ON THE ESTIMATION OF A CHANGEPOINT IN A TAIL INDEX

Kęstutis Gadeikis†, **Vygantas Paulauskas**‡

Abstract

In the paper we investigate application of a new estimator for tail index, proposed in [5] and [18]. Testing hypothesis of change at unknown place and detecting change in mean allow us to provide theoretical results on the estimation of a changepoint in a tail index. We demonstrate the applicability of these results in practice.

MSC: primary - 62G32, secondary - 62G05, 62G20, 60F17.
Keywords: tail index, changepoint.

1 INTRODUCTION AND FORMULATION OF RESULTS

During last decades many papers, devoted to various applications such as actuarial and financial mathematics, queing theory, internet traffic and so on, stressed the increasing role of heavy-tailed distributions.

Suppose that F is a distribution function (d.f.) and satisfies the following relation for large values of x:

$$1 - F(x) = x^{-\alpha} L(x),$$

where $\alpha > 0$ and L is slowly varying:

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1 \quad \forall t > 0.$$

Then α is called a tail index of the distribution.

Strictly speaking it should be called an index of the right tail, since we can define an index of the left tail of a distribution in a similar way. As we see, tail index reflects the fatness of the tail, determines the existence of the moments of some order (evidently, the left tail must not be heavier than the right one), is important in limit theorems for sum of i.i.d. random variables with this given d.f., etc.

Therefore during the last three decades there were a lot of papers which dealt with the estimation problem of the tail index. One of the first and probably the most popular estimator for the parameter $\gamma = 1/\alpha$ was proposed by B.M. Hill in [14]. This estimator is based on the order statistics $X_{n,1} \leq X_{n,2} \leq \ldots \leq X_{n,n}$

†The Faculty of Mathematics and Informatics, Naugarduko 24, Vilnius, gadeikis@ldr.lt
‡The Faculty of Mathematics and Informatics, Naugarduko 24, Vilnius, vygantas.paulauskas@maf.vu.lt, Institute of Mathematics and Informatics

from a sample $X_1, \dots, X_n$ of i.i.d. random variables with d.f. F. The estimator has the form

$$\gamma_{n,k} = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n,n-i} - \log X_{n,n-k},$$

where k is some number satisfying $1 \leq k \leq n$. The problem how to choose $k = k_n$ is not a simple one (see, for example, [4], [7], [8], [9], [12], [11]) and there were many papers devoted to this problem and to properties or modifications of Hill's estimator (see [1], [3], [13], [6], [19] and references therein).

In [5] and [18] it was proposed one more estimator which is based not on order statistics and, very interesting to note, came when considering rather abstract objects - random stable zonotopes (see [5]). The idea this new estimator is based on and construction of the estimator are quite simple.

From now let N stand for the sample size. Assuming $N = n \cdot m$ (in practice we choose m and then $n = [N/m]$, where $[\cdot]$ stands for the integer part of a positive number), we divide the sample into n groups $V_1, \dots, V_n$, each group containing m random variables. It's required that variables would be independent within a group and between the groups, for example $V_1 = \{X_1, \dots, X_m\}, V_2 = \{X_{m+1}, \dots, X_{2m}\}$ and so on. Let

$$M_{ni}^{(1)} = \max\{X_j \colon X_j \in V_i\}$$

and $M_{ni}^{(2)}$ - the second largest element in the same group V_i. Now let's denote

$$\kappa_{ni} = \frac{M_{ni}^{(2)}}{M_{ni}^{(1)}}, \qquad S_n = \sum_{i=1}^{n} \kappa_{ni}, \qquad Z_n = n^{-1} S_n.$$

Then under mild conditions on m and n that both n and $m \to \infty$ no slower than some positive degree of N (e.g., $n = CN^b$, $b \in (0;1)$, $C > 0$, and $m = [N/n] = [C^{-1}N^{1-b}]$) one can verify (see [18]) that

$$Z_n \xrightarrow{a.s.} \frac{\alpha}{\alpha + 1}.$$

This relation is nothing else as a strong law of large numbers for triangular array of random variables κ_{ni}, $i = 1, \dots, n$. Namely, in each row κ_{ni}, $i = 1, \dots, n$, are i.i.d. random variables and $E\kappa_{ni} \longrightarrow \frac{\alpha}{1+\alpha}$ with $N \to \infty$. Since $N = n \cdot m$, it is clear that one of the problems is how to choose m and n. This problem is similar to choose k in Hill's estimator. Assuming the so-called second-order asymptotic relation for tail of d.f., this problem was dealt with in [18].

Recently some effects in financial markets were explained by the change of tail index in some time series. It is well-known that some quantities like stock returns have heavy-tailed distributions. E.g., in the paper [20] we find investigation of the financial crisis during 1997-1998 in some Asian markets. Calculations given in [20] show that the data do not contradict the hypothesis that there was a change in tail index.

Mathematically the problem can be formulated as follows. Suppose that we are observing time series $X_1, X_2, \dots, X_N$ with distribution functions F_i, $i = 1, 2, \dots, N$. We want to test the hypothesis that there is a change in tail index in the sample, i.e. we want to check if there exist such number k, that

$F_1 = F_2 = \ldots = F_k$ with tail index α_1 and $F_{k+1} = F_{k+2} = \ldots = F_N$ with tail index α_2; of course, $\alpha_1 \neq \alpha_2$. If the data do not contradict hypothesis, then we would like to estimate the location of the changepoint, i.e. to estimate k.

In the paper [20] Hill's estimator is the main tool in testing this hypothesis, and several procedures were proposed for testing the hypothesis about the constancy of the tail index and location of the changepoint - recursive, rolling and sequential. Recursive estimator uses the first $[tn]$ elements of the sample, step by step increasing $t \in (0; 1)$. The rolling estimator fixes the subsample size and estimates α rolled through time. The sequential test is constructed from a pre-break and post-break estimators, i.e. the sequential pre-break estimator is just the recursive estimator and the post-break estimator is the reverse recursive estimator. In the same paper [20] one can find invariance principles for estimators and rates of convergence.

The main goal of this short note is to exploit the new tail index estimator from [18] for solving the problem of a changepoint in tail index. We propose the procedure which reduces estimation of the tail index to detecting the change in mean value. Therefore it is possible to apply well-known results, described in [15].

Suppose that we observe random variables $X_1^{(N)}, X_2^{(N)}, \ldots, X_N^{(N)}$ with distribution functions $F_i^{(N)}$ and there exists such unknown parameter $0 < \tau < 1$, such that $F_i^{(N)}$ for $i = 1, \ldots, [\tau N]$ are the same and have tail index α_1, and $F_i^{(N)}$ for $i = [\tau N] + 1, \ldots, N$ are identical and have tail index α_2. We can assume that all variables are positive, because we are investigating the index of the right tail only, so negative values can be dropped. We divide sample into n groups $V_1^{(N)}, \ldots, V_n^{(N)}$, each group containing m random variables, as we assume $N = n \cdot m$. Although all quantities below depend on N, dropping this index let us denote

$$M_{ni}^{(1)} = \max\{X_j \colon X_j \in V_i\}$$

and $M_{ni}^{(2)}$ let be the second largest element in the same group V_i. Finally, we denote

$$\kappa_{ni} = \frac{M_{ni}^{(2)}}{M_{ni}^{(1)}}.$$

Now we have the following sequence of independent random variables κ_{ni}: for $1 \leq i \leq k = [\frac{[\tau N]}{m}]$ random variables κ_{ni} are i.i.d. with expectation close to $\beta_1 = \alpha_1/(1+\alpha_1)$ and for $k+2 \leq i \leq n$ variables κ_{ni} are i.i.d. with expectation close to $\beta_2 = \alpha_2/(1+\alpha_2)$. There is only one random variable $\kappa_{n,k+1}$ which can be different from these of two groups, since it is constructed from the group, containing random variables with both tail indexes α_1 and α_2. Neglecting this variable we have standard problem of change in mean value, only in a triangular array scheme.

It is known (see, for example, [15]) that for this problem the following statistic can be used: for $1 \leq j \leq n-1$ and some parameter $0 \leq \zeta \leq 1$ let us define

$$U_j^{(n)} = \Big(\frac{j(n-j)}{n^2}\Big)^{1-\zeta}\Big(\frac{1}{j}\sum_{i=1}^{j}\kappa_{ni} - \frac{1}{n-j}\sum_{i=j+1}^{n}\kappa_{ni}\Big)$$

and

$$\tau_n = \frac{1}{n} \min \Big\{ i : \big|U_i^{(n)}\big| = \max_{1 \le l \le n-1} \big|U_l^{(n)}\big|, 1 \le i \le n-1 \Big\}.$$

Similarly to nonparametric change-point estimation (see [2]), we prove that τ_n is strongly consistent estimator for τ.

Theorem 1 *Suppose that $N \to \infty$ and $m \to \infty$. If there exist $b \in (0;1)$ and $C > 0$ such that $n \ge CN^b$, then $\tau_n \xrightarrow{a.s.} \tau$.*

We did not try to find the lowest possible growth of n and m in Theorem 1, since it was not an object of the article.

In order to get the rate of convergence in the above formulated theorem, we use general method from [10], where more general estimators of change-point are considered.

In [10] a triangular array $\xi_1^{(n)}, \dots, \xi_n^{(n)}$, $n \ge 2$, of row-wise independent random elements is considered. Suppose that $\xi_1^{(n)}, \dots, \xi_{[\tau n]}^{(n)}$ for some $\tau \in (0;1)$ have distribution ν_1 and $\xi_{[\tau n]+1}^{(n)}, \dots, \xi_n^{(n)}$ have distribution $\nu_{2,n}$. Let function $K : R^2 \to R$ be a measurable mapping (kernel) and define for $t \in [0;1]$

$$r_n(t) = \begin{cases} n^{-2} \sum\limits_{i=k+1}^{n} \sum\limits_{j=1}^{k} K\big(\xi_i^{(n)}, \xi_j^{(n)}\big), t = \frac{k}{n}, k = 1, \dots, n-1, \\ 0, \quad t = 0, 1, \\ r_n\big(\frac{k}{n}\big) + (nt - k)\Big(r_n\big(\frac{k+1}{n}\big) - r_n\big(\frac{k}{n}\big)\Big), t \in \big(\frac{k}{n}; \frac{k+1}{n}\big), k = 0, \dots, n-1. \end{cases}$$

Finally, the estimator of τ is defined as $\tau_n = \arg\max_{t \in [0;1]} w(t)|r_n(t)|$, where $w : [0;1] \to (0;\infty)$ is a weight function. By taking the smallest maximizer for convenience, τ_n will be unique.

In [10] it is required that the kernel K and the distributions ν_1 and $\nu_{2,n}$ satisfy the following conditions.

Antisymmetry (A): $K(x,y) = -K(y,x)\ \forall x, y \in R$.

Moment condition (M): There exist real numbers $a > 0$ and $M_0 < \infty$ such that

$$\int \big|K\big|^{2+a} d\mu_1 \otimes \mu_2 \le M_0, \quad \forall \mu_1, \mu_2 \in \{\nu_1, \nu_{2,n}\}.$$

It is also required that the following functions exist:

$$R_1(y) = \int K(x,y)\nu_1(dx), \quad R_n(y) = \int K(x,y)\nu_{2,n}(dx).$$

Stability condition (S):

$$\int f_n g_n \, dP_n \to \sigma^2, \quad \forall f_n, g_n \in \{R_1, R_n\},\ P_n \in \{\nu_1, \nu_{2,n}\},$$

where $\sigma^2 = \int R_1^2 d\nu_1$. Furthemore, set

$$\lambda_n = \int K d\nu_{2,n} \otimes \nu_1,$$

where it is assumed that $\lambda_n \ne 0\ \forall n \in N$ and $\lambda_n \to 0$ in such a way that $\lambda_n n^{\frac{1}{2}} \to \infty$ as $n \to \infty$.

Regularity condition (R): $w : [0;1] \to (0;\infty)$, w' is bounded on $[d;1-d]$ for any positive d and continuous at τ. Furthermore, $\inf_{0\le t\le\tau}\big(tw(t)\big)' > 0$ and $\inf_{1-\tau\le t\le 1}\big(w(t)(1-t)\big)' < 0$.

Remark: in original paper [10] w' is supposed to be bounded on $(0;1)$; from the proof of the theorem and examples given there it is easy to see that correct regularity condition must be stated as above.

Now we apply this result taking κ_{ni} instead of $\xi_i^{(n)}$. Setting $w(t) = \big(t(1-t)\big)^{-\zeta}$, $K(x,y) = y-x$ and taking only discrete time moments $t = k/n$, actually we have the definition of $U_j^{(n)}$. Let's check the conditions (A), (M), (S) and (R) in our case.

For the kernel $K(x,y) = y-x$ antisymmetry is evident.

Because $0 \le \kappa_{ni} \le 1$, $i = 1,\dots,n$, functions $R_1(y)$, $R_n(y)$ surely exist (we remind, that we assume that all X_j are positive). Moment condition is also satisfied, e.g. by taking $a = M_0 = 1$.

Because $w(t) = \big(t(1-t)\big)^{-\zeta}$, we have

$$w'(t) = -\zeta\frac{1-2t}{\big(t(1-t)\big)^{1+\zeta}},$$

$$\big(tw(t)\big)' = \frac{1-\zeta-t(1-2\zeta)}{t^{\zeta}(1-t)^{1+\zeta}},$$

$$\big(w(t)(1-t)\big)' = \frac{-\zeta-t(2\zeta-1)}{t^{1+\zeta}(1-t)^{\zeta}}.$$

It is easy to see that the weight function satisfies the conditions of regularity.

Finally, let's denote $\beta_i = \alpha_i/(1+\alpha_i)$ and $\sigma_i^2 = \alpha_i/\big((1+\alpha_i)^2(2+\alpha_i)\big)$, $i = 1,2$. β_1 and σ_1 are the limit mean value and variance of κ_{nj} for j before the change and β_2 and σ_2 are corresponding values after the change. This means that not only expectation, but also variances are changing in the sample. So in our case stability condition (S) is not satisfied. But this condition was essential in the case when $\nu_{2,n}$ weakly converges to ν_1 and $\lambda_n \to 0$ (this is more difficult case, allowing distributions before the change and after the change to become closer and closer), while in our case $\lambda_n = E\kappa_{n1} - E\kappa_{nn} \to \beta_1 - \beta_2 \ne 0$. Therefore by checking all steps in the proof of Theorem 1.1 from [10] we shall show that stability condition is not needed in our case.

Theorem 2 *Let κ_{ni} for $1 \le i \le k = [\tau n]$ be i.i.d. random variables with expectation $\beta_1^{(n)}$ and for $k+1 \le i \le n$ - with expectation $\beta_2^{(n)}$. If $\beta_1^{(n)} \to \beta_1$ and $\beta_2^{(n)} \to \beta_2$ as $n \to \infty$, then*

$$n(\beta_1-\beta_2)^2(\tau_n-\tau) \xrightarrow{d} \chi(Y),$$

where $\chi(Y)$ is the (a.s) unique maximizer of the process $Y(t) = W(t) + h(t)$, $t \in R$. Here W is a two-sided Brownian motion on R and the drift function

$$h(t) = \begin{cases} -\sigma_2^{-1}\big(\tau+\zeta(1-2\tau)\big)t, & t \ge 0,\\ \sigma_1^{-1}\big(1-\tau-\zeta(1-2\tau)\big)t, & t < 0.\end{cases}$$

Application of the theorem is described below.

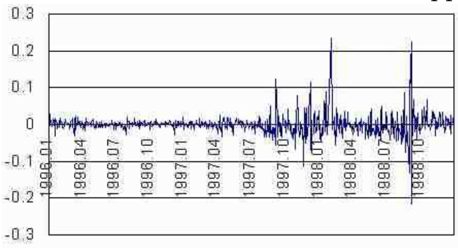
2 CHANGE OF TAIL INDEX DURING ASIAN FINANCIAL CRISIS

Crisis of Asian financial markets in the middle of 1997 is an interesting example of structural change. Rapid growth of economics stimulated investors to expand production and to take credit risks. However, such growth of economics couldn't last forever. On July 2, 1977 Thailand devalued its currency against the US dollar, and it was the beginning of a serious crisis in the whole Asian region.

Daily stock price indices of Asian countries during 1996-1998 allow us to investigate the situation. Since the data is easily available on internet, the main thing is a choice of the method for detecting the change point in tail index. To choose the best method is not a simple question.

During the crisis one can observe more intense fluctuation of stock market prices. To be more exact, daily returns are supposed to have heavy-tailed distribution. Despite limited number of observations, calculations usually indicate high probability of structural change. It is nothing else than evidence of change in a tail index of the distribution - decrease of the tail index means heavier tail.

In Picture 1 there are daily returns of Kuala Lumpur Composite Index (KLCI), calculated in Malaysian stock exchange. It is easy to see that there is structural change in the data. For this data we shall apply our results.



Picture 1. Daily returns of KLCI during 1996-1998

It is easy to notice that using the definition of $U_j^{(n)}$ and taking $\zeta = 0$ we have

$$\frac{j(n-j)}{n^2}\Big(\frac{1}{j}\sum_{i=1}^{j}\kappa_{ni} - \frac{1}{n-j}\sum_{i=j+1}^{n}\kappa_{ni}\Big) =$$

$$= \frac{j(n-j)}{n^2}\Big(\frac{1}{j}\sum_{i=1}^{j}\kappa_{ni} - \frac{1}{n-j}(\sum_{i=1}^{n}\kappa_{ni} - \sum_{i=1}^{j}\kappa_{ni})\Big) =$$

$$= \frac{j(n-j)}{n^2}\Big(\frac{n}{j(n-j)}\sum_{i=1}^{j}\kappa_{ni} - \frac{1}{n-j}\sum_{i=1}^{j}\kappa_{ni}\Big) =$$

$$= \frac{1}{n}\Big(\sum_{i=1}^{j}\kappa_{ni} - \frac{j}{n}\sum_{i=1}^{n}\kappa_{ni}\Big).$$

Now assuming that there is no change of distribution within $X_{n1}, \ldots, X_{nn}$ and denoting standard deviation of X_{ni} by σ, we see that

$$\frac{\sqrt{n}}{\sigma} r_n(t) \to W^0(t)$$

in distribution as $n \to \infty$, where $W^0(t)$ is a Brownian bridge, $t \in [0; 1]$. Therefore we have that

$$\sup_{t\in[0;1]} \frac{\sqrt{n}}{\sigma} |r_n(t)| \xrightarrow{d} \sup_{t\in[0;1]} |W^0(t)|.$$

Define

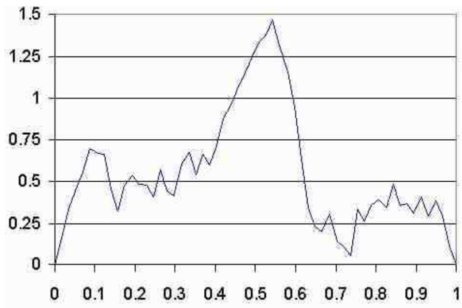
$$D = \sup_{t\in[0;1]} |W^0(t)|.$$

The distribution function of this random variable is well known:

$$F_D(x) = \begin{cases} 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2x^2}, x \geq 0, \\ 0, x < 0. \end{cases}$$

The series is extremely rapidly converging - usually a few terms suffice for very high accuracy.

Now we are ready to apply a test for the null hypothesis that the tail index of KLCI daily returns is constant over time. Having 744 daily observations of the index during 1996-1998, we divide them into n groups each having m random variables. For visual purposes taking $m = 13$, by procedure described at the begining of the article we get 57 random variables κ_{ni}, i.e. $i = 1, \ldots, 57$. Then we calculate $r_n(t)\sqrt{n}/\sigma$, the graph of this function is given below:



Picture 2. Graph of $r_n(t)n^{\frac{1}{2}}\sigma^{-1}$

As we see, at the point $t = 31/57$ we have maximum value 1.47. The 95% critical value of D is 1.36, so we able to reject the null of the constancy of the tail index for the distribution of KLCI daily returns.

Tabulating random variable $\chi(Y)$ as it is defined in Theorem 2, we should even be able to construct confidence intervals for the changepoint. However, tabulation process is not easy. It is much more important to notice that the detected changepoint $t = 31/57$ actually means the middle of 1997, which actually is the beginning of the Asian crisis.

Here one remark is appropriate. Both in [20] and in our paper the data of Asian financial crisis was analysed assuming the presence of one change point. It seems that a more realistic assumption would be to assume that at the beginning of the crisis the tail index had increased, but after short period, when some measures of stabilization in the financial market were taken, then the tail index had returned to the previous level (the so-called epidemic behaviour). But to detect such behaviour is more difficult and different approach is used.

This can be seen from the following effect: if we take the data corresponding to longer period, we shall not reject the null hypothesis. On the other hand, if we take simulated data with one change point, then it is easy to see that longer samples give the better results.

3 PROOFS

Proof of Theorem 1. From the well-known criteria (see, for example, [17]) we have that the statement $\tau_n \stackrel{a.s}{\to} \tau$ is equivalent to the following relation: for any $\epsilon > 0$

$$P\big(\omega : \sup_{m\geq l} |\tau_m(\omega) - \tau| \geq \epsilon\big) \to 0, \tag{1}$$

as $l \to \infty$.

Suppose that (1) does not hold, that is, there exist such $\epsilon > 0$, $\delta > 0$ and subsequence (l_j):

$$P\big(\omega : \sup_{m\geq l_j} |\tau_m(\omega) - \tau| \geq \epsilon\big) \geq \delta \quad \forall j. \tag{2}$$

Let's notice that sets $A_l^\epsilon = \big\{\omega : \sup_{m\geq l} |\tau_m(\omega) - \tau| \geq \epsilon\big\}$ are monotonous, i.e. $A_1^\epsilon \supset A_2^\epsilon \supset \ldots \supset A_l^\epsilon \supset A_{l+1}^\epsilon \supset \ldots$. So (2) can be written as follows: there $\exists \epsilon > 0$ and $\delta > 0$ such that for all l

$$P\big(A_l^\epsilon\big) \geq \delta. \tag{3}$$

Denote $V^\epsilon = \limsup_{l\to\infty} A_l$, $P(V^\epsilon) \geq \delta$.

Here and in what follows all limits are as $n \to \infty$. For $k = [\tau n]$ applying strong law of large numbers we have

$$\big|U_k^{(n)}\big| \stackrel{a.s.}{\to} A,$$

where $A = \big(\tau(1-\tau)\big)^{1-\zeta} |\beta_1 - \beta_2|$.

Taking $k = [tn]$, $t > \tau$, one can make sure that

$$\big|U_k^{(n)}\big| \stackrel{a.s.}{\to} \big(t(1-t)\big)^{1-\zeta} \ \frac{\tau}{t} \ |\beta_1 - \beta_2| = \Big(\frac{\tau}{t}\Big)^\zeta \Big(\frac{1-t}{1-\tau}\Big)^{1-\zeta} A. \tag{4}$$

Similarly taking $k = [tn]$, $t < \tau$, we have

$$\big|U_k^{(n)}\big| \stackrel{a.s.}{\to} \big(t(1-t)\big)^{1-\zeta} \ \frac{1-\tau}{1-t} \ |\beta_1 - \beta_2| = \Big(\frac{1-\tau}{1-t}\Big)^\zeta \Big(\frac{t}{\tau}\Big)^{1-\zeta} A. \tag{5}$$

If $\omega \in V$, there exists subsequence $\big(\tau_{m_i}(\omega)\big)$ such that $|\tau_{m_i}(\omega) - \tau| \geq \epsilon \ \forall i$. Since V is a set of non-zero probability, for any $\omega \in V$ due to relations (4) and (5) we have that $\forall \eta > 0$ there $\exists I > 0$ such that for $i \geq I$ and $\tau_{m_i}(\omega) \geq \tau + \epsilon$

$$\Big|U_{[\tau_{m_i}(\omega)n]}^{(m_i)} - \Big(\frac{\tau}{\tau_{m_i}(\omega)}\Big)^\zeta \Big(\frac{1-\tau_{m_i}(\omega)}{1-\tau}\Big)^{1-\zeta} A\Big| < \eta.$$

Similarly if $\tau_{m_i}(\omega) \leq \tau - \epsilon$, we have

$$\Big|U_{[\tau_{m_i}(\omega)n]}^{(m_i)} - \Big(\frac{1-\tau}{1-\tau_{m_i}(\omega)}\Big)^\zeta \Big(\frac{\tau_{m_i}(\omega)}{\tau}\Big)^{1-\zeta} A\Big| < \eta.$$

Because function $\big(\tau/x\big)^{\zeta}\big((1-x)/(1-\tau)\big)^{1-\zeta}$ is decreasing in x and function $\big((1-\tau)/(1-x)\big)^{\zeta}\big(x/\tau\big)^{1-\zeta}$ is increasing in x, therefore, if $\tau_{m_i}(\omega) \ge \tau+\epsilon$, then

$$\Big(\frac{\tau}{\tau_{m_i}(\omega)}\Big)^{\zeta}\Big(\frac{1-\tau_{m_i}(\omega)}{1-\tau}\Big)^{1-\zeta} \le \Big(\frac{\tau}{\tau+\epsilon}\Big)^{\zeta}\Big(\frac{1-\tau-\epsilon}{1-\tau}\Big)^{1-\zeta} < 1,$$

and if $\tau_{m_i}(\omega) \le \tau-\epsilon$, then

$$\Big(\frac{1-\tau}{1-\tau_{m_i}(\omega)}\Big)^{\zeta}\Big(\frac{\tau_{m_i}(\omega)}{\tau}\Big)^{1-\zeta} \le \Big(\frac{1-\tau}{1-\tau+\epsilon}\Big)^{\zeta}\Big(\frac{\tau-\epsilon}{\tau}\Big)^{1-\zeta} < 1.$$

Then two last inequalities mean that limits in relations (4) and (5) are strictly less than A, and we get contradiction to our assumption (3) (which was shown to be equivalent to (2)).

Proof of Theorem 2. Assume $\beta_1 < \beta_2$. Define $\lambda_n = \beta_1^{(n)} - \beta_2^{(n)}$ and $\Lambda_n = \big\{\frac{i}{n} : i = 1,2,\dots,n\big\}$. Since

$$\tau_n = \arg\max_{t\in\Lambda_n} w(t)|r_n(t)|$$

and $r_n(t)$ will be negative for large values of n (since the sign of $r_n(t)$ is defined by $\beta_1-\beta_2$), we introduce

$$\tilde\tau_n = \arg\min_{t\in\Lambda_n} w(t)r_n(t).$$

First we show that τ_n and $\tilde\tau_n$ are asymptotically equal in probability. Set

$$c_n = \frac{1}{2}w\Big(\frac{[n\tau]}{n}\Big)\big|\lambda_n\big|\frac{[n\tau]}{n}\Big(1-\frac{[n\tau]}{n}\Big).$$

Then

$$\begin{aligned} P\big(\tau_n \ne \tilde\tau_n\big) &\le P\Big(\max_{t\in\Lambda_n} w(t)r_n(t) \ge -\min_{t\in\Lambda_n} w(t)r_n(t)\Big) \le \\ &\le P\Big(-\min_{t\in\Lambda_n} w(t)r_n(t) \le c_n\Big) + P\Big(\max_{t\in\Lambda_n} w(t)r_n(t) \ge c_n\Big). \end{aligned}$$

Since

$$Er_n(\tau) = n^{-2}\sum_{i=[n\tau]+1}^{n}\sum_{j=1}^{[n\tau]} E(\kappa_{nj}-\kappa_{ni}) = \lambda_n\frac{[n\tau]}{n}(1-\frac{[n\tau]}{n})$$

and $\lambda_n < 0$ for large values of n, we see that $c_n = -\frac{1}{2}w\big(\frac{[n\tau]}{n}\big)Er_n(\tau)$. Now

$$\begin{aligned} &P\Big(\min_{t\in\Lambda_n} w(t)r_n(t) \ge -c_n\Big) \le P\Big(w(\frac{[n\tau]}{n})r_n(\tau) \ge -c_n\Big) = \\ =\ &P\Big(w(\frac{[n\tau]}{n})\big(r_n(\tau)-Er_n(\tau)\big) \ge w(\frac{[n\tau]}{n})\big|\lambda_n\big|\frac{[n\tau]}{n}(1-\frac{[n\tau]}{n}) - c_n\Big) = \\ =\ &P\Big(w(\frac{[n\tau]}{n})n^{1/2}\big(r_n(\tau)-Er_n(\tau)\big) \ge n^{1/2}c_n\Big). \end{aligned}$$

Because $n^{1/2}\big(r_n(\tau)-Er_n(\tau)\big)$ converges in distribution to a normal distribution and $n^{1/2}c_n \to \infty$, it follows that $P\big(\min_{t\in\Lambda_n} w(t)r_n(t) \ge -c_n\big) \to 0$ as $n\to\infty$.

In a similar way we show $P\big(\max_{t\in\Lambda_n} w(t)r_n(t) \geq c_n\big) \to 0$. This proves that $\lim\limits_{n\to\infty} P\big(\tau_n \neq \tilde{\tau_n}\big) = 0$.

For $t \in R$ introduce

$$Y_{n,w}(t) = n\lambda_n\big(r_{n,w}(\tau + \lambda_n^{-2}n^{-1}t) - r_{n,w}(\tau)\big)$$

with $r_{n,w}(t) = w(t)r_n(t)$.

We will show that

$$EY_{n,w}(t) \to \sigma(t)w(\tau)h(t), \tag{6}$$

where $\sigma(t) = \sigma_1 I_{\{t<0\}} + \sigma_2 I_{\{t\geq 0\}}$. Note that after multiplying $h(t)$ by $\sigma(t)$ there is no $\sigma(t)$ left in the expression, because actually we have $\sigma^{-1}(t)$ in the definition of $h(t)$. Such definition is used only to have the results in a simpler form.

The proof of (6) is straightforward. Denote $\Delta_n(t) = \tau + \lambda_n^{-2}n^{-1}t$. Then

$$\begin{aligned}
&Y_{n,w}(t) = n\lambda_n\Big(r_{n,w}\big(\Delta_n(t)\big) - r_{n,w}(\tau)\Big) = \\
=\ & n\lambda_n\Big(r_n\big(\Delta_n(t)\big)w\big(\Delta_n(t)\big) - r_n(\tau)w(\tau)\Big) = \\
=\ & n\lambda_n\Big(r_n\big(\Delta_n(t)\big)w\big(\Delta_n(t)\big) - r_n\big(\Delta_n(t)\big)w(\tau) + r_n\big(\Delta_n(t)\big)w(\tau) - r_n(\tau)w(\tau)\Big) = \\
=\ & n\lambda_n\Big(r_n\big(\Delta_n(t)\big)w'\big(\varphi_n(t)\big)\lambda_n^{-2}n^{-1}t + w(\tau)\big(r_n\big(\Delta_n(t)\big) - r_n(\tau)\big)\Big),
\end{aligned}$$

where $\varphi_n(t)$ is a point between τ and $\Delta_n(t)$. Note that when $n \to \infty$, we have $\Delta_n(t) \to \tau$, therefore $w'\big(\varphi_n(t)\big) \to w'(\tau)$ and $Er_n\big(\varphi_n(t)\big) \to Er_n(\tau) = \tau(1-\tau)\lambda_n$.

We also use equalities (1) and (2). It's easy to derive that for $t \geq 0$

$$Er_n\big(\Delta_n(t)\big) - Er_n(\tau) \to \Delta_n(t)\big(1-\Delta_n(t)\big)\frac{\tau}{\Delta_n(t)}\lambda_n - \tau(1-\tau)\lambda_n = -\tau\lambda_n^{-1}n^{-1}t,$$

and for $t < 0$ we have

$$\begin{aligned}
Er_n(\Delta_n(t)) - Er_n(t) \quad &\to \quad \Delta_n(t)(1-\Delta_n(t))\frac{1-\tau}{1-\Delta_n(t)}\lambda_n - \tau(1-\tau)\lambda_n = \\
&= \quad (1-\tau)\lambda_n^{-1}n^{-1}t.
\end{aligned}$$

From these relations we get

$$EY_{n,w}(t) \to g(t) = \begin{cases} -\tau t w(\tau) + \tau(1-\tau)tw'(\tau), & t \geq 0, \\ (1-\tau)tw(\tau) + \tau(1-\tau)tw'(\tau), & t < 0. \end{cases}$$

Taking $w(t) = \big(t(1-t)\big)^{-\zeta}$ we see that $g(t) = w(\tau)h(t)$ and (6) is proved.

Define $X_{n,w} = Y_{n,w} - EY_{n,w}$ and denote by Y_n and X_n the corresponding processes with the weight function $w(t) \equiv 1$. Then $\big\{X_n(t) : -d \leq t \leq d\big\}$, $n \geq N_0(d)$, is asymptotically C-tight for all $d > 0$. The proof is given in [10].

We prove that $X_n(t) \xrightarrow{d} \sigma(t)W(t)$. Define $n_t = [n\tau + \lambda_n^{-2}t]$. Then let us consider the case $t \geq 0$:

$$
\begin{aligned}
Y_n(t) &= \lambda_n n\Big(r_n(\tau + \lambda_n^{-2}n^{-1}t) - r_n(\tau)\Big) = \\
&= \lambda_n n^{-1}\Big(\sum_{i=n_t+1}^{n}\sum_{j=1}^{n_t} K(X_i, X_j) - \sum_{i=[n\tau]+1}^{n}\sum_{j=1}^{[n\tau]} K(X_i, X_j)\Big) = \\
&= \lambda_n n^{-1}\Big(\big(n_t - [n\tau]\big)\sum_{i=n_t+1}^{n} X_i - \big(n - n_t\big)\sum_{j=[n\tau]+1}^{n_t} X_j - \\
&- [n\tau]\sum_{i=[n\tau]+1}^{n_t} X_i + \big(n_t - [n\tau]\big)\sum_{j=1}^{[n\tau]} X_j\Big) = B_1^{(n)} + B_2^{(n)},
\end{aligned}
$$

where

$$B_1^{(n)} = \lambda_n n^{-1}\big(n_t - [n\tau]\big)\sum_{i \notin \big([n\tau]+1, n_t\big)} X_i$$

and

$$B_2^{(n)} = \lambda_n n^{-1}\big(n + [n\tau] - n_t\big)\sum_{i=[n\tau]+1}^{n_t} X_i.$$

Taking into account the definition of n_t, it is not difficult to verify that $VarB_1^{(n)} \to 0$ and $VarB_2^{(n)} \to t\sigma_2^2$ as $n \to \infty$. Since we have shown that $Y_n(t)$ is expressed as a sum of two summands with the variance of the first summand tending to zero, it is clear that finite-dimentional distributions of $Y_n(t)$ are defined by corresponding distributions of the second summand. Standard argument gives us that that in the limit we have Gaussian distribution. For $t < 0$ after identical calculations we have $VarY_n(t) \to t\sigma_1^2$. Combining tightness of $X_n(t)$ and the fact that $VarY_n(t) \to t\sigma^2(t)$, $X_n(t) \xrightarrow{d} \sigma(t)W(t)$ as $n \to \infty$.

The next step is to show that for all $d > 0$ $Y_{n,w}$ converges to the process $\tilde{Y} = \sigma w(\tau)\big(W + h\big)$ in distribution in the space $C[-d, d]$. We have

$$
\begin{aligned}
&Y_{n,w}(t) = w(\tau)Y_n(t) + n\lambda_n\big(w\big(\Delta_n(t)\big) - w(\tau)\big)r_n\big(\Delta_n(t)\big), \\
&X_{n,w}(t) = Y_{n,w}(t) - EY_{n,w}(t) = \\
&= w(\tau)X_n(t) + n\lambda_n\big(w\big(\Delta_n(t)\big) - w(\tau)\big)\Big(r_n\big(\Delta_n(t)\big) - Er_n\big(\Delta_n(t)\big)\Big). \quad (7)
\end{aligned}
$$

Applying mean-value theorem from analysis, the second summand is equal to

$$n^{-1/2}\lambda_n^{-1}w'\big(\varphi_n(t)\big)tn^{1/2}\Big(r_n\big(\Delta_n(t)\big) - Er_n\big(\Delta_n(t)\big)\Big),$$

where $\varphi_n(t)$ is a point between τ and $\Delta_n(t)$. The factor $n^{-1/2}\lambda_n^{-1}w'\big(\varphi_n(t)\big)t$ converges to zero uniformly on $[-d, d]$ and $n^{1/2}\Big(r_n\big(\Delta_n(t)\big) - Er_n\big(\Delta_n(t)\big)\Big)$ converges in distribution to a normal distribution. Therefore $X_{n,w}(t) = w(\tau)X_n(t) + o_P(1)$. Recall that $EY_{n,w}(t) \to \sigma(t)w(\tau)h(t)$ and $X_n \xrightarrow{d} \sigma W$. Since $Y_{n,w}(t) = X_{n,w}(t) + EY_{n,w}(t)$, we have

$$Y_{n,w} \xrightarrow{d} \tilde{Y} = w(\tau)\sigma W + \sigma w(\tau)h = \sigma w(\tau)\big(W + h\big),$$

as $n \to \infty$.

Having proved this relation the rest of the proof goes along the lines of the proof of Theorem 1.1 from [10]. Namely, we use lemmas 2.6, 2.7 and 2.8 taken from [10].

By lemma 2.6 the process $\tilde{Y}$ is continuous and has the unique minimizer almost surely.

Furthermore, we define $S(f) = \{t \in R : f(t) = \inf_{s\in R} f(s)\}$ for $f \in R^R$ and $M = \{f \in R^R : S(f) \neq \emptyset \text{ is closed}\}$. Let $\psi : M \to R$ be given by $\psi(f) = \min S(f)$. Let us choose t from the following relation $\tilde{\tau}_n = \tau + \lambda_n^{-2} n^{-1} t$. Then taking into account definition of $\tilde{\tau}_n$ we see that

$$n\lambda_n^2(\tilde{\tau}_n - \tau) = \psi(Y_{n,w}).$$

By lemma 2.7 in [10] the random variable $\psi_d(\tilde{Y})$ is continuous, here $\psi_d : M_d \to R$ and M_d is defined in the same way as M, only instead of R^R we take $C[-d, d]$.

Finally, for $d > 0$ we introduce $A_{n,d} = \{\psi(Y_{n,w}) \neq \psi_d(Y_{n,w})\}$. By lemma 2.8 in [10] $\forall d > 0$ there $\exists C > 0$, which does not depend on d or n, and $n_0 \in N$ such that

$$P(A_{n,d}) \leq Cd^{-1}, \quad \forall n \geq n_0.$$

Now we are ready to complete the proof of the theorem. Denote distribution functions of $\psi(\tilde{Y})$, $\psi(Y_{n,w})$, $\psi_d(Y_{n,w})$ and $\psi_d(\tilde{Y})$ by F, F_n, $F_{n,d}$ and F_d respectively. Then

$$|F_n(t) - F(t)| \leq |F_n(t) - F_{n,d}(t)| + |F_{n,d}(t) - F_d(t)| + |F_d(t) - F(t)|.$$

For $n \geq n_0$ and $d > 0$ we have

$$|F_n(t) - F_{n,d}(t)| = \Big|P\big(\psi(Y_{n,w}) \leq t\big) - P\big(\psi_d(Y_{n,w}) \leq t\big)\Big| \leq 2P(A_{n,d}) \leq Cd^{-1}.$$

Because $Y_{n,w} \xrightarrow{d} \tilde{Y}$ as $n \to \infty$, $|F_{n,d}(t) - F_d(t)| \to 0$ for all $t \in R$ as F_d is continuous.

Since $\psi_d(\tilde{Y})$ converges almost surely and therefore in distribution to $\psi(\tilde{Y})$ as $d \to \infty$, $|F_d(t) - F(t)| \to 0$ for every continuity point t of F.

So we have $\psi(Y_{n,w}) \xrightarrow{d} \psi(\tilde{Y}) \overset{d}{=} \chi(Y)$. Because

$$\psi(\tilde{Y}) = \psi\Big(\sigma w(\tau)\big(W + h\big)\Big) = \psi(W + h),$$

that ends the proof.

REFERENCES

[1] Beirlant, J., Teugels, J. L. and Vynckier, P.: Tail index estimation, Pareto quantile plots, and regression diagnostics, *J. Amer. Statist. Assoc.* **91** (1996), 1659–1667

[2] Carlstein, E.: Nonparametric Change-Point Estimation, *Ann. Statist.* **16** (1988), 188–197

[3] Csörgő, S., Deheuvels, P. and Mason, D.: Kernel estimates of the tail index of a distribution, *Ann. Statist.* **13** (1985), 1050–1077

[4] Danielsson, J., de Haan, L., Peng, L. and de Vries, C. G.: Using a bootstrap method to choose the sample fraction in tail index estimation, *J. Multivariate Anal.* **76** (2001), 226–248

[5] Davydov, Yu., Paulauskas, V., Račkauskas, A.: More on p-stable convex sets in Banach spaces, *J. Theoret. Probab.* **13** (2001), 39–64

[6] Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L.: A moment estimator for the index of an extreme-value distribution, *Ann. Statist.* **17** (1975), 1833–1855

[7] Dekkers, A. L. M. and de Haan, L.: Optimal choice of sample fraction in extreme-value estimation, *J. Multivariate Anal.* **47**(2) (1993), 172–173

[8] Draisma, G., de Haan, L. and Peng, L.: A Bootstrap-based method to achieve optimality in estimating the extreme-value index, *Extremes* **2**(4) (1999), 367–404

[9] Drees, H. and Kaufmann, E.: Selecting the optimal sample fraction in univariate extreme value estimation, *Stochastic Process. Appl.* **75**(2) (1998), 149–172

[10] Ferger, D.: Change-point estimators in case of small disorders, *J. Statist. Plann. Inference* **40** (1994), 33–49

[11] Geluk, J. L. and Peng, L.: An adaptive optimal estimate of the tail index for MA(1) time series, *Statist. Probab. Lett.* **46** (2000), 217–227

[12] Hall, P.: Using the bootstrap to estimate means squared error and select smoothing parameter in nonparametric problems, *J. Multivariate Anal.* **32** (1990), 177–203

[13] de Haan, L. and Resnick, S.: Method to choose the sample fraction in tail index estimation, *J. Roy. Statist. Soc. Ser. B* **42** (1980), 83–87

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

[15] Kokoszka, P., Leipus, R.: Change-point in the mean of dependent observations, *Statist. Probab. Lett.* **40** (1998), 385–393

[16] LePage, R., Woodroofe, M., Zinn, J.: Convergence to a stable distribution via order statistics, *Ann. Probab.* vol. 9, No. 4 (1981), 624–632

[17] Loeve, M.: Probability Theory, 4th edition, Springer (1977), 116

[18] Paulauskas V.: A New Estimator for Tail Index, *Acta Appl. Math.* **79** (2003), 55–67

[19] Pickands, J.: Statistical inference using extreme order statistics, *Ann. Statist.* **3**(1) (1975), 119–131

[20] Quintos, C., Fan, Zh., Phillips, P.: Structural Change Tests in Tail Behaviour and the Asian Crisis, *Rev. Econom. Stud.* **13** 2001, 633–663