 + \frac{4}{5}(\gamma_{g_3}-\gamma_{g_2})^2.
\end{aligned}
$$

Note that $x \mapsto \dfrac{x}{\lambda_{20}-\lambda_{10}+x}$ is strictly increasing for $x > 0$. By the construction of ϵ in (A.5), we have $\epsilon \leq \dfrac{1}{4}(\lambda_{20}-\lambda_{10})\dfrac{\gamma_{g_3}-\gamma_{g_2}}{\gamma_{g_2}-\gamma_{g_1}}$, and therefore,

$$
\frac{N\Big(S_{1N}\big(\tilde{\boldsymbol{\ell}}(1)\big) - S_{1N}\big(\tilde{\boldsymbol{\ell}}(2)\big)\Big)}{l_{20}-l_2} \geq -\frac{1}{16}\left(\frac{\gamma_{g_3}-\gamma_{g_2}}{\gamma_{g_2}-\gamma_{g_1}}\right)^2(\gamma_{g_2}-\gamma_{g_1})^2 + \frac{4}{5}(\gamma_{g_3}-\gamma_{g_2})^2 \geq \frac{3}{5}(\gamma_{g_3}-\gamma_{g_2})^2.
$$

To sum up, we see that (A.9) holds for the case $l_2 < l_{20}$ as well.

It remains to establish (A.7). For any random sequence $\{X_i\}$, we adopt the convention that

$\sum_{i=i_2}^{i_1} X_i = -\sum_{i=i_1}^{i_2} X_i$ for $i_1 < i_2$. We can write

$$\begin{aligned}\frac{NS_{2N}(\boldsymbol{\ell})}{2} &= \sum_{j=1}^{G} \frac{1}{(l_j - l_{j-1})} \left(\sum_{i=l_{(j-1)0}+1}^{l_{j0}} \gamma_i \right) \left(\sum_{i=l_{(j-1)0}+1}^{l_{j0}} \hat{\varepsilon}_i \right) \\ &\quad + \sum_{j=1}^{G} \frac{1}{(l_j - l_{j-1})} \left(\sum_{i=l_{j-1}+1}^{l_{(j-1)0}} \gamma_i + \sum_{i=l_{j0}+1}^{l_j} \gamma_i \right) \left(\sum_{i=l_{j-1}+1}^{l_j} \hat{\varepsilon}_i \right) \\ &\quad + \sum_{j=1}^{G} \frac{1}{(l_j - l_{j-1})} \left(\sum_{i=l_{(j-1)0}+1}^{l_{j0}} \gamma_i \right) \left(\sum_{i=l_{j-1}+1}^{l_{(j-1)0}} \hat{\varepsilon}_i + \sum_{i=l_{j0}+1}^{l_j} \hat{\varepsilon}_i \right).\end{aligned}$$

Note that $\sum_{i=l_{(j-1)0}+1}^{l_{j0}} \gamma_i = \big(l_{j0} - l_{(j-1)0}\big)\gamma_{g_j}$ and $NS_{2N}(\boldsymbol{\ell}_0)/2 = \sum_{j=1}^{G} \gamma_{g_j} \sum_{i=l_{(j-1)0}+1}^{l_{j0}} \hat{\varepsilon}_i$. We have

$$\begin{aligned}\frac{N\,|S_{2N}(\boldsymbol{\ell}_0) - S_{2N}(\boldsymbol{\ell})|}{2\|\boldsymbol{\ell} - \boldsymbol{\ell}_0\|_1} &\le \|\boldsymbol{\ell} - \boldsymbol{\ell}_0\|_1^{-1} \Bigg\{ \sum_{j=1}^{G} \left| \gamma_{g_j} \left(\frac{l_{j0} - l_{(j-1)0}}{l_j - l_{j-1}} - 1 \right) \right| \left| \sum_{i=l_{(j-1)0}+1}^{l_{j0}} \hat{\varepsilon}_i \right| \\ &\quad + \gamma_G \sum_{j=1}^{G} \big(|l_{j-1} - l_{(j-1)0}| + |l_j - l_{j0}| \big) \left| \frac{1}{(l_j - l_{j-1})} \sum_{i=l_{j-1}+1}^{l_j} \hat{\varepsilon}_i \right| \\ &\quad + \sum_{j=1}^{G} \gamma_{g_j} \frac{l_{j0} - l_{(j-1)0}}{(l_j - l_{j-1})} \left(\left| \sum_{i=l_{j-1}+1}^{l_{(j-1)0}} \hat{\varepsilon}_i \right| + \left| \sum_{i=l_{j0}+1}^{l_j} \hat{\varepsilon}_i \right| \right) \Bigg\}.\end{aligned}$$

By applying (2.9), (A.5), and using the bounds $|l_j - l_{j0}| \le \|\boldsymbol{\ell} - \boldsymbol{\ell}_0\|_1$, and

$$\frac{l_{j0} - l_{(j-1)0}}{l_j - l_{j-1}} \le \frac{\lambda_{j0} - \lambda_{(j-1)0}}{\lambda_{j0} - \lambda_{(j-1)0} - 2\epsilon} \le 2, \qquad \forall j = 1, \ldots, m,$$

we establish (A.7) for $i = 2$. The result for $S_{3N}(\cdot)$ term follows from the same reasoning. The details are omitted. □

In the proof below, equations and lemmas provided in the online supplement are referenced using numbers that begin with the prefix "S".

Proof of Theorem 2. Because $\big\|\hat{\boldsymbol{\ell}} - \boldsymbol{\ell}_0\big\|_1 \xrightarrow{p} 0$, it is sufficient to establish the asymptotic normality for $\hat{\gamma}_{g_j}\big(\boldsymbol{\ell}_0, \tilde{\boldsymbol{k}}\big) = \frac{1}{l_{j0} - l_{(j-1)0}} \sum_{i=l_{(j-1)0}}^{l_{j0}} \hat{\gamma}_i^H$. We give the proof for $j = 1$. Write $r_n = \frac{1}{\sqrt{\sum_{i=1}^{l_{10}} \frac{1}{k_i}}}$. Applying the decomposition in (S.1.1), we have

$$\frac{l_{10}}{\sqrt{\sum_{i=1}^{l_{10}} \frac{1}{k_i}}} \left(\hat{\gamma}_{g_1} - \gamma_{g_1} \right) = r_n S_{1,l_{10}} + r_n S_{2,l_{10}},$$

where $S_{1,l}$ and $S_{2,l}$ are defined in (S.1.1). It is straightforward to see that $r_n S_{1,l_{10}} \xrightarrow{d} N(0, \gamma_{g_1}^2)$. It remains to prove $r_n S_{2,l_{10}} \to B_1$.

Note that (2.10) is equivalent to

$$\lim_{s\to\infty} \frac{\log U_i(sx) - \log U_i(s) - \gamma_i \log x}{A_i(s)} = \frac{x^{\rho_i} - 1}{\rho_i}.$$

Moreover, it implies that for any $\epsilon > 0$, there exists an s_i such that for any $s > s_i$ and $x > 1$,

$$\left| \frac{\log U_i(sx) - \log U_i(s) - \gamma_i \log x}{\tilde{A}_i(s)} - \frac{x^{\rho_i} - 1}{\rho_i} \right| \leq \epsilon x^{\rho_i + \epsilon}, \quad i = 1, \ldots, N$$

where $\tilde{A}_i(s)/A_i(s) \to 1$ as $s \to \infty$; (see, e.g., de Haan and Ferreira, 2006, p. 60). By Lemma S.1.1 and $T_i/\tilde{k}_i \to \infty$, $P\left(\bigcap_{i=1}^N \left\{Y_{i,(T_i-\tilde{k}_i)} > s_i\right\}\right) \to 1$. We have uniformly for all $i = 1, \ldots, N$ and $j = 0, \ldots, \tilde{k}_i - 1$,

$$\begin{aligned}
&\log\left(\frac{U_i(Y_{i,(T_i-j)})}{U_i(Y_{i,(T_i-\tilde{k}_i)})}\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{-\gamma_i}\right) \\
=&\tilde{A}_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)\frac{\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{\rho_i} - 1}{\rho_i} + o_P(1)\tilde{A}_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{\rho_i+\epsilon}. \qquad \text{(A.14)}
\end{aligned}$$

By de Haan and Ferreira (2006, Theorem 2.3.3), Assumption 4 implies that $A_i(sx)/A_i(s) \to x^{\rho_i}$ for $x > 0$. Further, applying Lemma S.1.1, we have uniformly for all $i = 1, \ldots, N$,

$$\frac{\tilde{A}_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}{A_i\left(\frac{T_i}{\tilde{k}_i}\right)} = \frac{\tilde{A}_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}{A_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}\frac{A_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}{A_i\left(\frac{T_i}{\tilde{k}_i}\right)} = 1 + o_P(1). \qquad \text{(A.15)}$$

Finally combining (A.14) and (A.15), we obtain

$$\begin{aligned}
r_n S_{2,l_{10}} =& r_n \sum_{i=1}^{l_{10}} \frac{1}{\tilde{k}_i} \sum_{j=0}^{\tilde{k}_i-1} \log\left(\frac{U_i(Y_{i,(T_i-j)})}{U_i(Y_{i,(T_i-\tilde{k}_i)})}\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{-\gamma_i}\right) \\
=& r_n \sum_{i=1}^{l_{10}} \tilde{A}_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)\frac{1}{\tilde{k}_i}\sum_{j=0}^{\tilde{k}_i-1}\frac{\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{\rho_i} - 1}{\rho_i} \\
&+ o_P(1) r_n \sum_{i=1}^{l_{10}} \tilde{A}_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)\frac{1}{\tilde{k}_i}\sum_{j=0}^{\tilde{k}_i-1}\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{\rho_i+\epsilon} \\
=& r_n(1 + o_P(1)) \sum_{i=1}^{l_{10}} A_i\left(\frac{T_i}{\tilde{k}_i}\right)\frac{1}{\tilde{k}_i}\sum_{j=0}^{\tilde{k}_i-1}\frac{\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{\rho_i} - 1}{\rho_i} + o_P(1) r_n \sum_{i=1}^{l_{10}} A_i\left(\frac{T_i}{\tilde{k}_i}\right),
\end{aligned}$$

where the second term is $o_P(1)$, because $r_n \sum_{i=1}^{l_{10}} A_i\left(\frac{T_i}{\tilde{k}_i}\right)$ is bounded. For the first term, using (S.1.2), we obtain

$$\sum_{j=0}^{\tilde{k}_i-1} \frac{\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{\rho_i}-1}{\rho_i} \stackrel{d}{=} \sum_{j=1}^{\tilde{k}_i} \frac{e^{\rho_i E_{i,j}}-1}{\rho_i}.$$

Observe that $\mathbb{E}\left(\frac{e^{\rho_i E_{i,j}}-1}{\rho_i}\right) = \frac{1}{1-\rho_i}$. Thus,

$$P\left(\left| r_n \sum_{i=1}^{l_{10}} A_i\left(\frac{T_i}{\tilde{k}_i}\right) \frac{1}{\tilde{k}_i} \sum_{j=0}^{\tilde{k}_i-1} \frac{\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-\tilde{k}_i)}}\right)^{\rho_i}-1}{\rho_i} - r_n \sum_{i=1}^{l_{10}} \frac{A_i\left(\frac{T_i}{\tilde{k}_i}\right)}{1-\rho_i} \right| \geq \epsilon \right)$$

$$\leq \frac{1}{\epsilon^2} \operatorname{Var}\left(r_n \sum_{i=1}^{l_{10}} A_i\left(\frac{T_i}{\tilde{k}_i}\right) \frac{1}{\tilde{k}_i} \sum_{j=1}^{\tilde{k}_i} \frac{e^{E_{i,j}\rho_i}-1}{\rho_i} \right) = \frac{C}{\epsilon^2} \frac{\sum_{i=1}^{l_{10}} A_i^2\left(\frac{T_i}{\tilde{k}_i}\right)\frac{1}{\tilde{k}_i}}{\sum_{i=1}^{l_{10}} \frac{1}{\tilde{k}_i}} \to 0,$$

because $\max_{1\leq i\leq N}\left|A_i\left(\frac{T_i}{\tilde{k}_i}\right)\right| \to 0$. Now Theorem 2 follows from $r_n \sum_{i=1}^{l_{10}} \frac{A_i\left(\frac{T_i}{\tilde{k}_i}\right)}{1-\rho_i} \to B_1$ as required in Assumption 5. □

Online Appendix to:

Clustering extreme value indices in large panels

Chenhui Wang[1], *Juan Juan Cai*[1,2], *Yicong Lin*[1,2], *Julia Schaumburg*[1,2]

[1] *Vrije Universiteit Amsterdam, the Netherlands*

[2] *Tinbergen Institute, the Netherlands*

Contents

S.1 Proofs of auxiliary lemmas

In the subsequent proofs, we use $C, C_1, C_2, \ldots$ to denote constants whose values vary depending on the context.

Proof of Lemma 1. We will prove the case for $l_1 = 1$ and $l_2 = N$. The proof follows in the same manner for other values.

Let $\{Y_{i,t} : i = 1, \ldots, N, t = 1, \ldots, T_i\}$ be i.i.d. with a common distribution given by $x \mapsto 1 - 1/x$, $x \geq 1$. Then we have $\{Z_{i,t} : t = 1, \ldots, T_i,\ i = 1, \ldots, N\} =^d \{U_i(Y_{i,t}) : t = 1, \ldots, T_i,\ i = 1, \ldots, N\}$. For any $1 \leq l \leq N$, the partial sum in (2.9) can be written as

$$
\begin{aligned}
\sum_{i=1}^{l} \hat{\varepsilon}_i &= \sum_{i=1}^{l} \left(\hat{\gamma}_i^H - \gamma_i\right) \\
&= \sum_{i=1}^{l} \left(\frac{1}{k_i} \sum_{j=0}^{k_i-1} \log Z_{i,(T_i-j)} - \log Z_{i,(T_i-k_i)} - \gamma_i \right) \\
&=^d \sum_{i=1}^{l} \left(\frac{1}{k_i} \sum_{j=0}^{k_i-1} \log \frac{U_i(Y_{i,(T_i-j)})}{U_i(Y_{i,(T_i-k_i)})} - \gamma_i \right) \\
&= \sum_{i=1}^{l} \left(\frac{1}{k_i} \sum_{j=0}^{k_i-1} \left(\gamma_i \log \frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}} + \log \left(\frac{U_i(Y_{i,(T_i-j)})}{U_i(Y_{i,(T_i-k_i)})} \left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}} \right)^{-\gamma_i} \right) \right) - \gamma_i \right) \\
&= \sum_{i=1}^{l} \gamma_i \left(\frac{1}{k_i} \sum_{j=0}^{k_i-1} \log \frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}} - 1 \right) + \sum_{i=1}^{l} \frac{1}{k_i} \sum_{j=0}^{k_i-1} \log \left(\frac{U_i(Y_{i,(T_i-j)})}{U_i(Y_{i,(T_i-k_i)})} \left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}} \right)^{-\gamma_i} \right) \\
&=: S_{1,l} + S_{2,l}. \qquad \text{(S.1.1)}
\end{aligned}
$$

We shall show that $\sup_{1 \leq l \leq N} (Nl)^{-1/2} |S_{1,l}| = o_P(1)$ and $\sup_{1 \leq l \leq N} (Nl)^{-1/2} |S_{2,l}| = o_P(1)$. Let's deal with $S_{1,l}$ and $S_{2,l}$ separately. Denote $X_{1,i} = k_i^{-1} \sum_{j=0}^{k_i-1} \log \frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}} - 1$. Let $\{E_{i,j}, i = 1, \ldots, N, j = 1, \ldots, k_i\}$ be i.i.d standard exponential r.v.s. And $E_{i,(j)}$ denotes the j-th order statistics in row i, that is $E_{i,(1)} \leq \cdots \leq E_{i,(k_i)}$. By applying the Rényi's representation for the exponential order statistics (Rényi, 1953), one has

$$
\left\{ \log \frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}} \right\}_{j=0}^{k_i-1} =^d \left\{ E_{i,(k_i-j+1)} \right\}_{j=1}^{k_i}, \qquad \text{(S.1.2)}
$$

where the left-hand side does not depend on T_i. Thus, $\mathbb{E}(X_{1,i}) = \mathbb{E}\left(k_i^{-1} \sum_{j=1}^{k_i} E_{i,j} - 1\right) = 0$ and $\mathbb{E}\left(X_{1,i}\right)^2 = \mathrm{Var}\left(k_i^{-1} \sum_{j=1}^{k_i} E_{i,j} - 1\right) = k_i^{-1}$. For any $\epsilon > 0$, by Hájek and Rényi (1955)'s

inequality for independent r.v.s, as $N \to \infty$,

$$P\left(\sup_{1\le l\le N}\frac{1}{l}|S_{1,l}|\ge\epsilon\right)\le C\frac{1}{\epsilon^2}\sum_{i=1}^{N}\frac{1}{i^2}\frac{1}{k_i}=O\left(\max_{1\le i\le N}\frac{1}{k_i}\right)=o(1), \tag{S.1.3}$$

where $\max_{1\le i\le N}k_i^{-1}=o(1)$ is implied by $N\max_{1\le i\le N}e^{-k_i}=o(1)$ in Assumption 2.

We move on to $S_{2,l}$. For convenience, define

$$W_{i,j}:=\frac{U_i(Y_{i,(T_i-j)})}{U_i(Y_{i,(T_i-k_i)})\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}\right)^{\gamma_i}}-1.$$

We apply two inequalities for regularly varying functions, namely, Proposition B.1.9 (5) (Potter inequality) and Proposition B.1.10 in de Haan and Ferreira (2006) to obtain the lower and upper bounds of $W_{i,j}$. For any $\delta\in(0,1/4]$, there exists $t_{i,0}$ such that for any $t_i\ge t_{i,0}$ and $x>1$,

$$\max\left(\left((1-\delta)\,x^{-\delta}-1\right),-\delta x^{\delta}\right)\le\frac{U_i(t_ix)}{U_i(t_i)x^{\gamma_i}}-1\le\delta x^{\delta},\qquad i=1,\dots,N. \tag{S.1.4}$$

By Lemma S.1.1 and $T_i/k_i\to\infty$, $P\left(\bigcap_{i=1}^{N}\left\{Y_{i,(T_i-k_i)}>t_{i,0}\right\}\right)\to 1$. Thus, by replacing $t_i=Y_{i,(T_i-k_i)}$ and $x=\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}$ in (S.1.4), we have with probability tending to one,

$$\max\left(\left((1-\delta)\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}\right)^{-\delta}-1\right),-\delta\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}\right)^{\delta}\right)\le W_{i,j}\le\delta\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}\right)^{\delta},$$

for all $i=1,\dots,N$, $j=0,1,\dots,k_i-1$. Since $W_{i,j}>-1$ always holds, it follows almost surely that

$$|\log(1+W_{i,j})|\le\frac{|W_{i,j}|}{\sqrt{1+W_{i,j}}}\le\frac{\delta\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}\right)^{\delta}}{\sqrt{(1-\delta)\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}\right)^{-\delta}}}=\frac{\delta}{\sqrt{1-\delta}}\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}\right)^{3\delta/2}.$$

It leads to

$$|S_{2,l}|\le\sum_{i=1}^{l}\frac{1}{k_i}\sum_{j=0}^{k_i-1}|\log(1+W_{i,j})|\le\sum_{i=1}^{l}\frac{1}{k_i}\sum_{j=0}^{k_i-1}\frac{\delta}{\sqrt{1-\delta}}\left(\frac{Y_{i,(T_i-j)}}{Y_{i,(T_i-k_i)}}\right)^{3\delta/2}=:\sum_{i=1}^{l}X_{2,i}.$$

Applying (S.1.2), we obtain $\mathbb{E}(X_{2,i})=\frac{\delta}{\sqrt{1-\delta}}\mathbb{E}\left(\exp\left(\frac{3\delta}{2}E_{1,1}\right)\right)=\frac{\delta}{\sqrt{1-\delta}\,(1-\frac{3\delta}{2})}\le C_1\delta$ and

$$\mathrm{Var}\,(X_{2,i})=\frac{\delta^2}{k_i(1-\delta)}\mathrm{Var}\left(\exp\left(\frac{3\delta}{2}E_{1,1}\right)\right)=\frac{\delta^2}{k_i(1-\delta)}\left(\frac{1}{1-3\delta}-\frac{1}{(1-\frac{3\delta}{2})^2}\right)\le\frac{C_2}{k_i}\delta.$$

Note that the constants C_1 and C_2 above are independent of δ for any $\delta \in (0, 1/4]$. For any $\epsilon > 0$, choose $\delta \in (0, 1/4]$ such that $\epsilon - C_1\delta > 0$. Then, we have

$$\begin{aligned}
&P\left(\sup_{1\le l\le N}\frac{1}{l}\sum_{i=1}^{l}X_{2,i}\ge\epsilon\right)\\
\le&P\left(\sup_{1\le l\le N}\frac{1}{l}\sum_{i=1}^{l}(X_{2,i}-\mathbb{E}(X_{2,i}))+\sup_{1\le l\le N}\mathbb{E}(X_{2,i})\ge\epsilon\right)\\
\le&P\left(\sup_{1\le l\le N}\frac{1}{l}\sum_{i=1}^{l}(X_{2,i}-\mathbb{E}(X_{2,i}))\ge\epsilon-C_1\delta\right)\\
\le&\frac{1}{(\epsilon-C_1\delta)^2}\sum_{i=1}^{N}\frac{1}{i^2}\frac{C_2\delta}{k_i}=o(1),
\end{aligned}\tag{S.1.5}$$

where we apply again Hájek-Rényi inequality to obtain the second-to-last inequality. Thus (2.9) is proved by combining (S.1.3) and (S.1.5). □

Lemma S.1.1. *Let $\{Y_{i,t} : i = 1, \dots, N, t = 1, \dots, T_i\}$ be i.i.d. with a common distribution given by $x \mapsto 1 - 1/x$, $x \ge 1$.*

(a) If the sequence $\{k_i,\ i = 1, \dots, N\}$ fulfills Assumption 2, then

$$P\left(\bigcap_{i=1}^{N}\left\{Y_{i,(T_i-k_i)}\ge\sqrt{\frac{T_i}{k_i}}\right\}\right)\to 1.\tag{S.1.6}$$

(b) If the sequence $\{\tilde{k}_i,\ i = 1, \dots, N\}$ fulfills Assumption 5, then

$$\sup_{1\le i\le N}\left|\frac{A_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}{A_i\left(\frac{T_i}{\tilde{k}_i}\right)}-1\right|=o_P(1),\tag{S.1.7}$$

where A_i is defined in (2.10).

Proof of Lemma S.1.1. We first prove (S.1.6). For each i, by Hoeffding's inequality,

$$\begin{aligned}
P\left(Y_{i,(T_i-k_i)}\ge\sqrt{\frac{T_i}{k_i}}\right)&=P\left(\sum_{t=1}^{T_i}\mathbb{1}\left\{Y_{i,t}\ge\sqrt{\frac{T_i}{k_i}}\right\}>k_i\right)\\
&=1-P\left(\sum_{t=1}^{T_i}\left(\mathbb{1}\left\{Y_{i,t}\ge\sqrt{\frac{T_i}{k_i}}\right\}-\frac{k_i}{\sqrt{T_i}}\right)\le k_i-\sqrt{k_iT_i}\right)\\
&\ge 1-2\exp\left(-\frac{2}{T_i}\left(k_i-\sqrt{k_iT_i}\right)^2\right)\ge 1-2\exp(-k_i).
\end{aligned}$$

Using independence, we have

$$P\left(\bigcap_{i=1}^{N}\left\{Y_{i,(T_i-k_i)} \geq \frac{\sqrt{T_i}}{k_i}\right\}\right) = \prod_{i=1}^{N} P\left(Y_{i,(T_i-k_i)} \geq \frac{\sqrt{T_i}}{k_i}\right) \geq \prod_{i=1}^{N}\left(1-2\exp(-k_i)\right).$$

It suffices to prove that $\log\left(\prod_{i=1}^{N}\left(1-2\exp(-k_i)\right)\right) \to 0$, which follows from that

$$\sum_{i=1}^{N}\left|\log\left(1-2\exp(-k_i)\right)\right| \leq \sum_{i=1}^{N}\frac{2\exp(-k_i)}{1-2\exp(-k_i)} \leq 4\sum_{i=1}^{N}\exp(-k_i),$$

and the assumption that $N\max_{1\leq i\leq N} e^{-k_i} \to 0$ as $N\to\infty$.

Next, we prove (S.1.7). By de Haan and Ferreira (2006, Theorem 2.3.3), Assumption 4 implies that $A_i(sx)/A_i(s) \to x^{\rho_i}$ for $x>0$. If $\rho_i = -\infty$, which implies that F_i is the standard Pareto distribution, then (S.1.7) trivially holds. Thus, in the following, we assume that $|\rho_i| < \infty$ for all i. Applying Proposition B.1.10 in de Haan and Ferreira (2006), for any $\epsilon_1 > 0$, there exists $t_{i,0}$ such that for any $t_i \geq t_{i,0}$ and $x>0$,

$$\left|\frac{A_i(t_i x)}{A_i(t_i)} - x^{\rho_i}\right| \leq \epsilon_1 x^{\rho_i}\max\left(x^{\epsilon_1}, x^{-\epsilon_1}\right), \qquad i=1,\ldots,N.$$

Denote $I_N = \bigcap_{i=1}^{N}\left\{Y_{i,(T_i-\tilde{k}_i)} > t_{i,0}\right\}$. By the law of total probability, one has

$$\begin{aligned}
&P\left(\sup_{1\leq i\leq N}\left|\frac{A_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}{A_i\left(\frac{T_i}{\tilde{k}_i}\right)} - 1\right| \geq \epsilon\right)\\
\leq &P\left(\sup_{1\leq i\leq N}\left|\frac{A_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}{A_i\left(\frac{T_i}{\tilde{k}_i}\right)} - 1\right| \geq \epsilon \,\middle|\, I_N\right)P(I_N) + P(I_N^{\complement}).
\end{aligned}$$

Note that $P(I_N)\to 1$ due to (S.1.6). It suffices to show that conditioning on I_N,

$$\sup_{1\leq i\leq N}\left|\frac{A_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}{A_i\left(\frac{T_i}{\tilde{k}_i}\right)} - 1\right| = o_P(1).$$

Conditioning on I_N, we have for $i=1,\ldots,N$,

$$\left|\frac{A_i\left(Y_{i,(T_i-\tilde{k}_i)}\right)}{A_i\left(\frac{T_i}{\tilde{k}_i}\right)} - \left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)^{\rho_i}\right| \leq \epsilon_1\left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)^{\rho_i}\max\left(\left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)^{\epsilon_1}, \left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)^{-\epsilon_1}\right).$$

Then we only need to prove

$$\sup_{1\le i\le N}\left|\left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)^{\rho_i}-1\right|=o_P(1)\qquad\text{and}\qquad\sup_{1\le i\le N}\left|\left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)-1\right|=o_P(1).$$

We only prove the first statement, as the second follows with a more straightforward argument. Note that

$$\begin{aligned}
&P\left(\sup_{1\le i\le N}\left|\left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)^{\rho_i}-1\right|\ge\epsilon\right)\\
\le&\sum_{i=1}^{N}P\left(\left|\left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)^{\rho_i}-1\right|\ge\epsilon\right)\\
=&\sum_{i=1}^{N}P\left(\sum_{t=1}^{T_i}\mathbb{1}\left\{Y_{i,t}\ge\frac{T_i}{\tilde{k}_i}(1-\epsilon)^{1/\rho_i}\right\}\ge\tilde{k}_i+1\right)+\sum_{i=1}^{N}P\left(\sum_{t=1}^{T_i}\mathbb{1}\left\{Y_{i,t}\ge\frac{T_i}{\tilde{k}_i}(1+\epsilon)^{1/\rho_i}\right\}\le\tilde{k}_i+1\right)\\
=&\sum_{i=1}^{N}P\left(\widetilde{B}_{i,1}\ge\tilde{k}_i+1\right)+\sum_{i=1}^{N}P\left(\widetilde{B}_{i,2}\le\tilde{k}_i+1\right),
\end{aligned}\tag{S.1.8}$$

where $\widetilde{B}_{i,1}\sim\mathrm{Bin}\left(T_i,\frac{\tilde{k}_i}{T_i}(1-\epsilon)^{-1/\rho_i}\right)$ and $\widetilde{B}_{i,2}\sim\mathrm{Bin}\left(T_i,\frac{\tilde{k}_i}{T_i}(1+\epsilon)^{-1/\rho_i}\right)$. We shall apply the Chernoff bound for binomial random variables; see, e.g., Mohri et al. (2018, Appendix D). Specifically, suppose $\mathcal{B}\sim\mathrm{Bin}(n,p)$ with $n\ge0$ and $p\in(0,1)$, then, one has

$$\mathbb{P}\Big(\mathcal{B}\le(1-\epsilon)np\Big)\le\exp\left(-\frac{\epsilon^2}{2}np\right)\qquad\text{and}\qquad\mathbb{P}\Big(\mathcal{B}\ge(1+\epsilon)np\Big)\le\exp\left(-\frac{\epsilon^2}{3}np\right).$$

Thus,

$$\begin{aligned}
P\left(\widetilde{B}_{i,1}\ge\tilde{k}_i+1\right)&\le\exp\left(-\frac{\tilde{k}_i}{3}(1-\epsilon)^{-1/\rho_i}\Big((1+\tilde{k}_i^{-1})(1-\epsilon)^{1/\rho_i}-1\Big)^2\right)\\
&<\exp\left(-\frac{\tilde{k}_i}{3}(1-\epsilon)^{-1/\rho_i}\Big((1-\epsilon)^{1/\rho_i}-1\Big)^2\right)\\
&=\exp\left(-\frac{\tilde{k}_i}{3}\Big(\big[(1-\epsilon)^{1/\rho_i}-1\big]+\big[(1-\epsilon)^{-1/\rho_i}-1\big]\Big)\right)\\
&\le\exp\left(-\frac{\tilde{k}_i\epsilon^2}{3\rho_i^2}\right),
\end{aligned}\tag{S.1.9}$$

where the last inequality follows from the Taylor expansion for $(1-x)^{1/\rho_i}$ and $(1-x)^{-1/\rho_i}$. Similarly, we can obtain,

$$P\left(\widetilde{B}_{i,2}\le\tilde{k}_i+1\right)\le\exp\left(-\frac{\tilde{k}_i\epsilon^2}{2\rho_i^2}+\frac{3}{2}\right).\tag{S.1.10}$$

Substituting (S.1.9) and (S.1.10) into (S.1.8), we obtain

$$P\left(\sup_{1\le i\le N}\left|\left(\frac{\tilde{k}_i}{T_i}Y_{i,(T_i-\tilde{k}_i)}\right)^{\rho_i}-1\right|\ge\epsilon\right)\le 6N\max_{1\le i\le N}\exp\left(-\frac{\tilde{k}_i\epsilon^2}{3\rho_i^2}\right)$$
$$\le 6N\max_{1\le i\le N}\exp\left(-\tilde{k}_i/\log(\tilde{k}_i)\right)\to 0.$$

This completes the proof. □

S.2 Additional simulation results

We collect additional simulation outputs in this section.

S.2.1 Additional outputs of group estimator

Table S.2.1: Empirical MAE (in percent) of group and individual Hill estimators for $T = 3000$.

		$\widetilde{R}_k$			R_H				$\widetilde{R}_k$			R_H	
(G, S_G)	Group	2%	3%	4%	5%	7%	(G, S_G)	Group	2%	3%	4%	5%	7%
	1	3.58	4.49	5.32	6.54	5.71		1	3.58	4.49	5.32	6.54	5.71
$(3, 100)$	2	0.98	1.03	1.27	10.03	7.90	$(3, 300)$	2	0.69	0.90	1.24	10.03	7.90
	3	1.69	1.60	1.74	5.13	8.85		3	1.09	1.25	1.58	5.13	8.85
	1	2.53	3.16	3.75	6.54	7.90		1	2.90	3.64	4.33	6.54	7.90
	2	2.89	2.84	2.94	4.14	4.51		2	1.92	2.14	2.43	4.14	4.51
$(5, 100)$	3	0.98	1.02	1.25	5.72	5.14	$(5, 300)$	3	0.74	0.97	1.32	5.72	5.14
	4	1.59	1.64	1.86	7.86	6.94		4	1.27	1.52	1.85	7.86	6.94
	5	1.85	1.88	2.11	10.04	8.87		5	1.40	1.68	2.07	10.04	8.85

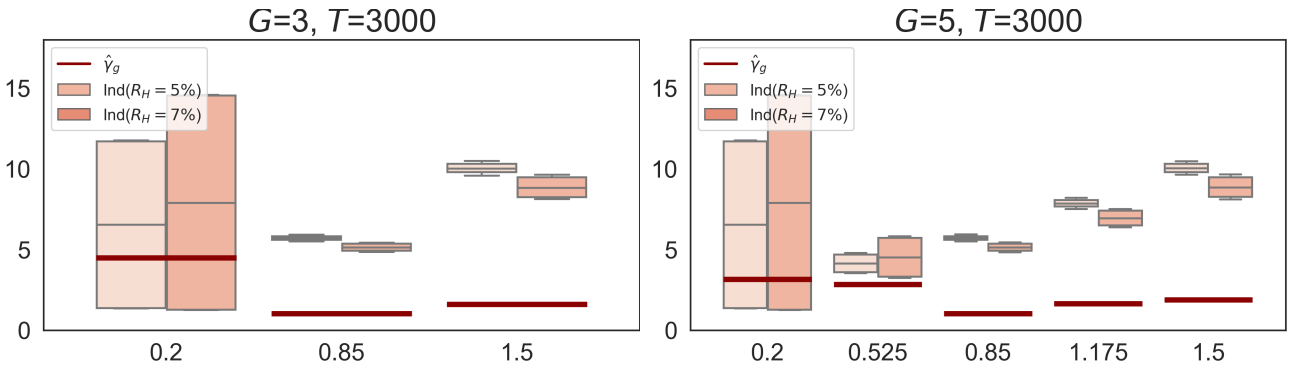


Figure S.2.1: Boxplots of group-wise MAEs (in percent) for the group Hill estimates constructed in (2.8) (solid red lines) and individual Hill estimates (2.3) The left figure corresponds to $G = 3$ groups, while the right one represents $G = 5$ groups. Here, $(T, S_G, \tilde{R}_k) = (3000, 100, 3\%)$.

S.2.2 Silhouette metric for selecting the number of segments

Another commonly used method for determining the number of clusters is the Silhouette index framework from the machine learning literature, originally introduced by Rousseeuw (1987) The Silhouette metric provides a comprehensive approach to assessing cluster validity. It takes into account within-cluster cohesion and its dissimilarity with other clusters. The Silhouette value for an observation i is:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

where $a(i) = \frac{1}{|C_I|-1} \sum_{j \in C_I, i \neq j} d(i,j)$ is the average dissimilarity of i to all other points within the same cluster C_I, and $b(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{j \in C_J} d(i,j)$ is the lowest average dissimilarity of i to points in any other cluster. The Silhouette value $S(i)$ ranges from -1 to 1, where a value close to 1 indicates that the observation is well clustered, a value near 0 suggests that the observation lies on the boundary between two clusters, and a negative value implies that the observation may have been misclassified.

The average Silhouette value for all observations provides an overall measure of clustering quality, $\bar{S} = N^{-1} \sum_{i=1}^{N} S(i)$, where N is the total number of observations. A higher average Silhouette value indicates a more appropriate clustering solution.

Table S.2.2: Empirical accuracy (in percent) of the elbow method and Silhouette metric for determining G.

			γ_S				
G	T	S_G	0.4	0.45	0.5	0.55	0.6
3	1000	100	99.12	100	100	100	100
		300	99.98	100	100	100	100
	3000	100	15.57	100	100	100	100
		300	2.14	100	100	100	100
5	1000	100	70.13	87.05	84.75	82.15	50.61
		300	89.55	98.50	97.90	97.60	58.99
	3000	100	0.34	100	100	100	93.22
		300	0	100	100	100	99.36

For applying the Silhouette method, we suggest a modification to the choice of k_i after the grouping process. When the estimated EVIs fall below a threshold $\gamma_S \in \{0.45, 0.5, 0.55\}$, a small threshold k_i of 1% of observations is chosen. The intuition behind this approach is to reduce bias for smaller EVI estimates, given their already small asymptotic variance. We assess

its performance through $M = 2000$ simulations, the results can be found in Table S.2.2. The modification performs well across different settings around $\gamma_S = 0.45$.

S.2.3 Empirical rejection rates of two-sample t-tests

Rejection probability is plotted against $\Delta\gamma$, representing the change in EVI within a given group while keeping others fixed. The DGP remains the same as in Section 3. We provide the results for $k = [12\%\, T]$, $S_G = 100$, $\tilde{k} = [3\%\, T]$. All three plots show empirical rejection rates that are close to 5% when group EVI remains unchanged ($\Delta\gamma = 0$), indicating that the test is correctly sized. As $|\Delta\gamma|$ increases, the empirical rejection probability (i.e., empirical power) rises. Group 1 shows the highest sensitivity to changes in $\Delta\gamma$, while group 3 exhibits the lowest. This suggests that small changes in $\Delta\gamma$ have a diminished impact on the empirical rejection probability for group 3, which could be indicative of progressively heavier tails as we move from group 1 to group 3.

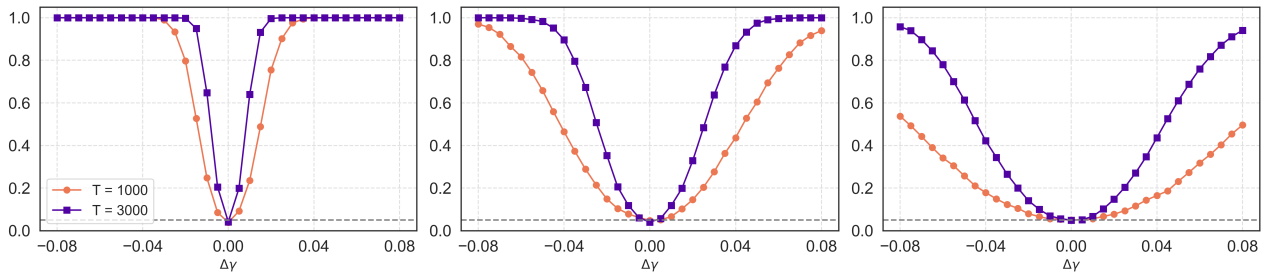


Figure S.2.2: Empirical rejection rates of two-sample t-tests for Group 1-3 (from left to right), with significance level 5%(dashed line).

S.3 Additional empirical results

Table S.3.3: Transition matrix for $G = 4$ between period 1 (rows) and period 2 (columns).

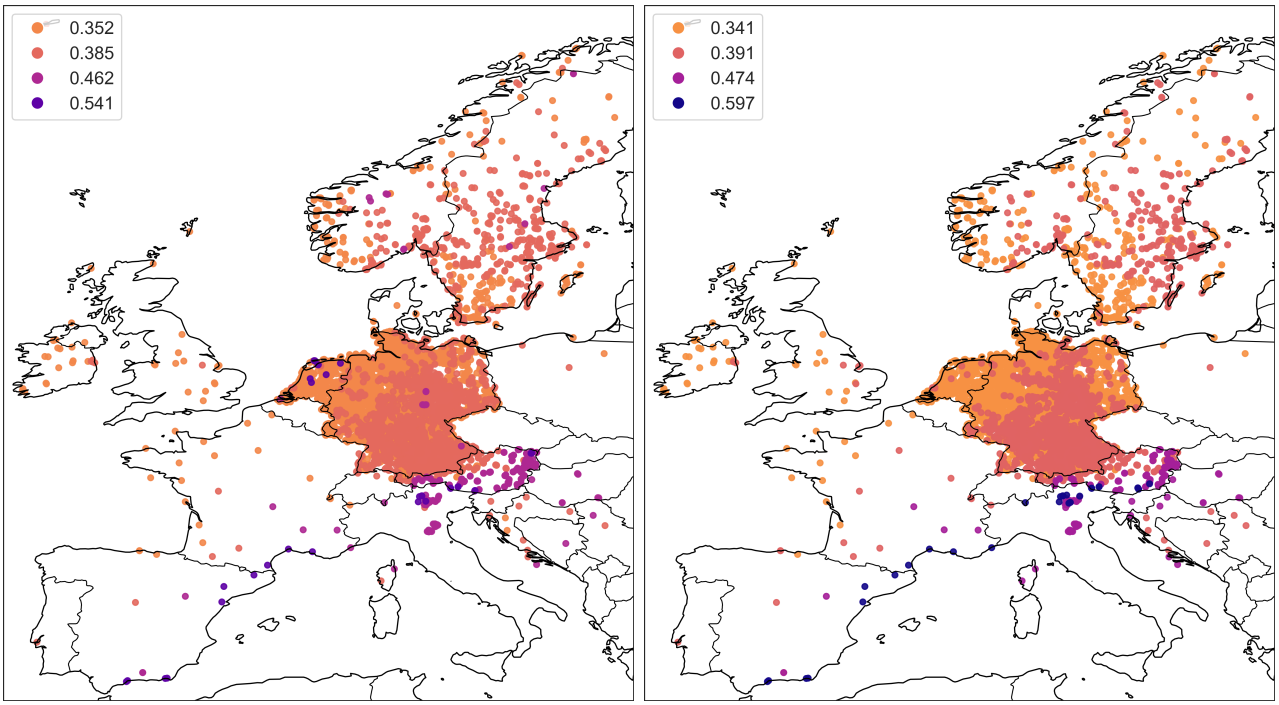
period 1 \ period 2	1	2	3	4
1	1998	516	0	0
2	589	1330	8	0
3	1	83	161	12
4	6	5	5	20

Table S.3.4: Transition matrix for $G = 5$ between period 1 (rows) and period 2 (columns).

period 1 \ period 2	1	2	3	4	5
1	1329	300	0	0	0
2	783	1202	14	0	0
3	95	655	129	6	1
4	0	8	50	124	7
5	6	3	2	1	19

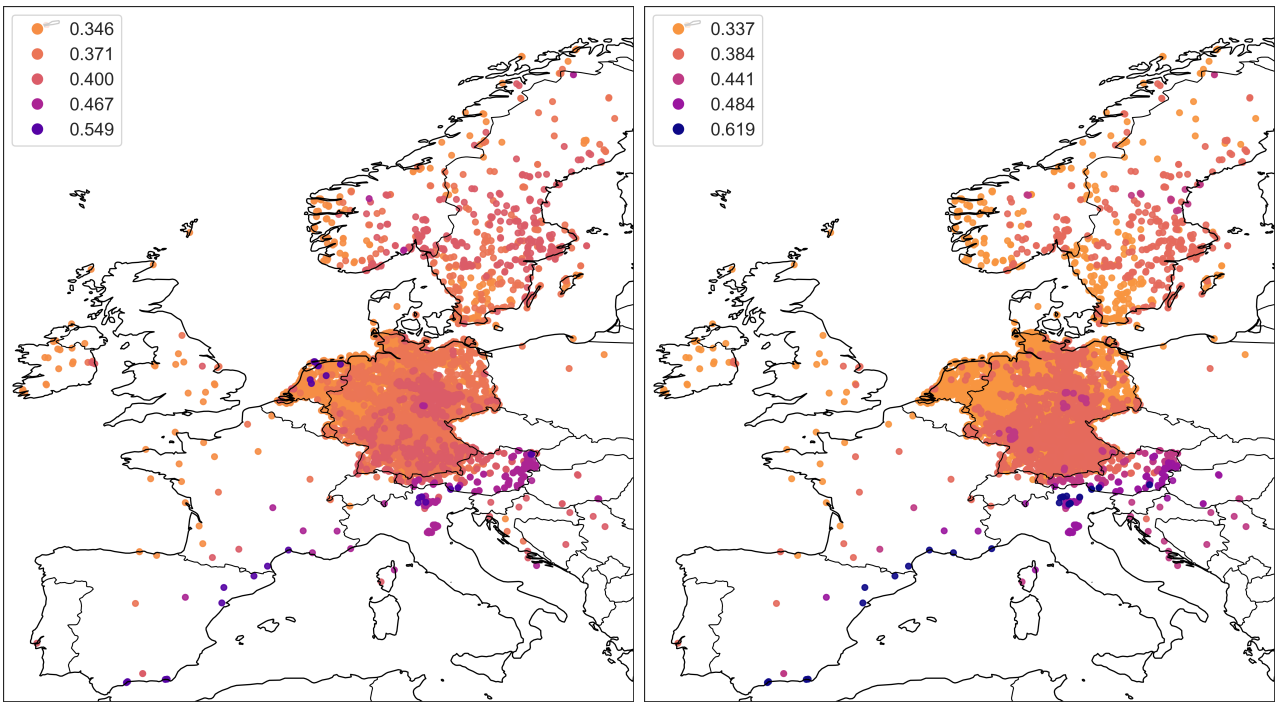
Table S.3.5: Test statistics with significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

	group 1	group 2	group 3	group 4	group 5
$G = 4$	10.451^{***}	5.050^{***}	2.419^{**}	3.943^{***}	–
$G = 5$	7.728^{***}	10.827^{***}	11.821^{***}	3.071^{**}	4.425^{***}



(a) Period 1 (1950–1985) **(b)** Period 2 (1986–2020)

Figure S.3.3: Maps showing the weather stations in our European sample for both periods. Dots depict grouped estimates ($G = 4$) for the extreme value index of precipitation volume. Darker shades refer to higher values.



(a) Period 1 (1950–1985) **(b)** Period 2 (1986–2020)

Figure S.3.4: Maps depicting the weather stations in our sample for both periods. Dots depict grouped estimates ($G = 5$) for the extreme value index of precipitation volume. Darker shades refer to higher values.

References in Appendix

Hájek, J. and A. Rényi (1955). Generalization of an inequality of Kolmogorov. *Acta Mathematica Hungarica 6*(3-4), 281–283.

Mohri, M., A. Rostamizadeh, and A. Talwalkar (2018). *Foundations of Machine Learning* (Second ed.). Cambridge: The MIT Press.

Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung 4*(2), 48–89.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics 20*, 53–65.