Ann Inst Stat Math (2010) 62:383–412
DOI 10.1007/s10463-008-0198-9

Procedure of test to compare the tail indices

Mathilde Mougeot · Karine Tribouley

Received: 13 June 2007 / Revised: 21 March 2008 / Published online: 25 September 2008

Abstract We propose a test procedure which compares the extreme value indices of two samples with heavy tail distributions. On a theoretical point of view, we adopt the minimax nonparametric point of view. We exhibit the separating rate between the null hypothesis and the alternative of our procedure. Next, we present a data driven test methodology and we evaluate its performance thanks to an extensive simulation study. As a practical real-life application, we compare the risk behaviors of a panel of different financial data.

Keywords Extreme value index · Nonparametric test · Minimax rates

1 Introduction

Our aim in this paper is to give methods which compare the behavior of two series with respect to their risk. We consider that the risk of a series is given by the tail parameter of the distribution of the log returns of the series. We assume that the log return $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ associated to the series have distributions denoted by F_X, F_Y. Their tails are modeled in the usual semiparametric way, following a Pareto distribution with parameter $\gamma > 0$ multiplied by an infinite dimensional parameter $L(\cdot)$

$$1 - F_X(x) = x^{-1/\gamma_X} L_X(x) \quad \text{and} \quad 1 - F_Y(x) = x^{-1/\gamma_Y} L_Y(x) \tag{1}$$

M. Mougeot (✉) · K. Tribouley
ModalX, Batiment G, Université Paris X-Nanterre,
200 avenue de la République, 92001 Nanterre, France
e-mail: mathilde.mougeot@u-paris10.fr

K. Tribouley
e-mail: karine.tribouley@u-paris10.fr

for x large enough. The nuisance parameter $L(\cdot)$ is supposed to be slowly varying: this first order condition is defined in (3). Our aim is to compare the tail indices γ_X and γ_Y of both series; roughly speaking, that means that we want to solve the test problem with the null hypothesis $H_0 : \gamma_X = \gamma_Y$.

The minimax test theory has been initiated by Ingster (1993) and developed in the classical models of the nonparametric estimation. The specificities are the following: first, the null hypothesis H_0 and the alternative H_1 have symmetric role when the error of decision is computed (the risk of the test is the sum of the first type risk and of the second type risk); secondly, we focus on the rate of separation of the hypotheses. It is a real challenge to determine the smaller distance allowing to distinguish between H_0 and H_1. Last, the alternative is precisely defined: generally, the hypothesis is the intersection between a regularity constraint and a geometrical constraint measuring the distance between H_1 and H_0 and depending on a norm chosen by the practitioner. This point of view is completely different from the other classical point of view where the nonparametric test is built on the distribution function. In this case, it is unusual to give the alternative; the law of the statistic of test is studied under H_0 but nothing is known under the alternative and then, the risk of the considered test can not be bounded but under H_0. This explains why the practitioner introduces some nonsymmetry between H_0 and H_1. These type of tests are generally excellent for monitoring problems because they allow to eliminate data which are not conform to H_0. But, few results on the power are given implying that it is impossible to compare different tests and to give optimality results.

For the comparison of the risks of series, a first idea is to apply Butucea and Tribouley (2006) who propose procedures to test the homogeneity of two samples: the null hypothesis is $H_0 : f_X = f_Y$ where f_X, f_Y are the densities associated to the distribution functions F_X, F_Y. This approach is not convenient here because we do not want to compare the entire density or distribution function but the behavior of the tails. We focus on the tail indices of the distribution functions, building test procedures associated to different estimation procedures of these tail indices. This problem of estimation of the tail index γ has been extensively explored: for a first approach, see Embrechts et al. (1997). The proposed estimators are depending on a parameter k whose the influence is similar to the influence of the smoothing parameter in the nonparametric estimation theory. The parameter k is the number of order statistics of the sample needed for giving information on the tail of the distribution. For choosing k, an analogue phenomenon with the usual trade-off between the bias term and the variance term when estimating curves appears. A second order condition is needed to determine the asymptotic optimal value of k.

When dealing with the test problem, our comparison procedure depends again on parameter k to be determined; a trade-off between bias term and variance term is needed leading to an optimal choice of parameter k^* which is the same as in the estimation problem. Next, we exhibit the separating rate of the null hypothesis H_0 from the alternative H_1 (to a prescribed level) which is the best achievable among the test family that we consider. To obtain uniform results, we need some restriction on the space of the considered distributions and introduce a kind of third order condition. Observe that an analogue is used in Fraga Alves et al. (2003) in the estimation problem.

Our procedure of test associated to k^* depends on extra parameters and then is not adaptive. Usually (see for instance Butucea and Tribouley 2006), in the theory of nonparametric tests, it is fundamental that the test statistic is unbiased under H_0. It allows to deduce adaptive procedures using Bonferroni's methods. Unfortunately, the test statistics based on estimators of the tail indices are always biased under H_0 as soon as we consider H_0 as a composite hypothesis.

On a practical point of view, we provide a data driven procedure of comparison using an estimator of k^* proposed by Hall and Welsh (1985). Many other recent and sophisticated procedures could be used (see by instance Fraga Alves et al. 2003) but we show that the practical qualities of our test are good even with the very simple and natural procedure introduced by Hall and Welsh (1985). The performances of our test procedure are studied on simulated samples of specific distributions. First, we provide a detailed analyze of the adaptive estimation of the tail parameters using an iterative improvement of the procedure by Hall and Welsh (1985). The improvement consists in reducing significatively the impact of initial values but is still biased. We then carry out an extensive simulation study to compute the test empirical levels and powers for various conditions for the first and second order parameters. The results are very encouraging: the estimators of the level of the test are generally good even in the case where the second order parameter is very small or when we consider the limit case where the second order parameter should be zero associated to a log function. We observe that the second order parameter has no significative impact on the estimation of the level. Nevertheless, the estimation of the power of the test does not provide such a good result. We think that we pay here the fact that the test statistic is biased even under the null hypothesis. From a theoretical point of view, that means that the separating rate appears to be very slow.

As a practical use, we also present an application to a real-life data set. Real financial data are compared using the previous test procedure on their log return series. Assuming that the model given in (1) is valid for the considered financial data, we presume that their risk is growing with the tail index γ. Different types of series with various associated risk are chosen as single share, indexes, bonds, raw material prices and currencies.

The paper is organized as follows. In Sect. 2, we present the test procedures. In Sect. 3, we define the functional spaces that we consider and we state a theoretical frame for the test problem. In Sect. 4, we give the main results about the rate and the optimality of the previous procedures. Section 5 is devoted to the empirical study. The proofs are postponed to Sect. 6.

2 Procedures

In this section, we propose procedures to test the null hypothesis $H_0 : \quad \gamma_X = \gamma_Y$ where γ_X (respectively γ_Y) is the tail index of the sample $X_1, \ldots, X_n$ (respectively $Y_1, \ldots, Y_m$). First we describe the methods for estimating the tail index: we consider the same methods as studied in de Haan and Peng (1998). Next, we give the test procedures based on these estimation methods.

Let p be either n or m and Z be X either Y and let k vary between 1 and p. Denote $Z_{(\cdot:p)}$ the sequence of the ordered variables:

$$Z_{(1:p)} \leq \cdots \leq Z_{(k:p)} \leq \cdots \leq Z_{(p:p)}.$$

We consider the usual estimators of the tail index suggested by Hill (1975), Dekkers et al. (1989), and de Vries (following de Haan and Peng 1998)

$$\hat{\gamma}_Z^1(k) = \frac{1}{k}\sum_{i=1}^{k} \log \frac{Z_{(p-i:p)}}{Z_{(p-k:p)}},$$

$$\hat{\gamma}_Z^2(k) = \hat{\gamma}_Z^1(k) + 1 - \frac{1}{2}\left(1 - \frac{(\hat{\gamma}_Z^1(k))^2}{M(k)}\right)^{-1},$$

$$\hat{\gamma}_Z^3(k) = \frac{M(k)}{2\hat{\gamma}_Z^1(k)},$$

where

$$M(k) = \frac{1}{k}\sum_{i=1}^{k}\left(\log \frac{Z_{(p-i:p)}}{Z_{(p-k:p)}}\right)^2.$$

Let now $\mathcal{K}$ be a subset of $\{1, \ldots, n\} \times \{1, \ldots, m\}$ and let us fix the smoothing parameters (k, k') in $\mathcal{K}$. For each method $l = 1, 2, 3$, we define the sequence of **test statistics** $D^l_{k,k'}$ comparing the estimator $T^l_{k,k'} = \hat{\gamma}^l_X(k) - \hat{\gamma}^l_Y(k')$ to a **critical value**, $t^l_{k,k'} > 0$

$$D^l_{k,k'} = \begin{cases} 0 \;\text{ if }\; |\hat{\gamma}^l_X(k) - \hat{\gamma}^l_Y(k')| \leq t^l_{k,k'} \\ \\ 1 \;\text{ if }\; |\hat{\gamma}^l_X(k) - \hat{\gamma}^l_Y(k')| > t^l_{k,k'}. \end{cases} \tag{2}$$

which means that we decide H_0 if $|T^l_{k,k'}| \leq t^l_{k,k'}$ and reject H_0 otherwise. This test procedure is depending on the smoothing parameter (k, k'). In Sect. 4, we explain how to choose (k, k') and the critical values $t^l_{k,k'}$ such that our procedures have good theoretical properties.

3 Test problem

We first state the assumptions on the considered distributions. The first and second order conditions are very usual and also needed to obtain results in the problem of estimation. We need extra assumptions: we add a third order condition. Our assumptions are more restrictive than the third order condition used in Fraga Alves et al. (2003) (in the case of the estimation problem) because we establish uniform results. Next, we define precisely our test problem (the null hypothesis and the alternative) and we describe the optimality criterion allowing to decide of the quality of the procedures.

3.1 Assumptions

First, we recall the definitions which are usual in the problem of the estimation of the tail index; these definitions lead to **the first order condition** and to **the second order condition**.

Definition 1 Let γ be a real. $G \in RV_{-1/\gamma}$ if there exists a function $g(z,t)$ tending to zero when t tends to infinity such that

$$\forall z > 0, \forall t > 0, \quad \frac{G(tz)}{G(t)} = z^{-1/\gamma}(1 + g(z,t)), \tag{3}$$

γ is called the first order parameter.

A second order condition is needed, specifying the rate of convergence in (3). When G is a survival function, it is usual to give the above definition in terms of the inverse of the distribution function $F = 1 - G$; U denotes the right continuous inverse of the function $(1 - F)^{-1}$.

Definition 2 Let $\gamma, \rho \geq 0$. $1-F \in RV_{-1/\gamma,\rho}$ if there exist a function $A(t)$ of constant sign and a function $R(z,t)$ tending to zero when t is going to infinity such that

$$\forall z > 1, \forall t > 1, \quad \frac{\frac{U(tz)}{U(t)} - z^{\gamma}}{A(t)} = z^{\gamma}\left(\frac{1 - z^{-\rho}}{\rho}\right)(1 + R(z,t)), \tag{4}$$

we call ρ the second order parameter.

For our purpose, we need to control more precisely the behavior of the functions A, g or R and we add two notations.

1. Let A, R be functions as described in Definition 2. We denote $RV_{-1/\gamma,\rho}(A,R)$ the set of the survival functions satisfying (4) for these given functions A, R.
2. Let $\tilde{g}$ be a nonnegative function tending to zero at infinity. We denote $RV_{-1/\gamma}(\tilde{g})$ the set of functions G satisfying (3) for some function g such

   $$\forall t > 0, \quad \sup_{z \in \mathcal{V}(1)} |g(z,t)| \leq \tilde{g}(t),$$

   where $\mathcal{V}(1)$ is a fixed neighborhood of 1.

We are now ready to describe the set of considered distributions. Let us fix $0 < \rho_1 < \rho_2$ and let us give functions $\tilde{a}_0, \tilde{r}_0$ tending to zero at infinity. Let A be in $RV_{-\rho}(\tilde{a}_0)$ and $\tilde{R}_0$ be in $RV_{\tau'}(\tilde{r}_0)$ for $\rho \in [\rho_1, \rho_2]$ and some $\tau' < 0$. Let us put

$$\Theta = \Bigg\{\theta = (\gamma, \rho, R),\ \gamma > 0, \rho \in [\rho_1, \rho_2], \\ \times\ R : \mathbb{R}^{+2} \to \mathbb{R}, \quad \sup_{z \geq 1} |R(z,t)| \leq \tilde{R}_0(t)\Bigg\}. \tag{5}$$

For any $\theta \in \Theta$, we define the set

$$\mathcal{F}(\theta) = \left\{ F \text{ d.f.}, \; 1 - F \in RV_{-1/\gamma,\rho}(A, R) \right\}. \tag{6}$$

Remark that the condition on $\tilde{R}_0(\cdot)$ is a third order condition but the third order parameter τ' will not play any role. Any other conditions are usual (observe that if the second order condition is satisfied, A has necessary to be in $RV_{-\rho}$) except that we need here some uniformity to have exact controls on the behavior of the tails. These assumptions could be compared with the assumptions on the modulus of continuity when we deal with curves estimation with regularity constraint (by instance, when the signal is supposed to belong to a Hölder class). Many distributions belong to the set $\mathcal{F}(\theta)$: by instance, our hypothesis are satisfied for the Burr(β, 1/2, λ) distribution

$$F(x) = 1 - \left(\frac{\beta}{\beta + \sqrt{x}} \right)^{\lambda}, \quad x > 0$$

for $\gamma = 2/\lambda, \rho = 1/\lambda, A(t) = 2/\lambda \; (t^{1/\lambda} - 1)^{-1}$ and $\tilde{a}_0(t) = \tilde{R}_0(t) = t^{-1/\zeta_2}$. The second order condition is satisfied by the Cauchy distribution for $\gamma = 1, \rho = 2, A(t) = 2\pi^2/3\, t^{-2}$ but our uniformity conditions are not satisfied because the function $R(z, t)$ is not uniformly bounded on $]1, \infty[$ but is uniformly bounded by ct^{-2} on $[z_0, +\infty[$ where $z_0 > 1$.

3.2 The test problem

Let $\kappa > 0$ and $\phi_{n,m} > 0$ be a function of n, m, called the **rate of convergence of the test**. Let $\tilde{a}_0, \tilde{r}_0$ be functions tending to zero at infinity and $0 < \rho_1 < \rho_2$. Let A_X (respectively A_Y) be in $RV_{-\rho_X}(\tilde{a}_0)$ for $\rho_X \in [\rho_1, \rho_2]$ (respectively in $RV_{-\rho_Y}(\tilde{a}_0)$ for $\rho_Y \in [\rho_1, \rho_2]$) and $\tilde{R}_0 \in RV_\tau(\tilde{r}_0)$ for some τ. We define the following semi parametric set of distribution functions

$$\Sigma_0 = \bigcup_{(\theta_X, \theta_Y) \in \Theta^2} \{(F_X, F_Y) \in \mathcal{F}(\theta_X) \times \mathcal{F}(\theta_Y), \; \gamma_X = \gamma_Y\} \tag{7}$$

and

$$\Sigma_{n,m}(\kappa) = \bigcup_{(\theta_X, \theta_Y) \in \Theta^2} \{(F_X, F_Y) \in \mathcal{F}(\theta_X) \times \mathcal{F}(\theta_Y), \; |\gamma_X - \gamma_Y| \geq \kappa \phi_{n,m}\}$$

for Θ defined in (5) and $\mathcal{F}(\theta)$ in (6). Our problem is to test the null hypothesis

$$H_0 : (F_X, F_Y) \in \Sigma_0$$

against the alternative

$$H_1 : (F_X, F_Y) \in \Sigma_{n,m}(\kappa).$$

The sequence $\phi_{n,m}$ measures the rate of separation between the hypothesis H_0 and H_1; it defines a marge where it is impossible to distinguish between the hypothesis. Our aim is to obtain tests with $\phi_{n,m}$ as fast as possible. A (nonrandomized) test $D_{n,m}$ is a measurable function from the space of the data $(X_1, \ldots, X_n, Y_1, \ldots Y_m)$ to $\{0, 1\}$ and we define the risk of $D_{n,m}$ as follows

$$\alpha(D_{n,m}, \kappa\phi_{n,m}) = \sup_{(F_X,F_Y)\in\Sigma_0} P_{F_X,F_Y}(D_{n,m} = 1) + \sup_{(F_X,F_Y)\in\Sigma_{n,m}(\kappa)} P_{F_X,F_Y}(D_{n,m} = 0).$$

3.3 Theoretical properties of our procedure

Let us fix a prescribed risk for the test problem stated in the previous section. Theorem 1 ensures that the test procedure described in Sect. 2 answers to the test problem and gives the rate separating the null hypothesis and the alternative achieved for a specific choice of smoothing parameter (k, k').

In the sequel, the various indices k are depending on n either m and are tending to infinity with $n \wedge m$.

Theorem 1 *Let* $0 < \rho_1 < \rho_2$ *and* A_X, A_Y *belonging respectively to* $RV_{-\rho_X}(\tilde{a}_0)$ *and* $RV_{-\rho_Y}(\tilde{a}_0)$ *for* $\rho_X, \rho_Y \in [\rho_1, \rho_2]$ *and for* $\tilde{a}_0$ *tending to zero at infinity; let* $\tilde{R}_0 \in RV_\tau(\tilde{r}_0)$ *for some* τ *and some* $\tilde{r}_0$ *tending to zero at infinity. Consider the test problem*

$$H_0 : (F_X, F_Y) \in \Sigma_0 \quad \textit{against} \quad H_1 : (F_X, F_Y) \in \Sigma_{n,m}(\kappa),$$

where Σ_0 *is defined in* (7)*. Let* $0 < \alpha < 1$ *be the prescribed risk of the test. Let* $\lambda > 0$ **to be chosen** *and denote* k_Z^* *the indices varying between* $\log(p)$ *and* $p/\log(p)$ *satisfying*

$$k_Z^* A_Z^2\left(\frac{p}{k_Z^*}\right) = \lambda. \tag{8}$$

We put $k^* = k_X^* \wedge k_Y^*$*. Then, the corresponding estimator*

$$T_{k_X^*,k_Y^*} = \hat{\gamma}_X^1(k_X^*) - \hat{\gamma}_Y^1(k_Y^*)$$

and the critical value $t_{k^*} = (k^*)^{-1/2}$ *provide test statistics*

$$D_{k_X^*,k_Y^*} = I_{\{|T_{k_X^*,k_Y^*}|>c\, t_{k^*}\}},$$

which achieve the rate of testing $\phi_{n,m}^* = (k^*)^{-1/2}$ *as soon as* κ *is larger than* $3c$ *for*

$$c \geq 4\,\alpha^{-1/2}\left[2((e_X)^2 + (e_Y)^2) + (d_X^2 + d_Y^2)\lambda\right]^{1/2},$$

where $e_Z = \gamma_Z$ *and* $d_Z = (1+\rho_Z)^{-1}$.

We emphasize that we are interested in the rates and that the constants are no optimal. Specifically, the positive constant λ can be everything implying that the rate $\phi^*_{n,m}$ is determined up to a constant.

3.4 A data driven procedure

Observe that in the specific case where $A_Z(z) = c_0 z^{-\rho_Z}$ for some constant $c_0 > 0$ (i.e. $\tilde{a}_0 \equiv 0$), we obtain (choosing in the previous theorem $\lambda = c_0^{-2}$)

$$k^* = n^{\frac{2\rho_X}{2\rho_X+1}} \wedge m^{\frac{2\rho_Y}{2\rho_Y+1}} \tag{9}$$

leading to the separating rate

$$\phi^*_{n,m} = n^{-\frac{\rho_X}{2\rho_X+1}} \vee m^{-\frac{\rho_Y}{2\rho_Y+1}}.$$

Even in this case, the previous result is not tractable in practice because the procedure is not adaptive: it depends on a crucial way on the functions A_X and A_Y which are generally unknown (or on the second order parameters ρ_X, ρ_Y). To build an estimator for k^*, we follow Hall and Welsh (1985). Since we are interested only with the order of k^*, we use their procedure for $\lambda = 1$. Suppose that $\rho_Z \in [\rho_1, \rho_2]$ where $0 < \rho_1 < \rho_2$ are **known**. Choose σ such that $0 < \sigma < 2\rho_1/(2\rho_1+1)$ and then τ_1, τ_2 such that

$$2\rho_2/(2\rho_2+1) \vee (1-\sigma/2\rho_2) < \tau_1 < \tau_2 < 1.$$

Put

$$s = [p^{\sigma}], \quad t_1 = [p^{\tau_1}], \quad t_2 = [p^{\tau_2}]$$

and

$$\hat{\rho}_Z = \left| \frac{\log(|\hat{\gamma}_Z(t_1) - \hat{\gamma}_Z(s)| / |\hat{\gamma}_Z(t_2) - \hat{\gamma}_Z(s)|)}{\log(t_1/t_2)} \right|. \tag{10}$$

An estimator of k^* defined in (9) is

$$\hat{k} = n^{\frac{2\hat{\rho}_X}{2\hat{\rho}_X+1}} \wedge m^{\frac{2\hat{\rho}_Y}{2\hat{\rho}_Y+1}}. \tag{11}$$

In the estimation problem of tail index with data $X_1, \ldots, X_n$, Hall consider distribution functions F which satisfy

$$1 - F(z) = C z^{-1/\gamma} \left(1 + D z^{-\rho/\gamma}(1 + R(z))\right), \tag{12}$$

where $\gamma, \rho > 0$, $C > 0$, $D \neq 0$ and R is tending to zero at infinity (γ is the parameter to be estimated). When the distribution function is assumed to belong to the Hall class defined in (12), Hall and Welsh (1985) prove that the method described below leads to an estimator $\hat{k}$ satisfying

$$\hat{k}/k^* = n^{\frac{2\hat{\rho}}{2\hat{\rho}+1}}/n^{\frac{2\rho}{2\rho+1}} \longrightarrow 1$$

in probability, as $p \to +\infty$.

Observe that there exists many other methods (more sophisticated) to estimate the parameter ρ. We choose the Hall and Welsh procedure because it is among the first one and we think that it is straightforward to understand. We propose in Sect. 5 an empirical study of the data driven procedure $D_{\hat{k}}$. Since the results are not so bad with the Hall and Welsh procedure, we do not investigate other methods to estimate ρ; obviously, the better is the procedure to estimate ρ, the better will be the applied results for our test methodology.

4 Simulations

The purpose of this section is to provide several examples to investigate the performances of the test procedure presented in Sect. 3.4. This part is not an illustration of the theoretical part. Instead to focus on the separating rate between the alternative and the null hypothesis, our aim is to study the test procedure on the risk point of view. We leave the theoretical framework by considering specific classes of distribution which are not necessary in the classes considered in Sect. 3.

4.1 Panel of distributions

Since most of tail parameters from economical data are between 1/3 and 1/5, we decide to restrict to indices γ laying from 0.2 to 1. However, the nuisance parameters ρ are free. First, we focus on a very simple class of distributions (called here Pareto–Hall distributions) containing distributions denoted P defined as follows

$$F(x) = 1 - C_{\gamma,\rho,\theta}\, x^{-1/\gamma}(1 + x^{-\rho/\gamma})1_{x>\theta}$$

for some γ, ρ, θ positive constant. By convention, the distribution P of parameters $(\gamma, \log)$ denote

$$F(x) = 1 - C_{\gamma,\rho,\theta}\, x^{-1/\gamma}(1 + \log(x))1_{x>\theta}.$$

Recall that these distributions do not belong to the Hall class. But these cases are very interesting because they appear as a limit case and they allow us to study the robustness of our procedure. Moreover, from a practical point of view, they are common in nonlife insurance mathematics. We tried various variations with respect to the Pareto–Hall distribution adding perturbations on the beginning of the support of the simulated data.

Since the empirical qualities of our procedure are the same for several perturbations, we limit the presentation with contaminations admitting the log normal distribution.

Then, as in Gomes and Martins (2002), we decide to study the following distributions (which belong to the Hall class).

– The Frechet distribution denoted F

$$F(x) = \exp\left(-x^{-1/\gamma}\right) 1_{x>0}$$

for which $(\gamma, \rho) = (1, 1)$.

– The Student distribution denoted t with $\nu = 4, 2, 1$ degrees of freedom, for which γ, ρ are $(0.25, 0.5)$, $(0.5, 1)$ and $(1, 2)$.

– the Burr distribution denoted B

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

Recall that our aim is to apply our methodology on financial series when the expected number of data is large. The procedure is obviously bad when p is very small (we try for $p \simeq 50$). Nevertheless, we decide to present results with smaller samples than what it is available for financial applications ($n = 700$, $m = 800$).

4.2 Tail index estimation

We focus here on the toy model (the Pareto–Hall distributions). First, we compute the adaptive estimator of γ using the Hall and Welsh standard procedure recalled in Sect. 3.4. We observe that $\hat{\rho}$ defined in (10) (and then $\hat{k}$ defined in (11)) is depending on the initialized values of σ, τ_1, τ_2. Without any specific assumptions, σ, τ_1, τ_2 are chosen randomly in accordance with the given constraints. For each considered γ, N replications are carry out with fixed parameters (ρ, n) but (randomly) changing parameters σ, τ_1, τ_2. In order to quantify the difference between $\hat{\gamma}(k^*)$ and $\hat{\gamma}(\hat{k})$, we compute the root mean square error (RMSE) defined as usual by

$$\mathrm{RMSE}(\hat{\gamma}(k^*), \hat{\gamma}(\hat{k})) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\gamma}_i(k^*) - \hat{\gamma}_i(\hat{k})\right)^2}.$$

Figure 1 (Left) shows the distribution of $\hat{\gamma}(\hat{k}) - \hat{\gamma}(k^*)$ where the Hill estimator is considered for different γ. For this example, the size of the $N = 500$ samples is $n = 800$; the samples are generated using a Pareto–Hall distribution with parameters $\gamma = 0.2, 0.4, 0.6, 0.8$ and $\rho = 1$. To estimate k^* by $\hat{k}$, the given interval for ρ is $[\rho 1, \rho 2] = [0.1, 3]$. Larger differences can be observed when γ is increasing: $\hat{\gamma}(\hat{k})$ can be also quite far from the optimal value $\hat{\gamma}(k^*)$. This is explained by the large variability of the estimations of ρ which could be observed when using the same sample, but with different initialized values σ, τ_1, τ_2. $\hat{\rho}$ could be quite far from the theoretical value, which induces strong variations in the computation of $\hat{\alpha}$.

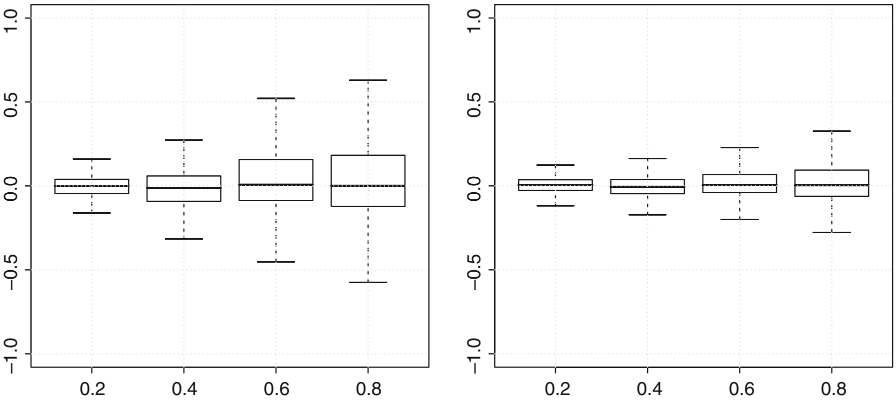

Fig. 1 *Left*: impact of Hall and Welsh initialized parameters on tail index variations. Distribution of $\hat{\gamma}(\hat{k}) - \hat{\gamma}(k^*)$, γ varying. *Right*: improvement of Hall and Welsh procedure on tail index variations. Distribution of $\hat{\gamma}(\tilde{k}) - \hat{\gamma}(k^*)$, γ varying

The RMSE (see Fig. 2 Left) increases linearly with γ and measures the differences induced by the initialization procedure. By instance, for $\gamma = 0.4$, we get RMSE $(\hat{\gamma}(k^*), \hat{\gamma}(\hat{k})) = 0.18$. Under the gaussian assumption and with a level of 95%, the corresponding confidence interval for the estimation values is [0.04, 0.78]. This interval seems to be large, especially for the parameters γ corresponding to financial data. In order to eliminate the impact of random numbers, we propose a practical improvement of the standard procedure to estimate k^*. This methodology is used for several simulating procedures when random initializations are needed (such as neural networks methods by instance). The procedure (denoted IHW for Improvement of the Hall and Welsh procedure) is iterative and computes the mean, denoted by $\tilde{k}$, of all unit estimations of k^*. Each unit estimation is computed with the Hall and Welsh standard procedure, starting from a new set of random conditions for σ, τ_1 and τ_2 parameters, in accordance with the given constraints.

$$\tilde{k} = \frac{1}{n_\rho} \sum_{i=1}^{n_\rho} \hat{k}_i,$$

where n_ρ is the number of iterations to estimate each underlaying $\hat{\rho}$. Figure 1 (Right) shows the distribution of $\hat{\gamma}(\tilde{k}) - \hat{\gamma}(k^*)$ when the Hill method is again considered. The chosen parameters for distribution are the same as before. The number of iterations for the Hall improved procedure is $n_\rho = 20$. Comparing Fig. 1 Right and Left, we observe that our improved method reduces significantly the distance between the estimators of γ built with $\tilde{k}$ and $\hat{k}$ and the estimator of γ built with the optimal index k^*. Figure 2 (Center) shows the RMSE computed between $\hat{\gamma}(\tilde{k})$ and $\hat{\gamma}(k^*)$: for any studied γ, RMSE$(\hat{\gamma}(\tilde{k}), \hat{\gamma}(k^*))$ is smaller than RMSE$(\hat{\gamma}(k^*), \hat{\gamma}(\hat{k}))$. Observe that the RMSE is reduced using the iterative method (solid line) compared to the standard method (dot line).

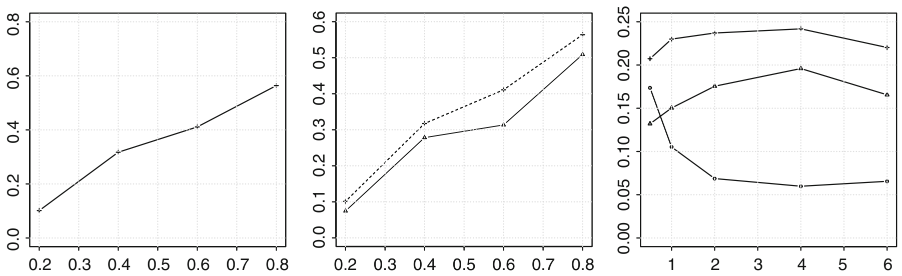

Fig. 2 *Left*: tail index variations due to initialized parameters of Hall and Welsh procedure. RMSE between Hill estimators for k^* and $\hat{k}$, γ varying. *Center*: Improvement of Hall and Welsh procedure on tail index variations. Comparison between the RMSE, ρ fixed, γ varying. Hall procedure (*dot line*); Hall improvement procedure (*regular line*). *Right*: Comparison of different tail index estimators. Comparison between the RMSE, γ fixed, ρ varying. k known ('o') ; k estimated by Hall procedure ('+'); k estimated by IHW ('Δ')

Similar results are observed for different values of $\rho > 0$. Figure 2 (Right) shows RMSE$(\gamma, \hat{\gamma}(\tilde{k}))$, RMSE$(\gamma, \hat{\gamma}(\hat{k}))$ and RMSE$(\gamma, \hat{\gamma}(k^*))$ computed for $\gamma = 0.4$ and various ρ. Obviously, the Hill estimator computed with k^* (γ, ρ known) leads to the smaller RMSE.

Table 1 quantifies the improvement of results for various γ, ρ values. The last column give the relative differences $I = (\text{RMSE}(\hat{k}) - \text{RMSE}(\tilde{k}))/\text{RMSE}(\hat{k})$. Different sizes of samples are studied: large samples ($n = 5{,}000$), medium samples ($n = 500$), and small samples ($n = 100$). In every case, the iterative procedure induces a significant decrease of the RMSE comparing to the standard Hall and Welsh procedure. The improvement is approximatively constant when γ is increasing. The size of the sample does not have any impact on the improvement. Notice that the nuisance parameter ρ does not perturb the results.

To conclude, we think that our iterative procedure is a good alternative to estimate the tail parameter γ without knowing any prior information on the nuisance parameter ρ. The only weak assumption is to assume than ρ belongs to a given interval, which can be chosen quite widely thanks to our improvement methodology.

4.3 Algorithm

Let us first describe the algorithm. The level $\alpha \in]0, 1[$, the number of data p and the parameters γ_Z, ρ_Z are fixed. In the case of the Pareto–Hall distribution, we choose $\theta_X = 1, \theta_Y = 2$. The main steps of our algorithm are the following:

1. Both samples are generated; in the case of the Pareto–Hall distributions, each sample Z^* of size p contains 20% of noisy observations generated using a log normal distribution ($\mu = 0.2, \sigma = 0.05$). Each data set Z^* is split randomly in two sub data sets of the same size: $Z^* = Z_1^* \oplus Z_2^*$.
2. Subsamples Z_1^* are used to compute $\tilde{k}_Z$ applying IHW (see the previous part). Then the index $\widehat{\gamma_Z}(\tilde{k}_Z)$ is computed leading to the statistic

$$T_{X_1^*, Y_1^*} = |\widehat{\gamma_{X_1^*}}(\tilde{k}_{X_1^*}) - \widehat{\gamma_{Y_1^*}}(\tilde{k}_{Y_1^*})|.$$

Table 1 Improvement of the RMSE for the Pareto–Hall distribution

n	γ	ρ/γ	RMSE($\hat{k}$)	RMSE($\tilde{k}$)	I (%)
100	0.2	1.0	0.092	0.899	2
–	0.4	1.0	0.148	0.11	23
–	0.6	1.0	0.259	0.199	23
–	0.8	1.0	0.377	0.260	31
500	0.2	1.0	0.084	0.075	11
–	0.4	1.0	0.174	0.125	28
–	0.6	1.0	0.257	0.202	21
–	0.8	1.0	0.392	0.283	27
5,000	0.2	1.0	0.09	0.070	21
–	0.4	1.0	0.252	0.207	18
–	0.6	1.0	0.356	0.297	16
–	0.8	1.0	0.486	0.410	15
100	0.2	2.0	0.093	0.067	28
–	0.4	2.0	0.162	0.107	34
–	0.6	2.0	0.315	0.260	17
–	0.8	2.0	0.448	0.348	22
500	0.2	2.0	0.07	0.05	23
–	0.4	2.0	0.186	0.142	23
–	0.6	2.0	0.320	0.263	18
–	0.8	2.0	0.435	0.357	18
5,000	0.2	2.0	0.104	0.088	16
–	0.4	2.0	0.239	0.201	15
–	0.6	2.0	0.345	0.277	19
–	0.8	2.0	0.363	0.317	12

3. Subsamples Z_2^* are used to estimate the variance of the statistics $T_{X_1^*,Y_1^*}$ using a bootstrap procedure with B replications: let us denote $\sigma_{X_2^*,Y_2^*}$ the square roots of the estimator. We deduce the critical value $t_{X_2^*,Y_2^*}(\alpha) = \sigma_{X_2^*,Y_2^*}/\alpha$.
4. Finally, the test statistic $D^{\alpha}_{X^*,Y^*}$ at the level α computed with the samples X^* and Y^* is given by:

$$D^{\alpha}_{X^*,Y^*} = \begin{cases} 0 \text{ if } |T_{X_1^*,Y_1^*}| \leq t_{X_2^*,Y_2^*}(\alpha) \\ \\ 1 \text{ if } |T_{X_1^*,Y_1^*}| > t_{X_2^*,Y_2^*}(\alpha). \end{cases} \tag{13}$$

5. The final result is $\hat{\alpha} = \overline{\mathcal{D}}_{X^*,Y^*}(\alpha)$ computed using the average of N decisions, N being the number of replications of the previous steps. When H_0 is true, $\hat{\alpha}$ is an estimation of the prescribed risk γ and when H_1 is true, $\hat{\alpha}$ is an estimation of the power of the test.

We made simulations for some smaller and larger values of α, and obtained similar results. So, we focus in the sequel on results for $\alpha = 0.1, 0.05$.

4.4 Performances of the test procedure

To estimate the first type error, couples of samples (X^*, Y^*) with the same theoretical tail parameter are generated ($\gamma_X = \gamma_Y$). To estimate the power of the test, the couple (X^*, Y^*) is generated with different tail parameters ($\gamma_X \neq \gamma_Y$). We take $n = 800$, $m = 700$, $N = 400$, $\rho_1 = 0.10$, $\rho_2 = 5.00$, $n_\rho = 20$. The constants for the bootstrap step are $B = 50$ replications with the rate 0.9.

The impact of the nuisance parameters ρ_X and ρ_Y is analyzed. Results on the first type error and power are presented.

4.4.1 First type error estimation

The estimations $\hat{\alpha}$ of the first type error are given in Table 2 when the nuisance parameters ρ are equal or different for both distributions X and Y.

The results are given in Table 2 and induce the following remarks

- We made extensive simulations for the toy model (the Pareto–Hall distributions) but give here a very short abstract. We observe that the results are excellent and that different values of the nuisance parameter (small or high) do not modify the estimations of the level. The behavior of the procedures associated to the Hill / Dekkers / de Vries estimators are similar. The results in the limit case are also excellent.
- The best results are obtained with the test methodology built with the Dekkers estimation procedure; in this case, they are globally good for any family of distributions.
- The level α is generally well fitted when the samples are providing of the same family of law. The test procedure can underestimate the prescribed level (see by instance when the samples are issue from the Frechet distribution).
- When the samples are issue from different families, the test methodology using the Hill estimator is the worst one. By instance, it gives bad result when the Student model is considered.
- Globally, it seems that the second order parameter has a small impact on the accuracy of the estimated level.

4.4.2 Power estimation

Results on the estimation of the power are given in Table 3.

- As expected, the estimated power is increasing with the difference between γ_X and γ_Y.
- In the case of the Pareto–Hall distributions, the rate of increasing of the power function is not very high, especially for the test procedure associated to the Dekkers estimator and in the limit cases.

Table 2 Empirical levels $\hat{\alpha}$

	$\gamma_X = \gamma_Y$	ρ_X	ρ_Y	Hill		Dekkers		Vries	
				$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$	$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$	$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$
PP	0.20	0.50	1.00	6	1	8	2	8	3
PP	0.30	0.50	1.00	8	2	8	4	10	3
PP	0.40	0.50	1.00	7	3	9	4	12	5
PP	0.50	0.50	1.00	6	2	8	3	7	3
PP	0.60	0.50	1.00	7	2	11	4	10	3
PP	0.70	0.50	1.00	9	4	9	4	9	4
PP	0.80	0.50	1.00	7	2	8	3	8	3
PP	0.20	log	1.00	5	1	7	2	7	2
PP	0.30	log	1.00	4	2	6	3	8	3
PP	0.40	log	1.00	7	1	9	4	8	4
PP	0.50	log	1.00	4	2	8	4	6	2
PP	0.60	log	1.00	7	3	12	5	9	4
PP	0.70	log	1.00	7	4	8	3	9	4
PP	0.80	log	1.00	8	4	10	4	9	3
tt	0.25	0.50	0.50	17	6	6	1	20	8
tt	0.50	1.00	1.00	25	11	7	1	13	5
tt	1.00	2.00	2.00	20	9	6	1	8	2
FF	0.25	1.00	1.00	8	1	6	2	6	2
FF	0.50	1.00	1.00	9	2	8	2	6	2
FF	1.00	1.00	1.00	10	3	8	3	9	2
BB	0.25	2.00	2.00	10	3	6	1	6	1
BB	0.25	1.00	2.00	28	14	8	1	15	6
BB	1.00	1.00	1.00	19	9	11	3	13	4
BB	1.00	1.00	2.00	29	14	14	5	14	7
Pt	1.00	0.50	0.50	32	15	13	7	26	11
Pt	1.00	0.50	1.00	37	17	13	4	22	10
PF	0.25	0.50	1.00	24	8	6	2	17	6
PF	0.50	0.50	1.00	26	8	5	2	20	7
PF	1.00	0.50	1.00	23	8	14	4	18	6
PB	0.25	0.50	1.00	52	28	6	4	34	16
PB	0.25	1.00	1.00	54	30	5	3	29	12
PB	1.00	1.00	1.00	49	26	23	10	29	12
PB	1.00	2.00	1.00	42	24	17	8	24	8
tF	0.50	1.00	1.00	57	33	5	2	26	12
tF	1.00	2.00	1.00	16	7	8	1	9	2
tB	0.50	1.00	0.50	36	16	13	4	28	10
tB	1.00	2.00	1.00	22	10	11	2	12	6
tB	1.00	2.00	2.00	20	8	10	2	8	3
FB	0.25	1.00	1.00	28	12	4	1	15	5

Table 2 continued

	$\gamma_X = \gamma_Y$	ρ_X	ρ_Y	Hill		Dekkers		Vries	
				$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$	$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$	$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$
FB	0.50	1.00	1.00	24	10	6	2	15	6
FB	1.00	1.00	1.00	26	14	11	4	15	6
FB	0.25	1.00	2.00	10	3	7	1	8	2
FB	0.50	1.00	2.00	9	2	5	1	9	4
FB	1.00	1.00	2.00	10	3	6	2	8	2

The prescribed levels are $\alpha = 10, 5\%$

- For the others distributions, the results are excellent when the Hill estimation procedure is considered. It is expected because the estimation of probability of the first type error is not very good in this case (see the previous part). The results obtained when the Vries method of estimation is used are similar.
- The Dekkers method of estimation produces tests with small power. When the distance between the tail indices is small, the test procedure has some difficulty to decide, see for instance the comparison between the $t(0.5, 1)$ and the $F(0.25, 1)$ leading to an estimated power 0.22 while this power is 0.99 with the Hill procedure and 0.98 with the Vries procedure. But when the difference between the indices is large, the results are very good; we observe that the second order parameter seems to have little influence. The Burr model and the Frechet model provide the best results.

4.4.3 Conclusion

Our test procedure provides accurate estimators of the first type error even in the limit case for our toy model and these estimators $\hat{\alpha}$ do not depend on the nuisance parameter. The test procedure admits similar behavior when the three estimation procedures are used. Remark that they are much better than in the nonparametric test setting (see for instance the simulations in Butucea and Tribouley 2006, where the entire densities are compared). The results on the power are less satisfying: as expected, the estimated power is increasing, but its growth towards one is very slow.

The results obtained in this empirical study confirm the interest of the test procedure for comparing unknown tail indices of various distribution functions even when the distribution functions do not belong to the same family of distributions.

5 A real life application: financial data

We apply the proposed test methodology on real economical data to compare the risk behavior of various financial series as companies shares, financial indexes, bonds or currencies, through their tail parameters. We consider a real-life example for which the data are not necessary independent. We emphasize that we do not want to test the validity of the model but decide H_0 or H_1.

Table 3 Empirical Power $\hat{\alpha}$

	$\gamma_X \neq \gamma_Y$	ρ_X	ρ_Y	Hill		Dekkers		Vries	
				$\hat{\alpha}$(10%)	$\hat{\alpha}$(5%)	$\hat{\alpha}$(10%)	$\hat{\alpha}$(5%)	$\hat{\alpha}$(10%)	$\hat{\alpha}$(5%)
PP	$0.50 \neq 0.20$	1.00	2.00	66	38	11	5	60	33
PP	$0.50 \neq 0.25$	1.00	2.00	48	19	9	4	44	20
PP	$0.50 \neq 0.30$	1.00	2.00	28	7	9	4	29	15
PP	$0.50 \neq 0.40$	1.00	2.00	10	2	11	4	14	6
PP	$0.50 = 0.50$	1.00	2.00	6	2	8	3	7	3
PP	$0.50 \neq 0.70$	1.00	2.00	22	9	10	4	17	6
PP	$0.50 \neq 1.00$	1.00	2.00	49	28	20	8	36	17
PP	$0.50 \neq 2.00$	1.00	2.00	91	75	76	57	82	63
PP	$0.50 \neq 3.00$	1.00	2.00	96	84	90	76	91	75
PP	$0.50 \neq 5.00$	1.00	2.00	96	88	95	84	92	81
PP	$0.50 \neq 10.00$	1.00	2.00	96	88	96	88	93	82
PP	$0.50 \neq 0.80$	log	2.00	9	4	10	3	10	4
PP	$0.50 \neq 1.00$	log	2.00	22	10	18	6	21	8
PP	$0.50 \neq 2.00$	log	2.00	58	42	56	38	56	37
PP	$0.50 \neq 3.00$	log	2.00	71	55	63	45	65	45
PP	$0.50 \neq 5.00$	log	2.00	76	57	75	59	72	53
tt	$0.25 \neq 0.50$	0.50	1.00	35	20	27	8	40	22
tt	$0.25 \neq 1.00$	0.50	2.00	82	62	92	70	91	74
tt	$0.50 \neq 1.00$	1.00	2.00	61	39	60	33	58	39
FF	$0.25 \neq 0.50$	1.00	1.00	85	61	20	9	68	48
FF	$0.25 \neq 1.00$	1.00	1.00	99	94	82	64	94	83
BB	$0.25 \neq 0.50$	0.50	0.50	94	82	54	25	95	80
BB	$0.25 \neq 1.00$	0.50	0.50	100	100	100	99	100	100
BB	$0.25 \neq 0.50$	1.00	0.50	100	98	63	38	99	98
BB	$0.25 \neq 1.00$	1.00	0.50	100	100	100	100	100	100
BB	$0.25 \neq 0.50$	1.00	2.00	67	40	34	15	64	39
BB	$0.25 \neq 1.00$	1.00	2.00	100	97	94	73	98	90
Pt	$0.25 \neq 0.50$	2.00	1.00	98	93	15	6	92	80
Pt	$0.25 \neq 1.00$	0.50	2.00	100	99	66	44	99	94
Pt	$0.50 \neq 0.25$	0.50	0.50	52	29	9	3	22	11
Pt	$0.50 \neq 1.00$	1.00	2.00	92	82	44	24	85	61
PF	$0.25 \neq 0.50$	0.50	1.00	89	69	16	6	77	54
PF	$0.25 \neq 0.50$	1.00	1.00	86	67	16	8	69	51
PF	$0.25 \neq 1.00$	0.50	1.00	99	96	59	37	97	86
PF	$0.25 \neq 1.00$	1.00	1.00	100	96	59	37	95	87
PF	$1.00 \neq 0.25$	0.50	1.00	58	36	34	20	53	33
PF	$1.00 \neq 0.25$	2.00	1.00	61	38	35	20	56	37
PF	$1.00 \neq 0.50$	0.50	1.00	24	10	14	5	20	10
PF	$1.00 \neq 0.50$	2.00	1.00	20	6	17	8	21	9

Table 3 continued

	$\gamma_X \neq \gamma_Y$	ρ_X	ρ_Y	Hill		Dekkers		Vries	
				$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$	$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$	$\hat{\alpha}(10\%)$	$\hat{\alpha}(5\%)$
PB	$0.25 \neq 1.00$	0.50	0.50	100	100	86	73	100	100
PB	$0.25 \neq 1.00$	0.50	1.00	100	100	72	52	100	97
PB	$0.25 \neq 1.00$	0.50	2.00	100	99	66	45	97	92
tF	$0.25 \neq 1.00$	0.50	1.00	68	45	82	61	72	52
tF	$0.50 \neq 0.25$	1.00	1.00	99	98	22	8	98	91
tF	$1.00 \neq 0.25$	2.00	1.00	100	100	91	71	99	93
tF	$1.00 \neq 0.50$	2.00	1.00	95	82	61	31	83	62

The prescribed levels are $\alpha = 10, 5\%$

The test procedures are applied on a panel of different financial data. Various types of series are chosen, as single share, indexes, financial bonds, raw material prices, and currencies. Financial indexes are selected as an indicator of stock market prices, based on the value of a set of shares belonging to France (*CAC40*), US (*dowjones*), New York (*sandP500*), UK (*ftse100uk*) and Japan (*nikkei255*). The single shares are Air Liquide (*AL*), General Electric (*GE*), fordmotor and exon mobil. The bonds are taken in US over 10 years (*US10yr*), 30 years (*US30yr*) and in Germany (*Germany*). The raw material price is *brent*. The exchange rates are *yen to dollar* and *deutchmark to dollar*. All these data correspond to daily closing market quotations.

For each time series, the log return sequences are computed and the tail parameters are estimated on the positive log return series. Tail indexes are estimated using Hill estimator associated to IHW (see Part 4.2) to estimate the optimal index k^*. For the initialization procedure, the given interval is $[\rho_1, \rho_2] = [0.1, 5]$. The number of replications is $n_\rho = 30$. Remark that we do not discuss about whether the series belong to the functional class: our aim is to decide if H_0 is more suitable than H_1 or not. To test whether the tail condition is not satisfied, see Drees et al. (2006).

Figure 3 shows the computed tail parameters estimations for these series. All these financial data exhibit heavy tail features with γ values from 0.3 to 0.7. Financial indexes (*cac, ftse*) and single shares exhibit the highest tail parameters with values around 0.7. Financial bonds (*US10yr, US30yr, Germany*) exhibit the lowest tail parameters with values around 0.3. All these results confirm the knowledge about these financial time series.

The algorithm described in page 15 is applied on all couples of financial series to compare their financial risk quantified by the tail index of the distribution of the log return. Recall that both samples are split randomly in two subsamples to compute independently the test value and its statistical variance. In order to eliminate the potential influence of this random procedure, an average decision $\overline{D^\alpha_{X^*,Y^*}}$ is computed. Observe that $\overline{D^\alpha_{X^*,Y^*}} > 1/2$ means that H_1 is accepted more often that H_0 (and conversely if $\overline{D^\alpha_{X^*,Y^*}} < 1/2$). So the final decision is taken as follows:

– low values of $\overline{D^\alpha_{X^*,Y^*}}$ means that both distributions have same tail parameter and then we accept that the risk of both series are the same,

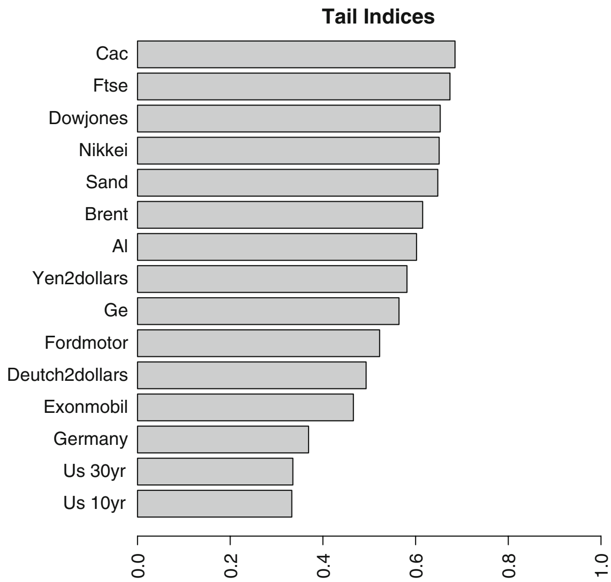

Fig. 3 Estimation of the tail indices for economical data

– high values of $\overline{D^{\alpha}_{X*,Y*}}$ means that the distribution have different tail parameters and we accept that one of the series have a smaller risk than the other one.

Figure 4 shows the average decisions for the bilateral test procedure computing for the studied financial series. Each subgraph presents $\overline{D^{\alpha}_{X*,Y*}}$ for $\alpha = 5\%$ when the series X mentioned in the title is tested against all the others series Y mentioned in the columns of each subgraph. The vertical dotted line is the level $1/2$: if $\overline{D^{\alpha}_{X*,Y*}}$ exceeds this line, we accept that $\gamma_X \neq \gamma_Y$ which means that X and Y do not have the same behavior with respect to the risk.

For instance, considering the last subgraph, we conclude that the behavior with respect to the risk of the *CAC* is not the same as the behavior of the bonds (*US and German*), of the share series *fordmotor, exonmobil* and of the currency series *DM/Dollar*. Observe that this conclusion is identical for the others indices *ftse, dowjones, nikkei, sand* and in a much curious way for the currency *Yen/Dollar* and the commodity *brent*. It is interesting to observe that the tail parameters of the series *yen/Dollar* and the series *DM/dollar* are accepted to be different. Moreover, the subgraph corresponding to *DM/Dollar* shows that the tail parameter corresponding is accepted to be different of the tail parameters of the indices series (*CAC, ftse, dowjones, nikkei, sand*) and *brent, AL* and also *Yen/Dollar*.

To analyze more precisely the difference of the tail parameters, we compare these parameters in an unilateral way. Figure 5 shows the average decision for the unilateral test to compare the financial risks of the series at level 5%. In this case, the procedure test is $H_0 : \gamma_X = \gamma_Y$ against $H_1 : \gamma_X > \gamma_Y$ where X is mentioned in the title of each subgraph and where Y is mentioned in the columns of each subgraph. If the quantity

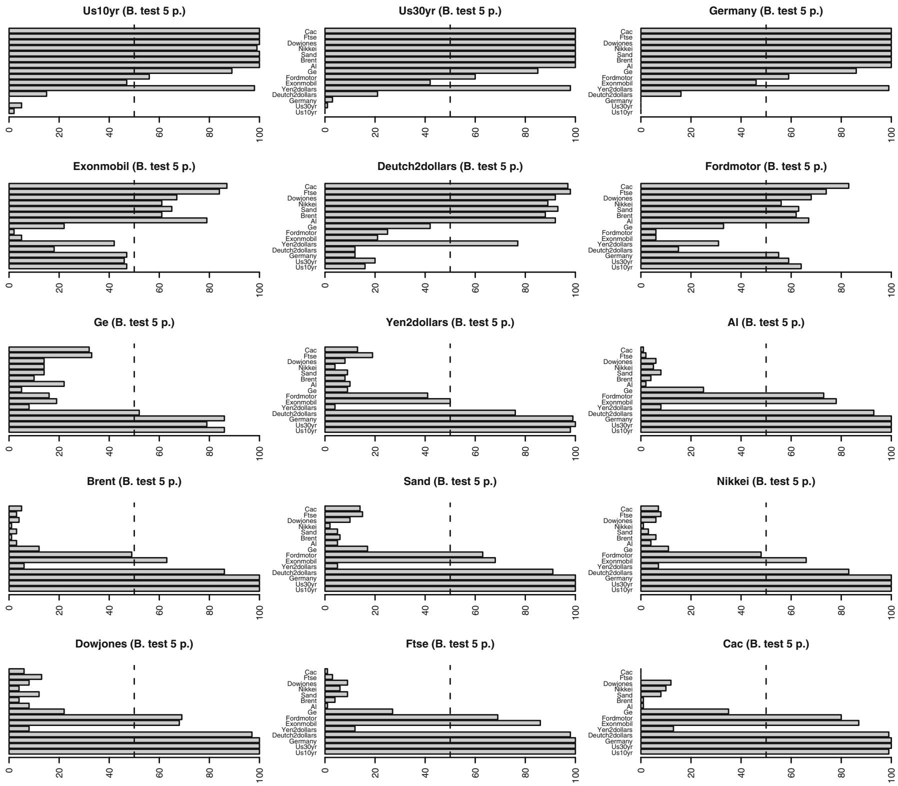

Fig. 4 Comparison of tail indices for economical data. Bilateral Test ($H_1 : \gamma_X \neq \gamma_Y$). Level 5%

$\overline{D^{\alpha}_{X*,Y*}}$ (represented by a horizontal stick) exceeds 0.5 (represented by the vertical dotted line), we conclude that the series X is more risked than the series Y.

We have four clusters of series.

$C_1 = \{Cac, Ftse, DowJones, Nikkei, Sand, AL\}$

$C_2 = \{Ge, Yen/Dollar, Brent\}$

$C_3 = \{Us10, US30, Germany, DM/Dollar\}$

$C_4 = \{ExonMobil, Fordmotor\}$

Let us denote by γ_i the tail indices of the series belonging to the cluster C_i for $i = 1, 2, 3, 4$. The first one contains all financial indices; Figure 5 shows that we accept that the γ_1's are larger than the γ_3's. Remark that C_3 contains the bonds. γ_2 is greater than the tail parameter of the bonds (see Fig. 5). Last, the series of C_4 have tail parameters different of the tail parameter of the indices when we use the bilateral test at level 5% but this difference is no more significative when we use unilateral tests at level 5% each. The financial indices (X) from France (*cac*), United Kingdom (*ftse*), US (*dowjones*), Japan (*nikkei*), have the higher tail parameters compared to all financial bonds (Y: *Us10yr, Us30yr, Germany*), with an average decision close to 100. The test computes also that the (*brent*) parameter is also one of the heaviest. *AL* shows a heavy tail compared to other single shares (it General electrics, fordmotor and exon mobil). For currencies, *yen to dollar* shows the heaviest tail compared to *DM/dollars*.

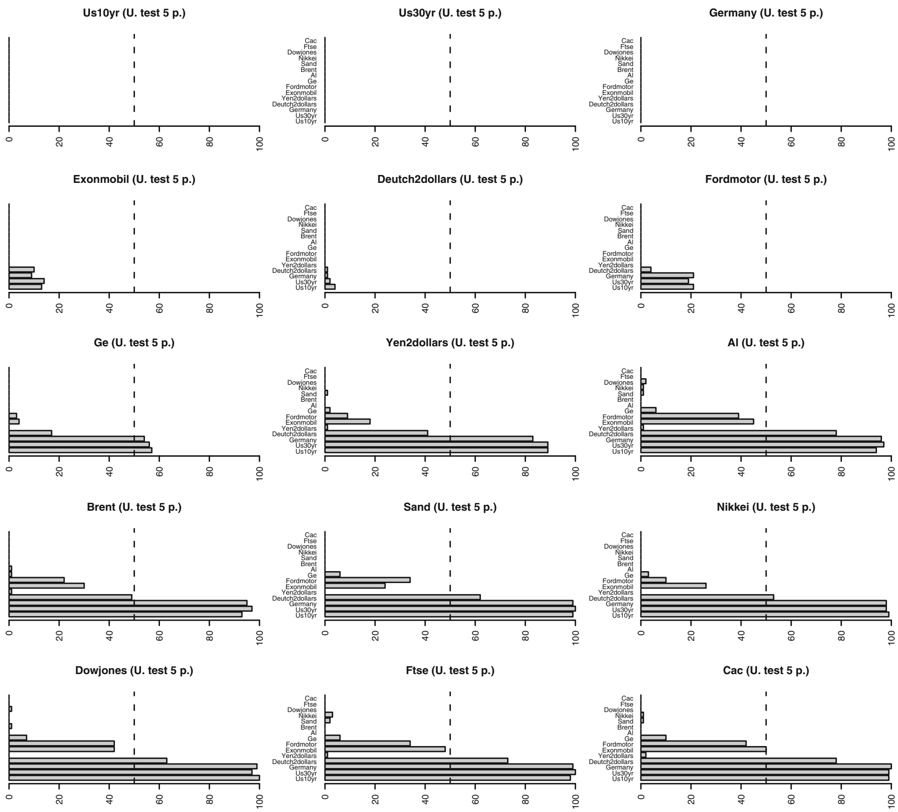

Fig. 5 Comparison of tail indices for economical data. Unilateral Test ($H_1 : \gamma_X > \gamma_Y$). Level 5%

Remark that, we never accept that the bonds and the currency *DM/ dollar* are more risked than any other series. The series *exonmobil* and *fordmotor* are never accepted to be more risked than any other series. Note also that our test does not allow us to separate between the indices series set (or the bonds series set). The bonds in US over 10 years (*US10yr*), 30 years (*US30yr*) and in Germany (*Germany*) show the lowest tail indexes. For currencies, the *deutchmark/dollars* shows the lowest tail.

The test procedure applied on this set of real financial series confirms objectively the knowledge of most financial experts. It reinforces also the strength of the test procedure and of its associated methodology to compare the risk behavior on both simulated and real data.

6 Proofs

6.1 Preliminaries

We first state a theorem giving an exact expansion of the estimator $\hat{\gamma}$. This theorem is given in de Haan and Peng (1998) with less precision in the control of the terms which are tending to zero in probability; the proof is given for the Hill's estimator $\hat{\gamma}^1$ and is rejected at the end of this part.

Theorem 2 *Let $0 < \rho_1 < \rho_2$. Let $Z_1, \ldots, Z_p$ be i.i.d. variables associated to the variable Z. Assume that the distribution function F_Z of Z satisfies the second order condition (4) for*

- *some γ_Z and ρ_Z such that $\rho_Z \in [\rho_1, \rho_2]$,*
- *a function A_Z belonging to $RV_{-\gamma_Z \rho_Z}(\tilde{a}_Z)$ for a function $\tilde{a}_Z$,*
- *a function R_Z such that $\sup_{z \geq 1} |R(z,t)| \leq \tilde{R}_Z(t)$ for a function $\tilde{R}_Z$ belonging to $RV_\tau(\tilde{r}_Z)$ for some τ and some function $\tilde{r}_Z$.*

Then there exists a sequence $(P_p(k))$ which is equal in distribution to

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k}(E_i - 1)\right), \tag{14}$$

where $E_1, \ldots, E_p$ are i.i.d. from a standard exponential distribution such that

$$\hat{\gamma}_Z(k) \overset{d}{=} \gamma_Z + e_X \frac{P_p(k)}{\sqrt{k}} + d_Z A_Z\left(\frac{p}{k}\right)(1 + W_Z(k))$$

for

$$e = \gamma_Z, \quad d_Z = \tfrac{1}{1+\rho_Z}$$

and where, for any $u > 0$,

$$P(|W_Z| > u) \leq C_1 \left(\frac{1}{k} + A_Z\left(\frac{p}{k}\right) + \frac{1}{k}\tilde{R}_Z^2\left(\frac{p}{k}\right) \right.$$
$$\left. + 1_{\{\tilde{R}_Z(\frac{p}{k}) \vee \tilde{R}_Z(\frac{p}{k})\tilde{r}_Z(\frac{p}{k}) \vee \tilde{a}_Z(\frac{p}{k}) \vee \tilde{a}_Z^2(\frac{p}{k}) > C_2\}}\right). \tag{15}$$

C_1, C_2 are positive constants depending on ρ_1, ρ_2, τ and u.

6.2 Proof of Theorem 1

Using the expansion given in Theorem 2, we get

$$\hat{\gamma}_X(k) - \hat{\gamma}_Y(k') = (\gamma_X - \gamma_Y) + \frac{1}{\sqrt{k \wedge k'}} R_{n,m}(k,k') + b(k,k') + \Omega(k) - \Omega(k'),$$

where the random term is

$$R_{n,m}(k,k') = \left(e_X \frac{P_n(k)}{\sqrt{\frac{k}{k \wedge k'}}} - e_Y \frac{P_m(k')}{\sqrt{\frac{k'}{k \wedge k'}}} \right)$$

and the bias terms are

$$b(k,k') = \left[d_X A_X\left(\frac{n}{k}\right) - d_Y A_Y\left(\frac{m}{k'}\right) \right]$$

and

$$\Omega(k) = d_X A_X\left(\frac{n}{k}\right) W_X, \ \ \Omega(k') = d_Y A_Y\left(\frac{m}{k'}\right) W_Y, \tag{16}$$

where $W_X(k)$, $W_Y(k')$ verifies (15). Using (14), we obviously have

$$E R_{n,m}(k,k') = 0 \quad \text{ and } \quad E R^2_{n,m}(k,k') \le 2(e_X^2 + e_Y^2) = \upsilon_{X,Y}.$$

Let us study the error of the first type and determine the critical value. Under H_0, we have $\gamma_X - \gamma_Y = 0$ and

$$\begin{aligned} P_{\Sigma_0}\left(D_{k,k'} = 1\right) &= P_{\Sigma_0}\left(|T_{k,k'}| \ge t_{k,k'} \right) \\ &\le P_{\Sigma_0}\left(\left| \frac{1}{\sqrt{k \wedge k'}} R_{n,m}(k,k') + \ b(k,k') \right| \ge \ t_{k,k'}/2 \right) \\ &\quad + P_{\Sigma_0}\left(|\Omega(k)| \ge \ t_{k,k'}/4 \right) + P_{\Sigma_0}\left(\left|\Omega(k')\right| \ge \ t_{k,k'}/4 \right) \\ &= p_1 + p_2 + p_3. \end{aligned}$$

Using Markov Inequality, we bound the first term

$$\begin{aligned} p_1 &\le \frac{8\left[E R^2_{n,m}(k,k') + (k \wedge k') b^2(k,k') \right]}{(k \wedge k')\, t^2_{k,k'}} \\ &\le 8\left(\upsilon_{X,Y} + d_X^2 k A_X^2\left(\frac{n}{k}\right) + d_Y^2 k' A_Y^2\left(\frac{m}{k'}\right) \right) \left((k \wedge k')\, t^2_{k,k'} \right)^{-1}. \end{aligned}$$

Fix $\lambda > 0$. Choosing the indices $k(\lambda)$ and $k'(\lambda)$ such that

$$k(\lambda) A_X^2\left(\frac{n}{k(\lambda)}\right) = \lambda \quad \text{and} \quad k'(\lambda) A_Y^2\left(\frac{m}{k'(\lambda)}\right) = \lambda \tag{17}$$

and the critical value

$$t_{k(\lambda),k'(\lambda)} = 4\sqrt{\frac{\upsilon_{X,Y} + (d_X^2 + d_Y^2)\lambda}{\alpha}}\, (k(\lambda) \wedge k'(\lambda))^{-1/2}, \tag{18}$$

we bound p_1 by $\alpha/2$. Using (17) and (16), we obtain

$$p_2 \le P_{\Sigma_0}\left(|\Omega(k(\lambda))| \ge \sqrt{\frac{d_X^2}{\alpha}}\sqrt{\frac{\lambda}{k(\lambda)}}\right)$$
$$= P_{\Sigma_0}\left(|W_X(k(\lambda))| \ge \frac{1}{\sqrt{\alpha}}\right).$$

Using (15), since $\log(n) \le k(\lambda) \le n(\log n)^{-1}$

$$p_2 \le C_1\left(\frac{\lambda}{\sqrt{\log(n)}} + 1_{\{\tilde{R}_0(\log n)\vee\tilde{a}_0(\log n)>C_2\}}\right) = C_3\ o(1)$$

(as soon as n is large enough) where C_1, C_2, C_3 are positive constants depending on $\rho_1, \rho_2, \tau, \alpha$ and on the functions $\tilde{R}_0, \tilde{a}_0, \tilde{r}_0$. Similarly, we bound p_3 by $C_3\ o(1)$. We deduce that the risk of the first type is bounded by $\alpha/2 + o(1)$. Let us study the second type error term. Let us denote the optimal index $k^* = k(\lambda) \wedge k'(\lambda)$ where $k(\lambda), k'(\lambda)$ are defined in (17) and t_{k^*} the optimal critical value defined in (18). Let us denote

$$B = b(k(\lambda), k'(\lambda)) + \Omega(k(\lambda)) + \Omega(k'(\lambda)).$$

We get, setting $R_{n,m}(k(\lambda), k'(\lambda)) = R_{n,m}$,

$$\begin{aligned}
&P_{F_X,F_Y}(D_{k(\lambda),k'(\lambda)} = 0) = P_{F_X,F_Y}(|T_{k(\lambda),k'(\lambda)}| \le t_{k^*})\\
&\quad\le P_{F_X,F_Y}\Big(\sqrt{k^*}\left[-t_{k^*} - |B| - (\gamma_X - \gamma_Y)\right] \le R_{n,m}\\
&\quad\le \sqrt{k^*}\left[t_{k^*} + |B| - (\gamma_X - \gamma_Y)\right]\Big)\\
&\quad\le P_{F_X,F_Y}\Big(\sqrt{k^*}\left[-t_{k^*} - |B| - (\gamma_X - \gamma_Y)\right] \le R_{n,m} \text{ and } -\gamma_X + \gamma_Y \ge \phi_{n,m}\Big)\\
&\qquad+P_{F_X,F_Y}\Big(R_{n,m} \le \sqrt{k^*}\left[t_{k^*} + |B| - (\gamma_X - \gamma_Y)\right] \text{ and } \gamma_X - \gamma_Y \ge \phi_{n,m}\Big)\\
&\quad\le 2\ P_{F_X,F_Y}\left(|R_{n,m}| + \sqrt{k^*}|B| \ge \phi_{n,m}\sqrt{k^*}\left[1 - \frac{3t_{k^*}}{2\phi_{n,m}}\right]\right).
\end{aligned}$$

Let us consider rates satisfying

$$\frac{3t_{k^*(\lambda)}}{2\phi_{n,m}} \le 1/2.$$

We finish as for the error of the first type:

$$P_{F_X,F_Y}(D_{k(\lambda),k'(\lambda)} = 0) \le p_1 + p_2 + p_3,$$

where

$$p_1 = 2\,P_{F_X,F_Y}\left(|R_{n,m}(k(\lambda),k'(\lambda))| + \sqrt{k^*}|b(k(\lambda),k'(\lambda)| \geq \phi_{n,m}\sqrt{k^*}/2\right)$$
$$\leq 16\,\frac{\upsilon_{X,Y} + \lambda(d_X^2 + d_Y^2)}{k^*\,\phi_{n,m}^2}$$

and

$$p_2 + p_3 = 2\,P_{F_X,F_Y}\left(|\Omega(k(\lambda))| \geq \phi_{n,m}/4\right) + 2\,P_{F_X,F_Y}\left(|\Omega(k'(\lambda))| \geq \phi_{n,m}/4\right).$$

We deduce that the second type error is bounded by $\alpha/2+o(1)$ for rates $\phi_{n,m}$ satisfying

$$\phi_{n,m} \geq 3t_{k^*} \vee 8\sqrt{\frac{\upsilon_{X,Y} + \lambda(d_X^2 + d_Y^2)}{\alpha\,k^*}},$$

which leads to the announced result.

6.3 Proof of Theorem 2

First we state the following lemma proven at the end of this section

Lemma 1 *Let $G(\cdot)$ be a function in $RV_\tau(\tilde{g})$ for some τ and $\tilde{g}(\cdot)$. Then, for any $u > 0$*

$$P\left(\left|G\left(1/\xi_{(k:p)}\right) - G\left(\frac{p}{k}\right)\right| > u\right) \leq \frac{c}{k}\left(1 + (2^{-\tau}\tau)^2\frac{G^2\left(\frac{p}{k}\right)}{u^2}\right)$$
$$+1_{\{\tilde{g}(\frac{p}{k})G(\frac{p}{k}) > u(2^\tau \wedge (3/2)^\tau)\}},$$

where c is an universal constant.

Next, note that $Z_{(p-k:p)} =^d U(1/\xi_{(k:p)})$ where $\xi_1,\ldots,\xi_p$ are independent uniform random variables. Put $F_i = \xi_{(k:p)}/\xi_{(i:p)}$ and denote

$$\begin{cases} d^1 = E\left(\frac{1-(F_i)^{-\rho}}{\rho}\right) = \frac{1}{1+\rho}, \\ g^1 = V\left(\frac{1-(F_i)^{-\rho}}{\rho}\right) = \frac{1}{\rho^2}\left[1 + \frac{1}{1+2\rho} - \frac{2}{1+\rho}\right]. \end{cases}$$

The second order condition (4) implies that

$$\forall x > 0, \forall t > 0, \quad \log(U(tx)) - \log(U(t)) = \gamma\log(x) + S(x,t) + \tilde{S}(x,t),$$

where

$$\begin{cases} S(x,t) = A(t)\left(\frac{1-x^{-\rho}}{\rho}\right)(1 + R(x,t)), \\ \tilde{S}(x,t) = \log(1 + S(x,t)) - S(x,t) \leq S^2(x,t). \end{cases}$$

Taking $t = 1/\xi_{(k:p)}$ and $x = F_i = \xi_{(k:p)}/\xi_{(i:p)}$, we obtain

$$\hat{\gamma}^1 = \gamma + \gamma \frac{1}{k}\sum_{i=1}^{k}\left[\log(F_i) - 1\right] + \frac{1}{k}\sum_{i=1}^{k}\left[S(F_i, 1/\xi_{(k:p)})\right] + \frac{1}{k}\sum_{i=1}^{k}\left[\tilde{S}(F_i, 1/\xi_{(k:p)})\right].$$

Remark that $F_i \geq 1$ and then $|R(F_i, 1/\xi_{(k:p)})| \leq \tilde{R}_0(1/\xi_{(k:p)})$. Writing

$$\begin{aligned} S(F_i, 1/\xi_{(k:p)}) &= A(1/\xi_{(k:p)})\left(\frac{1-(F_i)^{-\rho}}{\rho}\right)(1 + R(F_i, 1/\xi_{(k:p)})) \\ &= \Bigg\{A(p/k)E\left(\frac{1-(F_i)^{-\rho}}{\rho}\right) \\ &\quad + \left[A(1/\xi_{(k:p)}) - A(p/k)\right]\left[\left(\frac{1-(F_i)^{-\rho}}{\rho}\right) - E\left(\frac{1-(F_i)^{-\rho}}{\rho}\right)\right] \\ &\quad + \left[A(1/\xi_{(k:p)}) - A(p/k)\right]E\left(\frac{1-(F_i)^{-\rho}}{\rho}\right) \\ &\quad + A(p/k)\left[\left(\frac{1-(F_i)^{-\rho}}{\rho}\right) - E\left(\frac{1-(F_i)^{-\rho}}{\rho}\right)\right]\Bigg\} \\ &\quad \times(1 + R(F_i, 1/\xi_{(k:p)})), \end{aligned}$$

it follows

$$\begin{aligned} \frac{1}{k}\sum_{i=1}^{k}\left[S(F_i, 1/\xi_{(k:p)})\right] &= d^1\, A(p/k)(1 + W_3) + W_1 W_2 \\ &\quad + d^1 W_1(1 + W_3) + A(p/k)W_2 \\ \left|\frac{1}{k}\sum_{i=1}^{k}\left[\tilde{S}(F_i, 1/\xi_{(k:p)})\right]\right| &\leq \frac{1}{k}\sum_{i=1}^{k}\left[S^2(F_i, 1/\xi_{(k:p)})\right] \\ &\leq 2A(p/k)^2(1 + W_6^2)W_4 + 2(1 + W_6^2)W_4 W_5 \end{aligned}$$

for

$$\begin{aligned} W_1 &= A(1/\xi_{(k:p)}) - A(p/k), \\ W_5 &= A(1/\xi_{(k:p)})^2 - A(p/k)^2, \\ W_6 &= \tilde{R}_0(1/\xi_{(k:p)}), \\ |W_2| &= \frac{1}{k}\sum_{i=1}^{k}\left\{\left[\left(\frac{1-(F_i)^{-\rho}}{\rho}\right) - E\left(\frac{1-(F_i)^{-\rho}}{\rho}\right)\right]\left[1 + R(F_i, 1/\xi_{(k:p)})\right]\right\} \end{aligned}$$

$$\leq \left[1+\tilde{R}_0(1/\xi_{(k:p)})\right] \left|\frac{1}{k}\sum_{i=1}^{k}\left[\left(\frac{1-(F_i)^{-\rho}}{\rho}\right)-E\left(\frac{1-(F_i)^{-\rho}}{\rho}\right)\right]\right|,$$

$$W_4 = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{1-(F_i)^{-\rho}}{\rho}\right),$$

$$|W_3| = \left|\frac{1}{k}\sum_{i=1}^{k} R(F_i, 1/\xi_{(k:p)})\right| \leq W_6.$$

For any $t > 0$, we get

$$P(|W_4| \geq t) \leq \frac{g^1}{t}. \tag{19}$$

Applying Lemma 1 for $G := \tilde{R}$, we have

$$\begin{aligned} P(|W_3| \geq t) \vee P(W_6^2 \geq t) &\leq P\left(\tilde{R}_0(1/\xi_{(k:p)}) - \tilde{R}_0(p/k) > (t \wedge \sqrt{t})/2\right) \\ &\leq \frac{c}{k}\left(1+(2^{-\tau}\tau)^2\frac{\tilde{R}_0^2\left(\frac{p}{k}\right)}{(t^2\wedge t)}\right) \\ &\quad +1_{\left\{\tilde{r}_0\left(\frac{p}{k}\right)\tilde{R}_0\left(\frac{p}{k}\right)>(t\wedge\sqrt{t})(2^\tau\wedge(3/2)^\tau)/2\right\}} \end{aligned} \tag{20}$$

as soon as $t > 2\tilde{R}(p/k) \vee 4\tilde{R}^2(p/k)$. Similarly, for any $t \geq 4\tilde{R}^2(p/k)$,

$$\begin{aligned} P(|W_2| \geq 2t) &\leq 2P\left(\left|\frac{1}{k}\sum_{i=1}^{k}\left(\frac{1-(F_i)^{-\rho}}{\rho}\right)-E\left(\frac{1-(F_i)^{-\rho}}{\rho}\right)\right| > (t\wedge\sqrt{t})\right) \\ &\quad +P\left(\tilde{R}_0(1/\xi_{(k:p)}) - \tilde{R}_0(p/k) > \sqrt{t}/2\right) \\ &\leq g^1\frac{1}{k(t^2\wedge t)} + \frac{c}{k}\left(1+(2^{-\tau}\tau)^2\frac{\tilde{R}_0^2\left(\frac{p}{k}\right)}{t}\right) \\ &\quad +1_{\left\{\tilde{r}_0\left(\frac{p}{k}\right)\tilde{R}_0\left(\frac{p}{k}\right)>\sqrt{t}(2^\tau\wedge(3/2)^\tau)/2\right\}}. \end{aligned} \tag{21}$$

Using again Lemma 1 for $G := A$ and $G := A^2$, we get

$$P(|W_1| \geq t) \leq \frac{c}{k}\left(1+(2^\rho\rho)^2\frac{A^2\left(\frac{p}{k}\right)}{t^2}\right) + 1_{\{\tilde{a}_0\left(\frac{p}{k}\right)A\left(\frac{p}{k}\right)>t(3/2)^{-\rho}/2\}} \tag{22}$$

$$\begin{aligned} P(|W_5| \geq t) &\leq \frac{c}{k}\left(1+(2^{2\rho}\rho)^2\frac{A^4\left(\frac{p}{k}\right)}{t^2}\right) \\ &\quad +1_{\{\tilde{a}_0\left(\frac{p}{k}\right)(2+\tilde{a}_0\left(\frac{p}{k}\right))A^2\left(\frac{p}{k}\right)>t(3/2)^{-2\rho}/2\}}. \end{aligned} \tag{23}$$

It follows that

$$\hat{\gamma}^1 = \gamma + \gamma \ \frac{P_p(k)}{\sqrt{k}} + d^1 A(p/k)(1 + \ W),$$

where

$$|W| \le |W_3 + (A(p/k)d^1)^{-1} W_1 W_2 + A^{-1}(p/k) W_1 (1 + W_3) + (d^1)^{-1} W_2| \\ +2|(d^1)^{-1} A(p/k)(1 + W_6^2) W_4 + (A(p/k)d^1)^{-1}(1 + W_6^2) W_4 W_5|$$

satisfies, for any $t > 0$

$$\begin{aligned} P(|W| > 9t) \le\ & P(|W_3| > t) + P(|W_1 W_2| > d^1 A(p/k)t) + P(|W_1| > A(p/k)t) \\ & +P(|W_1 W_3)| > A(p/k)t) + P(|W_2| > d^1 t) \\ & +P(|W_4| > d^1 A(p/k)^{-1} t/2) \\ & +P(|W_4 W_6^2| > d^1 A(p/k)^{-1} t/2) + P(|W_4 W_5| > d^1 A(p/k)t/2) \\ & +P(|W_4 W_5 W_6^2| > d^1 A(p/k)t/2) \\ \le\ & 3P(|W_1| > A(p/k)(t \wedge 1)) + 2P(|W_3| > t) + 2P(|W_2| > d^1 t) \\ & +4P(|W_4| > d^1 A(p/k)^{-1} t/2) + 2P(|W_5| > A(p/k)^2) \\ & +2P(|W_6| > 1) \\ \le\ & C_1 \left(\frac{1}{k} + A\left(\frac{p}{k}\right) + \frac{1}{k} \tilde{R}_0^2 \left(\frac{p}{k}\right) \right. \\ & \left. +1_{\{\tilde{R}_0(\frac{p}{k}) \vee \tilde{R}_0(\frac{p}{k}) \tilde{r}_0(\frac{p}{k}) \vee \tilde{a}_0(\frac{p}{k}) \vee \tilde{a}_0^2(\frac{p}{k}) > C_2\}} \right), \end{aligned}$$

where C_1, C_2 are positive constants depending on ρ_1, ρ_2, τ and t. This ends the proof.

6.4 Proof of Lemma 1

Let us recall the following inequality (see by instance Shorack and Wellner, p 723).

Lemma 2 *Let $\xi_1, \ldots, \xi_p$ be independent variable which are uniform on* $[0, 1]$. *Then, there exists some universal constant $c > 0$ such that, for any $k = 1, \ldots, p$*

$$E\left(\xi_{k:p} - \frac{k}{p}\right)^2 \le c\frac{k}{p^2}.$$

For any $t > 0$, we get

$$\begin{aligned} p &= P\left(\left|G\left(1/\xi_{(k:p)}\right) - G\left(\frac{p}{k}\right)\right| > t\right) \\ &= P\left(\left|\frac{G\left(\frac{(k/p)}{\xi_{(k:p)}} * (p/k)\right)}{G\left(\frac{p}{k}\right)} - 1\right| > \frac{t}{G\left(\frac{p}{k}\right)}\right). \end{aligned}$$

Since $G(\cdot) \in RV_\tau(\tilde{g})$, there exists g such that

$$\forall x > 0, \forall t > 0, \frac{G(tx)}{G(t)} = x^\tau(1 + g(x,t)) \text{ with } \sup_{x \in \nu(1)} |g(x,t)| \leq \tilde{g}(t).$$

It follows

$$\begin{aligned}
p \leq\ & P\left(\left|\left(\frac{(k/p)}{\xi_{(k:p)}}\right)^\tau - 1\right| > \frac{t}{2G\left(\frac{p}{k}\right)}\right) \\
& + P\left(\left(\frac{(k/p)}{\xi_{(k:p)}}\right)^\tau \left|g\left(\frac{(k/p)}{\xi_{(k:p)}}, \frac{p}{k}\right)\right| > \frac{t}{2G\left(\frac{p}{k}\right)}\right) \\
\leq\ & P\left(\left|\left(\frac{(k/p)}{\xi_{(k:p)}}\right)^\tau - 1\right| > \frac{t}{2G\left(\frac{p}{k}\right)} \text{ and } \left(\frac{(k/p)}{\xi_{(k:p)}}\right) \in [1/2, 3/2]\right) \\
& + P\left(\left(\frac{(k/p)}{\xi_{(k:p)}}\right)^\tau \left|g\left(\frac{(k/p)}{\xi_{(k:p)}}, \frac{p}{k}\right)\right| > \frac{t}{2G\left(\frac{p}{k}\right)} \text{ and } \left(\frac{(k/p)}{\xi_{(k:p)}}\right) \in [1/2, 3/2]\right) \\
& + 2P\left(\left(\frac{(k/p)}{\xi_{(k:p)}}\right) < 1/2 \text{ or } \left(\frac{(k/p)}{\xi_{(k:p)}}\right) > 3/2\right).
\end{aligned}$$

Remark that for any u and v varying between $1/2$ and $3/2$, we have $|u^\tau - v^\tau| \leq M_\tau |1/u - 1/v|$ where $M_\tau = |\tau| \left((2)^{1+\tau} \vee (2/3)^{1+\tau}\right)$. Since $|g(x,t)| \leq \tilde{g}(t)$ for any $x \in [1/2, 3/2]$, it follows

$$\begin{aligned}
p \leq\ & P\left(\left|\left(\frac{\xi_{(k:p)}}{(k/p)}\right) - 1\right| > \frac{t}{2M_\tau G\left(\frac{p}{k}\right)}\right) \\
& + 1_{\left\{(2^{-\tau} \vee (3/2)^{-\tau})\tilde{g}\left(\frac{p}{k}\right) > \frac{t}{2G\left(\frac{p}{k}\right)}\right\}} \\
& + 4P\left(\left(\frac{\xi_{(k:p)}}{(k/p)} - 1\right) > 1 \text{ or } \left(\frac{\xi_{(k:p)}}{(k/p)} - 1\right) < -1/3\right) \\
\leq\ & P\left(\left|\xi_{(k:p)} - \frac{k}{p}\right| > \frac{tk/p}{2M_\tau G\left(\frac{p}{k}\right)}\right) + 1_{\left\{\tilde{r}\left(\frac{p}{k}\right) > t\, 2G\left(\frac{p}{k}\right)^{-1} (2^{-\tau} \vee (3/2)^{-\tau})^{-1}/2\right\}} \\
& + 4P\left(\left(\xi_{(k:p)} - \frac{k}{p}\right) > k/p \text{ or } \left(\xi_{(k:p)} - \frac{k}{p}\right) < -k/3p\right).
\end{aligned}$$

Applying Lemma 2, we deduce the announced inequality and Lemma 1 is proved.

References

Butucea, C., Tribouley, K. (2006). Nonparametric homogeneity tests. *Journal of Statistical Planning and Inference, 3*, 597–639.

Dekkers, A., Einmahl, J., De Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics, 17*, 1833–1855.

Drees, H., de Haan, L., Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *Journal of Statistical Planning and Inference, 136*, 3498–3538.

Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling extremal events for insurance and finance. Applications of mathematics* (Vol. 33). Berlin; New York: Springer.

Fraga Alves, I., Gomes, I., De Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica, 60*, 193–213.

Gomes, I., Martins, J. (2002). Asymptotically Unbiased estimators of the tail index based on external estimation of the second order parameter. *Extremes, 5:1*, 5–31.

de Haan, L., Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica, 52*, 60–70.

Hall, P., Welsh, A. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics, 13*, 331–341.

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

Ingster, Y. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives, I, II, III. *Mathematical Methods of Statistics, 2*, 85–114, 171–189, 248–268.

Shorack, G.R., Wellner, J.A. (1986). *Empirical processes with applications to statistics*. New York: Wiley.