A test for comparing tail indices for heavy-tailed distributions via empirical likelihood

Julien Worms, Rym Worms

To cite this version:

HAL Id: hal-00733240
https://hal.archives-ouvertes.fr/hal-00733240
Submitted on 18 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A test for comparing tail indices for heavy-tailed distributions via empirical likelihood

Julien Worms (1) & Rym Worms (2)

(1) Université de Versailles-Saint-Quentin-En-Yvelines
Laboratoire de Mathématiques de Versailles (CNRS UMR 8100),
UFR de Sciences, Bât. Fermat,
45 av. des Etats-Unis, 78035 Versailles Cedex,
e-mail: julien.worms@uvsq.fr

(2) Université Paris-Est-Créteil
Laboratoire d’Analyse et de Mathématiques Appliquées (CNRS UMR 8050),
61 av. du Général de Gaulle, 94010 Créteil cedex,
e-mail: rym.worms@u-pec.fr

AMS Classification. Primary 62G32; Secondary 62G10.

Keywords and phrases. Extreme values, Tail index, Comparison Test, Hill Estimator, Empirical Likelihood.

Abstract

In this work, the problem of testing whether different (≥ 2) independent samples, with (possibly) different heavy tailed distributions, share the same extreme value index, is addressed. The test statistic proposed is inspired by the empirical likelihood methodology and consists in an ANOVA-like confrontation of Hill estimators. Asymptotic validity of this simple procedure is proved and efficiency, in terms of empirical type I error and power, is investigated through simulations under a variety of situations. Surprisingly, this topic had hardly been addressed before, and only in the two sample case, though it can prove useful in applications.

1. Introduction

In the topic of univariate extreme value analysis, the interest lies generally in the study of a single sample and inference in the tail of its underlying distribution. The main parameter describing the tail behavior of a continuous distribution function $F$ is its extreme value index. In many fields of applications, distributions of interest are those exhibiting a heavy tail phenomenon. In this case the extreme value index appears as the positive number $\gamma$ (1/$\gamma$ is then called the tail index) such that the survival function $1 - F$ is regularly varying with order $-1/\gamma$, which means that $F(x)$ grossly behaves like $x^{-1/\gamma}$ for large $x$ (precise definition of regular variation is given below), i.e. tail decreases to 0 at a polynomial rate. Therefore, the greater the value of $\gamma$ is, the greater is the chance that samples drawn from $F$ exhibits extreme values.
Most of the literature on univariate extreme value statistics focuses on methods for estimating as accurately as possible the extreme value index, which appears as the natural measure of heaviness of the tail of the underlying distribution. This work is devoted to a different topic, namely testing whether different distributions could share the same value of the tail index, in the heavy tail framework, on the basis of the observation of independent samples of these distributions. More precisely, we suppose that we observe $K$ independent i.i.d. samples $(X_1^{(1)})_{i \leq n_1}, \ldots, (X_K^{(K)})_{i \leq n_K}$, respectively coming from heavy-tailed distribution functions $F_1, \ldots, F_K$, i.e. such that for each $1 \leq j \leq K$, $F_j$ satisfies

$$\frac{1 - F_j(tx)}{1 - F_j(t)} \to x^{-1/\gamma_{0j}} \text{, as } t \to +\infty,$$

for all $x > 0$, where the parameter $\gamma_{0j} > 0$ is the extreme value index of the $j$-th sample.

The aim of this paper is to provide a simple yet effective test procedure for comparison of $\gamma_{01}, \ldots, \gamma_{0K}$, for instance by testing the equality hypothesis

$$H_0 : \gamma_{01} = \ldots = \gamma_{0K}$$

This problem has not been so much addressed in the literature: it nonetheless has practical applications in the usual fields where heavy tail phenomena occur, such as insurance, finance, or teletraffic data analysis. For instance, the data studied in Bottolo et al (2003) consists of insurance claims having different origins, and it is of interest to decide whether the claim type can have an effect on the tail behavior of the claim distribution, and if so, which types can be found to be equivalent in this sense. In Mougeot and Tribouley (2010), the authors consider various financial data of different firms, and address the problem of comparing (pairwisely) their associated financial risks through their corresponding tail indices. For this purpose, they proposed a data-driven procedure of comparison of (only) two positive tail indices: in Section 3, we will compare their results to those corresponding to the test statistic we propose, especially in terms of coverage accuracy and power, through a simulation study.

The organization of the paper is classical: in Section 2, the methodology, assumptions, and results are stated, and Section 3 is devoted to a simulation study which shows satisfactory power and coverage accuracy of our method. The proofs are delegated to the Appendix.

2. Methodology and statement of the results

We suppose that we observe $K$ independent samples $(X_1^{(1)})_{i \leq n_1}, \ldots, (X_K^{(K)})_{i \leq n_K}$ satisfying (1), as presented previously. In order to derive our asymptotic result, we need a slightly stronger condition, which specifies the rate of convergence in (1). Denoting by $U_j$ the inverse function of $1/(1 - F_j)$ ($j = 1, \ldots, K$), we suppose that there exists a function $A_j$ tending to 0 at infinity such that

$$U_j(tx)/U_j(t) - x^{\gamma_{0j}} \to x^{\rho_j} \frac{x^{\rho_j} - 1}{\rho_j} \text{, as } t \to +\infty,$$

for all $x > 0$, where $\rho_j < 0$. This so-called second order condition is classical in the extreme value theory framework, and is known to hold for most commonly encountered
heavy-tailed distributions.

The method we propose for testing the null hypothesis $H_0 : \gamma_0 = \ldots = \gamma_0 K$, is inspired by a version of the classical ANOVA test based on the empirical likelihood methodology, proposed by Owen (1991). The starting point is the famous Hill estimator for $\gamma_0 = (\gamma_0, \ldots \gamma_K)$ which is defined as

$$\hat{\gamma}_j := \frac{1}{k_j} \sum_{i=1}^{k_j} Y_i^{(j)},$$

where, denoting by $X^{(j)}_{(1)} \leq \ldots \leq X^{(j)}_{(n_j)}$ the order statistics associated to the $j$-th sample, the log spacings $Y_i^{(j)}$ are defined by

$$Y_i^{(j)} := \log \frac{X^{(j)}_{(n_j - i + 1)}}{X^{(j)}_{(n_j - i)}} \quad (i = 1, \ldots, n_j)$$

and $k_j$, the sample fraction of observations to keep from $(X^{(j)}_{(i)})_{i=1}^{n_j}$, satisfies

$$k_j \to +\infty, \quad k_j / n_j \to 0, \quad \text{and} \quad \sqrt{k_j} A_j n_j / k_j \to 0, \quad \text{as} \quad n_j \to +\infty. \quad (3)$$

Let $n = \sum_{j=1}^{K} n_j$ and $k = \sum_{j=1}^{K} k_j$. Note that we do not use the index $n$ in the definition of the $k_j$s (nor in that of other quantities depending on $n$) in order to lighten the notations.

We define, for any $\gamma = (\gamma_j)_{1 \leq j \leq K}$, the empirical likelihood ratio

$$ELR(\gamma) := \sup_{(p_{ij})} \left\{ \prod_{j=1}^{K} \prod_{i=1}^{k_j} (k p_{ij}) : p_{ij} \geq 0, \quad \sum_{i=1}^{k_j} p_{ij} = 1, \quad \sum_{i=1}^{k_j} p_{ij} (Y_i^{(j)} - \gamma_j) = 0 \right\}.$$ 

Clearly, this function is maximum at $\hat{\gamma} = (\hat{\gamma}_j)_{1 \leq j \leq K}$ and $ELR(\hat{\gamma}) = 1$. Thus, if we note

$$l(\gamma) := -2 \log ELR(\gamma),$$

testing $H_0$ can be based on the statistic

$$\inf_{\gamma \in A} l(\gamma) = -2 \log \frac{\sup_{\gamma \in A} ELR(\gamma)}{\sup_{\gamma} ELR(\gamma)},$$

where $A = \{ \gamma \in [0, +\infty)^K, \quad \gamma_1 = \ldots = \gamma_K \}$.

The first part of the following theorem yields the asymptotic distribution of this statistic and thus provides a critical region for this empirical likelihood ratio test of $H_0$, with prescribed asymptotic level. The second part shows that the method extends to more general linear hypotheses on the extreme value indices.

**Theorem 1.** Under assumptions (1)-(3), if $\min_{j \leq K} k_j / k$ is bounded away from 0 then, under $H_0$,

$$\inf_{\gamma \in A} l(\gamma) \xrightarrow{d} \chi^2(K - 1), \quad \text{as} \quad n \to +\infty.$$ 

Moreover, if $H'_0 : " C \gamma_0 = 0 "$, where $\gamma_0 = (\gamma_{0j})_{1 \leq j \leq K}$ and $C$ is a $d \times K$ matrix with full
rank 1 ≤ d < K, then under $H'_0$,

$$\inf_{\gamma \in A'} l(\gamma) \xrightarrow{d} \chi^2(d), \quad \text{as } n \to +\infty$$

where $A' = \{ \gamma \in [0, +\infty]^K : C\gamma = 0 \}$. 

In order to prove Theorem 1, we have to introduce some important quantities and provide their asymptotic behavior in Lemma 1 stated below: for any $\gamma = (\gamma_j)_{1 \leq j \leq K}$ with positive components,

$$G_n(\gamma) := (\hat{\gamma}_j - \gamma_j)^2$$

$S_j^2(\gamma) := \frac{1}{k_j} \sum_{i=1}^{k_j} (Y_i^{(j)} - \gamma_j)^2$

$\tilde{G}_n(\gamma) := \left(\sqrt{k_j}(\hat{\gamma}_j - \gamma_j)\right)_{1 \leq j \leq K}$

$B_n(\gamma) := \text{diag}(S_1^2(\gamma), \ldots, S_K^2(\gamma))$

$M_n(\gamma) := \max_{1 \leq j \leq K} \max_{1 \leq k \leq k_j} |Y_i^{(j)} - \gamma_j|$

Note that the proof of Theorem 1 will make it clear that the statistic $\inf_{\gamma \in A} l(\gamma)$, seemingly lengthy to compute, is very close to the quadratic quantity $\sum_{j=1}^{K} k_j (\hat{\gamma}_j - \gamma_j)^2 / S_j(\hat{\gamma})$ where $\tilde{\gamma} = \left(\tilde{\gamma}_1, \ldots, \tilde{\gamma}_K\right)$ and $\tilde{\gamma}$ is defined in Equation (4).

**Lemma 1.** Under the conditions of Theorem 1, if $B := \text{diag}(\gamma_0^2, \ldots, \gamma_0^2)$, then, as $n \to +\infty$, we have

$$G_n(\gamma_0) \xrightarrow{P} 0, \quad \tilde{G}_n(\gamma_0) \xrightarrow{d} N(0, B), \quad B_n(\gamma_0) \xrightarrow{P} B, \quad M_n(\gamma_0) = o_P(\sqrt{k}).$$

The results contained in this Lemma are due to the asymptotic normality of the Hill estimator (see de Haan and Peng (1998) for example) and to equations (8)-(9) of J.C. Lu and L. Peng (2002).

**Proof of Theorem 1**

- **First step**: we shall prove in Subsection 5.1 that for any $C > 0$,

$$l(\gamma) = Q(\gamma) + o_P(1), \quad \text{uniformly for } \gamma \in B_{n,C},$$

where

$$Q(\gamma) := \tilde{G}_n(\gamma)(B_n(\hat{\gamma}))^{-1}\tilde{G}_n(\gamma) = \sum_{j=1}^{K} k_j (\hat{\gamma}_j - \gamma_j)^2 / S_j(\hat{\gamma}).$$

and

$$B_{n,C} := \{ \gamma : ||\gamma - \gamma_0|| \leq Ck^{-1/2} \}$$

- **Second step**: we shall prove in Subsection 5.2 that

$$\inf_{\gamma \in A} Q(\gamma) = Q(\hat{\gamma}) \xrightarrow{d} \chi^2(K - 1), \quad \text{as } n \to +\infty.$$
since this implies the contiguity of $l(\bar{\gamma})$ and $Q(\bar{\gamma})$. Using the assumption that
$
\min_{0 \leq k} k_j/k$ is bounded away from 0, we can find a positive constant $c$ (not de-
pendent on $C$) such that

$$
\mathbb{P} [ \bar{\gamma} \notin B_{n,C} ] \leq \mathbb{P} [ |\bar{\gamma} - \gamma_0| > K^{-1/2} C k^{-1/2} ]
\leq \mathbb{P} [ \sum_{j=1}^{K} |\bar{\gamma}_j - \gamma_{0j}| > K^{-1/2} C k^{-1/2} ]
\leq \sum_{j=1}^{K} \mathbb{P} [ \sqrt{k_j} |\bar{\gamma}_j - \gamma_{0j}| > cC ]
$$

By the asymptotic normality of the Hill estimators of the $K$ extreme value indices
(stated in Lemma 1), the right-hand side of the last inequation converges (as $n \to \infty$)
to a quantity which vanishes as $C$ goes to $+\infty$.

Remark 1. According to (5) and (6), we can use in practice indifferently $l(\bar{\gamma})$ or $Q(\bar{\gamma})$
as a statistic for our test. As expected and can be seen from assumption (3) and Figure
1, the accuracy of our procedure depends on the choice of the sample fractions $k_j$, for
$j = 1, \ldots, K$, as for the estimation problem. We shall present, in Section 3 below, the
method we used for choosing adequate values for these tuning parameters.

3. Simulations

The purpose of this section is to investigate, through extensive simulations, the per-
formance of our test procedure in the case of the comparison of tail indices of two samples
$X_1$ and $X_2$ (case $K = 2$). In this testing framework, we naturally focus our attention on
both the empirical type I error and power function of the procedure.

As was made in Mougeot and Tribouley (2010), we made simulations based on 2000
random samples of size $n = 800$ (for the first sample $X_1$) and $m = 700$ (for the second
sample $X_2$) generated from the following families of distributions which satisfy the first
and second order conditions (1) and (2) :

- The Fréchet distribution (denoted by $F$) with parameter $\gamma > 0$ given by $F(x) = \exp(-x^{-1/\gamma}) 1_{x > 0}$, for which $\rho = -1$.
- The Student distribution (denoted by $t$) with $\nu$ degrees of freedom, for which
$\gamma = 1/\nu$ and $\rho = -2\gamma$.
- The Burr distribution (denoted by $B$) with parameters $\gamma > 0$ and $\rho < 0$, given by
$F(x) = 1 - (1 + x^{\rho/\gamma})^{-1/\rho} 1_{x > 0}$.

In Subsections 3.1 and 3.2, we apply the method proposed by Hall and Welsh (1985)
for the estimation of the optimal (in the sense of minimization of the mean square error)
sample fractions $k_1$ and $k_2$ given by :

$$
\hat{k}_1 = n^{-2\rho_j/(1-2\rho_j)} \quad \text{and} \quad \hat{k}_2 = m^{-2\rho_j/(1-2\rho_j)},
$$

where $\hat{\rho}_j$ is the estimator of the second order parameter $\rho_j$ ($j = 1$ or 2). We used the one
proposed in Fraga Alves et al (2003) for comparison reasons. Recent alternatives can be
found in Worms (2012), Ciuperca and Mercadier (2010), Goegebeur et al (2010) and
de Wet et al (2012). Results for the type I error are given in Subsection 3.1 and for the
power function in Subsection 3.2.
In Subsection 3.1, we also investigated the effect of the choice of \(k_1\) and \(k_2\) on the type I accuracy of the test, by performing simulations under a large range of values for the sample fractions \(k_1\) and \(k_2\).

3.1. Type I error estimation

Table 1 provides the estimated type I error \(\hat{\alpha}\) at the optimal sample fractions mentioned above, for nominal risks \(\alpha = 5\%\) and \(\alpha = 10\%\), and for different combinations of the above mentioned distributions, where the second order parameters \(\rho_1\) and \(\rho_2\) can be equal or different. We add to our results those of Mougeot and Tribouley (2010) taken from their Table 2, when using the Hill estimator. In the following table, our results are noted WW and those of Mougeot and Tribouley (2010), MT.

**Table 1. Type I error \(\hat{\alpha}\)**

<table>
<thead>
<tr>
<th>Distr.</th>
<th>(\gamma_1 = \gamma_2)</th>
<th>(\rho_1)</th>
<th>(\rho_2)</th>
<th>WW (\hat{\alpha}) (5%)</th>
<th>WW (\hat{\alpha}) (10%)</th>
<th>MT (\hat{\alpha}) (5%)</th>
<th>MT (\hat{\alpha}) (10%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tt</td>
<td>0.25 -0.5 -0.5</td>
<td>4.50</td>
<td>8.65</td>
<td>6</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tt</td>
<td>0.5 -1 -1</td>
<td>5.10</td>
<td>10.85</td>
<td>11</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tt</td>
<td>1 -2 -2</td>
<td>18.10</td>
<td>25.30</td>
<td>9</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>0.25 -1 -1</td>
<td>6.10</td>
<td>12.25</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>0.5 -1 -1</td>
<td>6.50</td>
<td>10.70</td>
<td>2</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FF</td>
<td>1 -1 -1</td>
<td>6.65</td>
<td>10.10</td>
<td>3</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>0.25 -2 -2</td>
<td>4.25</td>
<td>9.00</td>
<td>3</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>0.25 -1 -2</td>
<td>5.00</td>
<td>10.00</td>
<td>14</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>1 -1 -1</td>
<td>5.40</td>
<td>9.95</td>
<td>9</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>1 -1 -2</td>
<td>5.30</td>
<td>11.35</td>
<td>14</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tF</td>
<td>0.5 -1 -1</td>
<td>6.15</td>
<td>10.80</td>
<td>33</td>
<td>57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tF</td>
<td>1 -2 -1</td>
<td>7.55</td>
<td>12.60</td>
<td>7</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tB</td>
<td>0.5 -1 -0.5</td>
<td>15.00</td>
<td>24.80</td>
<td>16</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tB</td>
<td>1 -2 -1</td>
<td>5.65</td>
<td>11.00</td>
<td>10</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tB</td>
<td>1 -2 -2</td>
<td>6.60</td>
<td>14.00</td>
<td>8</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB</td>
<td>0.25 -1 -1</td>
<td>5.60</td>
<td>10.55</td>
<td>12</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB</td>
<td>0.5 -1 -1</td>
<td>5.50</td>
<td>10.85</td>
<td>10</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB</td>
<td>1 -1 -1</td>
<td>5.80</td>
<td>10.20</td>
<td>14</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB</td>
<td>0.25 -1 -2</td>
<td>5.30</td>
<td>11.90</td>
<td>3</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB</td>
<td>0.5 -1 -2</td>
<td>5.80</td>
<td>11.15</td>
<td>2</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FB</td>
<td>1 -1 -2</td>
<td>6.75</td>
<td>12.35</td>
<td>3</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We see, through these simulations, that the procedure we propose, though simple to implement, yields generally sharper, more satisfactory results than those from Mougeot and Tribouley (2010) (except for a few cases), even without using any bootstrap calibration for our test. The same conclusion holds when, in their method, the Hill estimator is replaced by the other estimators they considered in their simulations.

We also present, in Figure 1, some contour graphs of the empirical type I error \(\hat{\alpha}\), for a wide range of the sample fractions \(k_1\) and \(k_2\). We can observe that the accuracy of course depends on the choice of these fractions, but there is some kind of permissibility for this choice before obtaining unsatisfactory results (here “satisfactory” was arbitrarily defined as “being included in the interval \([3\%, 7\%]\)”, for a nominal risk \(\alpha = 5\%)\).
(a) $X_1 \sim \text{Burr}(1/4, -2), X_2 \sim \text{Burr}(1/4, -2)$

(b) $X_1 \sim \text{Burr}(1/4, -1), X_2 \sim \text{Burr}(1/4, -2)$

(c) $X_1 \sim \text{Frechet}(1/2), X_2 \sim \text{Burr}(1/2, -1)$

(d) $X_1 \sim \text{Frechet}(1/2), X_2 \sim \text{Burr}(1/2, -2)$

Figure 1: Estimated type I error $\hat{\alpha}$ as a function of the sample fractions $k_1$ and $k_2$, for $\alpha = 5\%$. White areas correspond to cases where $\hat{\alpha}$ turned out to be greater than 7\%
3.2. Power function estimation

In this part, we get the extreme value index $\gamma_1$ of $X_1$ fixed and calculate the power associated to our test procedure for different values of $\gamma_2$, the extreme value index of $X_2$. For example, if $\gamma_1 = 1/2$, we take $\gamma_2$ between 0.1 and 1. Moreover, $\rho_1$ and $\rho_2$ can be taken equal or different. We use again the method of Hall and Welsh (1985) to estimate the optimal $k_1$ and $k_2$ values. Figure 2 gives two examples of estimated power functions for Burr and Fréchet distributions.

In both examples, the graphic on the left is such that $\rho_1$ is always equal to $\rho_2$, whereas the graphic on the right is such that $\rho_1$ is always different from $\rho_2$. We remarked through these two examples and others not presented here, that when the two distributions of $X_1$ and $X_2$ have the same second order parameters $\rho_1$ and $\rho_2$, the estimated power function is, as expected, minimal when the tail indices $\gamma_1$ and $\gamma_2$ are equal. This is not the case when the second order parameters $\rho_1$ and $\rho_2$ are different.

The empirical results look rather satisfactory, though no comparison with other methods can be made due to the fact that the extreme value indices comparison topic has hardly been addressed in the literature before.

4. Conclusion

In this work, we presented a first attempt to address the problem of comparing extreme value indices of different heavy-tailed samples, through some statistical test with prescribed theoretical accuracy. Exception made of Mougeot and Tribouley (2010), we know of no work addressing explicitly and methodologically this topic in the extreme value literature. The method we propose (i) appears to perform relatively well in terms of empirical type I error accuracy and power, (ii) is valid for comparison of more than 2 samples, and (iii) allows for testing linear hypotheses more general than just the equality of all the extreme value indices. Its inspiration stems from the empirical likelihood work on linear models in Owen (1991) and from J.C. Lu and L. Peng (2002), which yields confidence intervals for the tail index based on the Hill estimator.

One possibility to generalize this work would be to define a similar test statistic as ours, but based on more efficient tail index estimators than the Hill estimator; it should though be stressed that the particular structure of the Hill estimator has been used here to prove the asymptotic distribution of the test statistic (the Hill estimator is the mean of log-spacings and its asymptotic variance is a simple function of the tail index alone). Another path to pursue the study of this topic could be to use known results on regression models for the tail index.

5. Appendix

5.1. Proof of (5)

Let $m_0 = 0$ and $m_j = k_1 + \ldots + k_j$ (for $j \leq K$). We denote by $(Z_{l,n}(\gamma))_{l \leq K}$ the $K$-dimensional vector for which, if $l \in \{m_{j-1} + 1, \ldots, m_j\}$,

$$(Z_{l,n}(\gamma))_i := \begin{cases} 0 & \text{if } t \neq j \\ Y^{(j)}_i - \gamma_j & \text{if } t = j, \text{ with } i = l - m_{j-1} \end{cases}$$
(a) $X_1 \sim B(1/2, -1)$, and $X_2 \sim B(1/2, \rho_2)$, $\rho_2 = -1$ or $-1/2$

(b) $X_1 \sim B(1/2, \rho_1)$, $\rho_1 = -1$ or $-1/2$, and $X_2 \sim F(1/2)$

Figure 2: Estimated power function as $\gamma_2$ varies, when $\gamma_1$ is kept fixed
ELR(γ) can then be rewritten as (remind that \( k = \sum_{j=1}^{K} k_j \))

\[
ELR(\gamma) := \sup_{(p_l)} \left\{ \frac{1}{k} \prod_{l=1}^{k} (k p_l) : p_l \geq 0, \sum_{l=1}^{k} p_l = 1, \sum_{l=1}^{k} p_l \bar{Z}_{l,n}(\gamma) = 0 \right\},
\]

The classical Lagrange multipliers method yields the following optimal weights \( p_l, n(\gamma) \) in the above description of \( ELR(\gamma) \)

\[
p_{l,n}(\gamma) := \frac{1}{k(1 + \gamma_{l,n}(\gamma), Z_{l,n}(\gamma))}
\]

where \( \gamma_{n}(\gamma) \) is determined as the solution of

\[
\sum_{l=1}^{k} (1 + \gamma_{l,n}(\gamma), Z_{l,n}(\gamma))^{|-1|} Z_{l,n}(\gamma) = 0,
\]

so that

\[
l(\gamma) = 2 \sum_{l=1}^{k} \log(1 + \gamma_{l,n}(\gamma), Z_{l,n}(\gamma)) > 0.
\]

Now, if we note \( \Delta_{n} := \text{diag}(k_1/k, \ldots, k_K/k) \) and

\[
\tilde{Z}_{n}(\gamma) := \frac{1}{k} \sum_{l=1}^{k} Z_{l,n}, \quad \hat{V}_{n}(\gamma) := \frac{1}{k} \sum_{l=1}^{k} Z_{l,n} Z_{l,n}^{t}, \quad Z_{n}^{*}(\gamma) := \max_{1 \leq i \leq k} ||Z_{i,n}(\gamma)||,
\]

then we have (remind the definitions of \( G_{n}, \hat{G}_{n}, B_{n} \) and \( M_{n} \) before the statement of Lemma 1), for any given \( C > 0 \) and \( \gamma \in B_{n,C} \),

\[
\tilde{Z}_{n}(\gamma) = \Delta_{n} G_{n}(\gamma) = \Delta_{n} G_{n}(\gamma_{0}) + \Delta_{n}(\gamma_{0} - \gamma), \quad \hat{V}_{n}(\gamma) = \Delta_{n} B_{n}(\gamma) = \Delta_{n} B_{n}(\gamma_{0}) + \Delta_{n} \sqrt{\kappa}(\gamma_{0} - \gamma),
\]

\[
Z_{n}^{*}(\gamma) = M_{n}(\gamma) \leq M(\gamma_{0}) + \kappa ||\gamma_{0} - \gamma||
\]

where \( O_{n} \) is a \( K \times K \) matrix such that \( ||O_{n}|| = O_{p}(1) \), and \( \kappa \) is an absolute constant.

Please note that, in the sequel, all the asymptotic results (in \( o_{p} \) or \( O_{p} \)) are established \( uniformly \) for \( \gamma \in B_{n,C} \).

Now, as usual in empirical likelihood methodology, starting from the definition of \( \lambda_{n}(\gamma) \) we obtain

\[
||\lambda_{n}(\gamma)|| (u_{n}^{t}(\gamma) \hat{V}_{n}(\gamma) u_{n}(\gamma) - Z_{n}^{*}(\gamma)||Z_{n}(\gamma)||) \leq ||\tilde{Z}_{n}(\gamma)||
\]

where \( u_{n}(\gamma) := \lambda_{n}(\gamma)/||\lambda_{n}(\gamma)|| \). It is clear from Lemma 1, relations (9) and (11), and the definition of the ball \( B_{n,C} \) that

\[
\sqrt{\kappa}||\tilde{Z}_{n}(\gamma)|| = O_{p}(1) \quad \text{and} \quad Z_{n}^{*}(\gamma) = o_{p}(\sqrt{\kappa}).
\]

Consequently,

\[
||\lambda_{n}(\gamma)|| = O_{p}(k^{-1/2})
\]

since \( u_{n}^{t}(\gamma) \hat{V}_{n}(\gamma) u_{n}(\gamma) \) is bounded away from 0, which is due to the assumption that \( k_{j}/k \) is bounded away from 0 (which implies that \( ||\Delta_{n}|| = O_{p}(1) \)) and that \( \hat{V}_{n}(\gamma) = \Delta_{n}^{2} B \Delta_{n}^{1/2} + o_{p}(1) \), with \( B \) invertible (since the indices \( \gamma_{0j} \) are positive).
Therefore, setting \( g_{1,n}(\gamma) := <\lambda_n(\gamma), Z_{1,n}(\gamma)> \) we have

\[
g_n^* := \sup\{||g_{l,n}(\gamma)||, \gamma \in B_{n,C}\} = o_P(1). \tag{12}
\]

It follows classically (from the definition of \( \lambda_n(\gamma) \)) that

\[
\hat{V}_n(\gamma)\lambda_n(\gamma) = Z_n(\gamma) + R_n(\gamma),
\]

where \( ||R_n(\gamma)|| \leq 2(g_n^*)^2||Z_n(\gamma)|| = o_P(k^{-1/2}). \) By Lemma 1, it is simple to see that \( B_n(\gamma) = B_n(\hat{\gamma}) + \mathcal{O}_P(1) \) and thus (10) leads to

\[
\sqrt{k}\lambda_n(\gamma) = (B_n(\hat{\gamma}))^{-1}\Delta_n^{-1/2}\hat{G}_n(\gamma) + \mathcal{O}_P(1). \tag{13}
\]

Using (12), a Taylor expansion yields

\[
l(\gamma) = 2\sum_{l=1}^{k} \log(1 + g_{l,n}(\gamma)) = 2\sum_{l=1}^{k} g_{l,n}(\gamma) - \sum_{l=1}^{k} (g_{l,n}(\gamma))^2 + R'_n(\gamma),
\]

where \( R'_n(\gamma) = \frac{2}{3} \sum_{l=1}^{k} (g_{l,n}(\gamma))^3(1 + \xi_{l,n}(\gamma))^{-3} \) for some \( \xi_{l,n}(\gamma) \in ]0, g_{l,n}(\gamma)\]. Moreover, by (9), (10) and (13),

\[
\sum_{l=1}^{k} g_{l,n}(\gamma) = (\sqrt{k}\lambda_n(\gamma))^4(\sqrt{k}Z_n(\gamma)) = (\hat{G}_n(\gamma))^4(B_n(\hat{\gamma}))^{-1}\hat{G}_n(\gamma) + \mathcal{O}_P(1)
\]

and

\[
\sum_{l=1}^{k} (g_{l,n}(\gamma))^2 = (\sqrt{k}\lambda_n(\gamma))^4(\hat{V}_n(\gamma))(\sqrt{k}\lambda_n(\gamma)) = (\hat{G}_n(\gamma))^4(B_n(\hat{\gamma}))^{-1}(\Delta_n^{-1/2}B_n(\hat{\gamma})) + \mathcal{O}_P(1))(B_n(\hat{\gamma}))^{-1}\Delta_n^{-1/2}(\hat{G}_n(\gamma) + \mathcal{O}_P(1))
\]

Finally, by (12) it comes \( |R'_n(\gamma)| \leq \mathcal{O}_P(1) \sum_{l=1}^{k} (g_{l,n}(\gamma))^2 = \mathcal{O}_P(1) \) and

\[
l(\gamma) = (\hat{G}_n(\gamma))^4(B_n(\hat{\gamma}))^{-1}\hat{G}_n(\gamma) + \mathcal{O}_P(1) = Q(\gamma) + \mathcal{O}_P(1),
\]

which concludes the first step of the proof of Theorem 1. Note that in the last equation, \( B_n(\hat{\gamma}) \) can be replaced indifferently by \( B_n(\gamma) \) or \( B_n(\gamma_0) \) without invalidating the approximation.

5.2. Proof of (6)

Recalling that \( \mathcal{A} = \{\gamma \in ]0, +\infty[^K, \gamma_1 = \ldots = \gamma_K\} \) and that \( \delta_j = \sqrt{k_j/S_j(\gamma)} \), we readily have

\[
\inf_{\gamma \in \mathcal{A}} Q(\gamma) = Q(\bar{\gamma}, \ldots, \bar{\gamma}), \quad \text{where} \quad \bar{\gamma} = \sum_{j=1}^{K}\delta_j^2\hat{\gamma}_j/\sum_{j=1}^{K}\delta_j^2.
\tag{14}
\]

Moreover,

\[
\inf_{\gamma \in \mathcal{A}} Q(\gamma) = ||W - W'||^2,
\]

where \( W = (\sqrt{k_j}(\hat{\gamma}_j - \gamma_{0j})/S_j(\bar{\gamma}))_{j=1...K} \) and \( W' = (\delta_j(\hat{\gamma}_j - \gamma_{0j}))_{j=1...K} \). By Lemma 1, positiveness of the \( \gamma_{0j} \), and the fact that \( S_j^2(\bar{\gamma}) - S_j^2(\gamma_0) \stackrel{p}{\to} 0 \), we see that \( W \) is asymptotically distributed as \( NK(0, I) \).
Moreover, setting (under $H_0$) $\gamma^* = \gamma_{01} = \ldots = \gamma_{0K}$, if we introduce $p = \sum_{j=1}^{K} \delta_j^2$, $\delta = (\delta_j)_{j=1, \ldots, K}$ and $v_n = \frac{\delta}{\sqrt{p}}$, then, under $H_0$,
\[ < W, v_n > v_n = \sum_{j=1}^{K} \left( \frac{\delta_j}{\sqrt{p}} \delta_j (\hat{\gamma}_j - \gamma_{0j}) \right) v_n = \hat{\gamma} - \gamma^* \delta = W', \]
so that $W'$ is the orthogonal projection of $W$ on a subspace of dimension 1. Equation (6) follows.

Finally, if the hypothesis tested is $H'_0$: “$C\gamma_0 = 0$” as in the second part of Theorem 1, then let us set $\Delta = \text{diag}(\delta_1, \ldots, \delta_K)$, $\tilde{C} = C\Delta^{-1}$, $w_0 = \Delta \gamma_0$. Since $W = \Delta (\hat{\gamma} - \gamma_0)$ is asymptotically distributed as $N_K(0, I)$ and, under $H'_0$, we have $\tilde{C}w_0 = 0$, it comes
\[ \inf_{\gamma \in \mathcal{A}'} Q(\gamma) = \inf_{\gamma} \{ \| \Delta (\hat{\gamma} - \gamma) \| ; C\gamma = 0 \} = \inf_{w} \{ \| \Delta (\hat{\gamma} - \gamma_0) - w \| ; \tilde{C}w = 0 \} \]
Because the subspace $\{ w ; \tilde{C}w = 0 \}$ is $(K - d)$-dimensional, the result thus comes by Cochran’s theorem.

References