IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 95–104

DOI: 10.1214/10-IMSCOLL710

Simple sequential procedures for change in distribution

Marie Hušková[*,1] **and Ondřej Chochola**[1]

Charles University of Prague, Department of Statistics

Abstract: A simple sequential procedure is proposed for detection of a change in distribution when a training sample with no change is available. Its properties under both null and alternative hypothesis are studied and possible modifications are discussed. Theoretical results are accompanied by a simulation study.

1. Introduction

We assume that the observations $X_1, \dots, X_n, \dots$ are arriving sequentially, X_i has a continuous distribution function F_i, $i = 1, 2, \dots$ and the first m observations have the same distribution function F_0, i. e.,

$$F_1 = \dots = F_m = F_0,$$

where F_0 is unknown. $X_1, \dots, X_m$ are usually called training data. We are interested in testing the null hypothesis

$$H_0 : F_i = F_0, \quad \forall\, i \geq m,$$

against the alternative hypothesis

$$\begin{aligned} H_A &: \text{there exists } k^* \geq 0 \text{ such that } F_i = F_0,\ 1 \leq i \leq m + k^*, \\ &F_i = F^0,\ m + k^* < i < \infty, \quad F_0 \neq F^0. \end{aligned}$$

In case of independent observations there are no particular assumptions on the distribution functions F_i except their continuity. In case of dependent observations certain dependency among observations is assumed. Such a problem was considered by [2, 11] and [12]. Mostly such testing problems concern a change in finite dimensional parameter, see, [3, 8, 1] among others. They developed and studied sequential tests for a change in parameters in regression models.

Our test procedure is described by the stopping rule:

$$\tau_{m,N} = \inf\{1 \leq k \leq N : |Q(m, k)| \geq c\, q_\gamma(k/m)\} \tag{1}$$

[*]*Acknowledgement.* The work of the first author has been supported by grants GAČR 201/09/J006, MSM 0021620839 and LC06024. The work of the second author has been supported by grants SVS 261315/2010 and GAUK 162310.
[1]*Address:* MFF UK Sokolovská 83, CZ – 186 75 Praha 8, Czech Republic
e-mail: marie.huskova@karlin.mff.cuni.cz, e-mail: chochola@karlin.mff.cuni.cz
AMS 2000 subject classifications: 62G20, 62E20, 62L10.
Keywords and phrases: change in distribution, sequential monitoring.

with $\inf\emptyset := \infty$ and either $N = \infty$ or $N = N(m)$ with $\lim_{m\to\infty} N(m)/m = \infty$. $Q(m,k)$ is a detector depending on $X_1,\dots,X_{m+k}$, $k = 1,2,\dots$, $q_\gamma(t)$, $t\in(0,\infty)$ is a boundary function with $\gamma\in[0,1/2)$ (a tuning parameter) and c is a suitably chosen positive constant.

We require under H_0 for $\alpha\in(0,1)$ (fixed) and, under H_A,

$$\lim_{m\to\infty} P_{H_0}\big(\tau_{m,N} < \infty\big) \;=\; \alpha, \tag{2}$$

and, under H_A,

$$\lim_{m\to\infty} P_{H_A}\big(\tau_{m,N} < \infty\big) \;=\; 1. \tag{3}$$

The request (2) means that the test has asymptotically level α and (3) corresponds to consistency of the test. We usually choose the detectors $Q(m,k)$'s and the boundary function $q_\gamma(\cdot)$, then constant c has to fulfill under H_0

$$\lim_{m\to\infty} P\left(\max_{1\le k\le N}\frac{|Q(m,k)|}{q_\gamma(k/m)}\ge c\right)=\alpha.$$

In the present paper we choose

$$Q(m,k)=\frac{1}{\widehat{\sigma}_m\sqrt{m}}\sum_{i=m+1}^{m+k}(\widehat{F}_m(X_i)-1/2),\quad k=1,2,\dots, \tag{4}$$

where $\widehat{F}_m$ is an empirical distribution function based on $X_1,\dots,X_m$ and $\widehat{\sigma}_m$ is a suitable standardization based on $X_1,\dots,X_m$. We put

$$q_\gamma(t)=(1+t)(t/(1+t))^\gamma,\ t\in(0,\infty),\quad 0\le\gamma<1/2. \tag{5}$$

Two sets of assumptions on the joint distribution of X_i's are considered. One set assumes that $\{X_i\}_i$ are independent random variables and X_i has continuous distribution function F_i, $i=1,2,\dots$, i. e., under H_0 they are independent identically distributed (i.i.d.) with common unknown continuous distribution function F_0. The other set of conditions admits dependent observations.

Notice that the detector $Q(m,k)$ can be expressed through empirical distribution function based on $X_1,\dots,X_m$ and observations $X_{m+1},\dots,X_{m+k}$. Different test procedures for our problem based on empirical distribution functions were proposed by [11] and [2]. In these papers there are rather strict restrictions on N and independent observations are assumed. The paper [12] focuses the sequential detection of a change in the error distribution in time series. The studied procedure is based on empirical distribution functions of residuals. One can develop rank based procedures along the above lines but we do not pursue it here. Certain class of rank based are considered in [12] while U-statistics based sequential procedures are studied in [5] and [6].

The rest of the paper is organized as follows. Section 2 contain theoretical results together with discussions. Section 3 presents results of a simulation study. The proofs are in Section 4.

2. Main Results

Here we formulate assertions on limit behavior of our test procedure under both null hypothesis as well under some alternatives and discuss various consequences. Under the null hypothesis we consider two sets of assumptions:

(H_1) $\{X_i\}_i$ are independent identically distributed (i.i.d.) random variables, X_i has continuous distribution function F_0.

(H_2) $\{X_i\}_i$ is a strictly stationary α- mixing sequence with $\{\alpha(i)\}_i$ such that for all $\delta > 0$

$$P(|X_1 - X_{1+i}| \leq \delta) \leq D_1\delta, \quad i = 1, 2, \ldots, \tag{6}$$

$$\alpha(i) \leq D_2 i^{-(1+\eta)3}, \quad i = 1, 2, \ldots \tag{7}$$

for some positive constants η, D_1, D_2. X_i has continuous distribution function F_0. Here the coefficient $\alpha(i)$'s are defined as

$$\alpha(i) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|$$

where sup is taken over $A \in \sigma(X_j,\, j \leq n)$ and $A \in \sigma(X_j,\, j \geq n+i)$.

Next the assertion on limit behavior of the functional of $Q(m,k)$ under H_0 is stated.

Theorem 1
(I) Let the sequence $\{X_i\}_i$ fulfill the assumption (H_1) and put $\widehat{\sigma}_m^2 = 1/12$. Then

$$\lim_{m\to\infty} P\left(\sup_{1\leq k<N} \frac{|Q(m,k)|}{q_\gamma(k/m)} \leq x\right) = P\left(\sup_{0\leq t\leq 1} \frac{|W(t)|}{t^\gamma} \leq x\right) \tag{8}$$

for all x, where $q_\gamma(\cdot)$ is defined in (5) and $\{W(t);\, 0 \leq t \leq 1\}$ is a Wiener process.

(II) Let the sequence $\{X_i\}_i$ fulfill the assumption (H_2) and let, as $m \to \infty$,

$$N(m)/m \to \infty, \quad (\log N(m))^2/m \to 0. \tag{9}$$

Moreover, let estimator $\widehat{\sigma}_m$ be such that, as $m \to \infty$, $\widehat{\sigma}_m^2 - \sigma^2 = o_P(1)$, where

$$\sigma^2 = \frac{1}{12} + 2\sum_{j=1}^{\infty} cov\{F_0(X_1), F_0(X_{j+1})\}.$$

Then (8) holds true.

Concerning alternatives we consider either of the following setups:

(A_1) $\{X_i\}_i$ are independent random variables, X_i has continuous distribution function F_0 for $i = 1, \ldots, m+k^*$ and F^0 for $i = m+k^*+1, \ldots$, such that $\int F_0(x)\,\mathrm{d}F^0(x) \neq 1/2$.

(A_2) For some integer $k^* \leq N\eta$, $\eta \in [0,1)$ $\{X_i\}_{i=1}^{m+k^*}$ is a strictly stationary α-mixing sequence with $\{\alpha_0(i)\}_i$ with continuous distribution function F_0 and satisfying (6) and (7). Given $X_1, \ldots, X_{m+k^*}$ the sequence $\{X_i\}_{m+k^*}^{\infty}$ is a strictly stationary α- mixing sequence with $\{\alpha^0(i)\}_i$ with continuous distribution function F^0 and such that for all $\delta > 0$

$$P(|X_{m+k^*+1} - X_{m+k^*+1+i}| \leq \delta) \leq D_3\delta, \quad i = 1, 2, \ldots, \tag{10}$$

$$\alpha^0(i) \leq D_4 i^{-(1+\kappa)3}, \quad i = 1, 2, \ldots \tag{11}$$

for some positive constants κ, D_3, D_4. Also $\int F_0(x)\,\mathrm{d}F^0(x) \neq 1/2$ is assumed.

Alternative hypotheses cover a change in parameters like location but also a change in the shape of distribution. Additionally, alternative $(A.2)$ is sensitive w.r.t. a change in dependence among observations.

Theorem 2 *Let $\{X_i\}_i$ fulfill either (A_1) or (A_2), let $k^* < N\eta$ for some $0 \geq \eta < 1$, let (5) be satisfied. Then, as $m \to \infty$,*

$$\sup_{1 \leq k < N} \frac{|Q(k, m)|}{q_\gamma(k/m)} \xrightarrow{P} \infty.$$

Proofs of both theorems are postponed to Section 4.

Theorem 1 provides approximation for critical value c so that the test procedure fulfills (2) under the null hypothesis (H_1) or (H_2), i. e., c is the solution of the equation

$$P\left(\sup_{0 \leq t \leq 1} \frac{|W(t)|}{t^\gamma} \leq c\right) = 1 - \alpha. \tag{12}$$

Notice that under (H_1) the test procedure is distribution free and hence approximation for c can be obtained by simulation for arbitrary continuous F_0.

Both theorems certainly hold under more general assumptions but their proofs become much more technical and quite long.

The basic idea of the proof under the null hypothesis is to show that the limit distribution of the process $\{V_m(t),\, t > 0\}$, where

$$V_m(t) = \frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+\lfloor mt \rfloor} (\widehat{F}_m(X_i) - 1/2)$$

is the same as of $\{Z_m(t),\, t > 0\}$ with

$$Z_m(t) = \frac{1}{\sqrt{m}} \Big(\sum_{i=m+1}^{m+\lfloor mt \rfloor} (F_0(X_i) - 1/2) - \frac{k}{m} \sum_{j=1}^{m} (F_0(X_j) - 1/2) \Big).$$

Moreover, as $m \to \infty$ the process $\left\{ \frac{1}{\sqrt{m}} \Big(\sum_{i=m+1}^{m+\lfloor mt \rfloor} (F_0(X_i) - 1/2),\, t > 0 \right\}$ converges to a Gaussian process in a certain sense and $\frac{1}{\sqrt{m}} \sum_{j=1}^{m} (F_0(X_j) - 1/2)$ converges in distribution to $N(0,\, \sigma^2)$, where

$$\sigma^2 = \frac{1}{12} + 2 \sum_{j=1}^{\infty} cov\{F_0(X_1), F_0(X_{j+1})\}.$$

In case of independent observations $\sigma^2 = 1/12$ while for dependent ones the second term in σ^2 is generally nonzero and also unknown. As an estimator of σ^2 we use the estimator

$$\widehat{\sigma}_m^2 = \widehat{R}(0) + 2 \sum_{k=1}^{\Lambda_m} w(k/\Lambda_m) \widehat{R}_m(k), \tag{13}$$

$$\widehat{R}_m(k) = \frac{1}{n} \sum_{i=1}^{n-k} (\widehat{F}_m(X_i) - 1/2)(\widehat{F}_m(X_{i+k}) - 1/2), \tag{14}$$

where $w(\cdot)$ is a weight function. Usual choices are either

$$w_1(t) = 1I\{0 \le t \le 1/2\} + 2(1-t)\{1/2 < t \le 1\}$$

or

$$w_2(t) = 1 - tI\{0 \le t \le 1\}.$$

The weight $w_1(\cdot)$ is called the flat top kernel, while $w_2(\cdot)$ is the Bartlett kernel.

Theorem 3 *Let the sequence $\{X_i\}_i$ fulfill the assumption (H_1) and let*

$$\Lambda_m \to \infty, \quad \Lambda_m (\log m)^{-\beta} \to 0$$

for some $\beta > 2$. Then, as $m \to \infty$,

$$\widehat{\sigma}_m^2 - \sigma^2 = o_P(1).$$

Proof. It is omitted since it very similar to the proof of Theorem 1 (II). □

3. Simulations

In this section we report the results of a small simulation study that is performed in order to check the finite sample performance of the monitoring procedure considered in the previous section. The simulations were performed using the R software.

All results are obtained for the level $\alpha = 5\%$ where the critical values c were set using the limit distribution as indicated in (12). Unfortunately the explicit form for the distribution of $\sup_{0 \le t \le 1} |W(t)|/t^\gamma$ is known only for $\gamma = 0$ otherwise the simulated critical values are used. They are reported in [8] for example. We choose three different length of the training data $m = 50$, 100 and 500 to asses the approximation based on asymptotics. The estimate $\widehat{\sigma}_m^2$ is set to 1/12 for independent observations and it is calculated according to (13) with flat top kernel for dependent ones. We also comment on a common situation when we do not have the apriori information about the independence and the estimate of σ_m^2 is calculated also for the independent observations. The symbol t_k stands for t-distribution with k degrees of freedom.

The empirical sizes of the procedure under the *null hypothesis* are based on 10 000 replications and monitoring period of length 10 000. They are reported in Table 1 for both independent and dependent observations, where dependent ones form an AR(1) sequence with a coefficient ρ. Since the procedure make use of the empirical distribution function it is convenient also for distributions with heavier tails. Two such examples are shown in the table, as well as a skewed distribution (demeaned Log-normal one). We use different values of a tuning constant γ and since we will later examine an early change, we are mostly interested in γ close to 1/2.

We can see that for independent observations the level is kept and the prolongation of the training period has no significant effect. This is not the case when we do not make use of the independence information (figures are not reported here). The reason is that we need more data to estimate σ^2 precisely enough and therefore the prolongation will bring the empirical size closer to the required level. Similar reasoning holds for dependent observations as well. For γ in question (0.49), the results are satisfactory. Typically, the results for more regular distributions (e. g. normal one) are better than those reported here.

TABLE 1
Empirical sizes for 5% level for different distribution of errors being either independent ($\rho = 0$) or forming AR(1) sequence with coefficient ρ.

		t_1				t_4				LN(0,1)-$e^{-1/2}$			
ρ	$m \setminus \gamma$	0	0.25	0.45	0.49	0	0.25	0.45	0.49	0	0.25	0.45	0.49
	50	4.7	4.5	2.9	1.7	4.4	4.3	2.8	1.7	4.3	4.1	3.0	1.7
0	100	4.6	4.7	3.4	2.2	4.7	4.3	3.2	2.0	4.7	4.3	3.1	2.0
	500	4.5	4.5	4.2	3.0	4.2	4.4	3.8	2.8	4.4	4.5	4.0	3.0
	50	9.4	9.0	6.7	4.6	8.6	8.7	6.6	4.6	9.0	8.8	6.7	4.6
0.2	100	7.5	7.6	5.7	3.8	6.6	6.4	5.3	3.8	7.5	7.5	5.8	3.9
	500	5.7	5.8	5.3	4.1	5.0	5.3	4.7	3.5	5.2	5.6	5.1	3.8
	50	12.1	12.2	8.7	5.6	10.3	10.4	7.6	5.2	11.0	10.9	7.9	5.4
0.4	100	10.9	11.0	8.3	5.9	9.0	9.3	6.9	4.8	8.9	8.8	6.4	4.2
	500	8.8	9.6	8.8	6.6	6.7	7.2	6.4	4.9	7.2	7.4	6.8	5.0

Now we focus on *alternatives*. We take $k^* = 0$, i. e. the change occurs right after the end of training period. Therefore we use $\gamma = 0.49$, which is the most convenient choice for an early change. The maximal length of the monitoring period is 500 and the number of replications is 2500.

Table 2 summarizes results for stopping times for independent observation when change is in the location with zero location before the change and μ^0 afterwards. For comparison there are $k^* = 0$ and also $k^* = 9$. The latter case leads to a small increase in the delay of detection, otherwise the results are analogous, so we will report only results for $k^* = 0$ onwards. The detection delays are quite small even for a smaller change. The prolongation of the training period leads mainly to reducing extremes of the delay. However when we do not have the apriori information about the independence i. e. the estimate of σ^2 need to be calculated, the delays are monotonically decreasing in m. The results are generally a bit worse even for the largest m (figures are not reported here). In some simulations where the max value equals to 500 the change was not detected, however this is quite rare in this setting.

The results for dependent observations are shown in Table 3. In the the upper part there are stopping times for a unit change in mean, when errors form AR(1) sequence. For dependent observations the positive impact of increased m is clearly visible. With an increasing dependence amongst the data, the performance of the procedure is worsening. However the results for $m = 500$ are satisfactory even with $\rho = 0.4$. The lower part of the table presents the results for change in distribution of innovations from t_4 to demeaned Log-normal one. The procedure detects the change for larger m, however the performance is not satisfactory. This pair of distributions was chosen because it fulfills the requirement on F_0 and F^0 as described in (A_1). That requirement excludes the possibility of change from a symmetric distribution to another symmetric one. Simulations confirmed that the procedure is insensitive to this type of change.

Table 4 shows the results for a change in variance of independent observations. Due to the requirement of (A_1) we choose two skewed distributions, Log-normal and χ^2_2 ones, which were again demeaned. We consider doubling either the variance or the standard deviation. The results are generally better for Log-normal distribution because it is more skewed. One can see an improvement in delay with an increasing m. A longer training period is crucial mainly for a smaller change.

TABLE 2
Summary of the stopping times for independent observations with different distributions when change in location of μ^0 occurs, $\widehat{\sigma}_m^2 = 1/12$ and $k^* = 0$ (if not stated otherwise).

		t_4			t_1			LN(0,1)-$e^{-1/2}$			t_4, k^*=9		
μ^0	$\backslash m$	50	100	500	50	100	500	50	100	500	50	100	500
	Min.	5	5	5	4	4	4	5	4	5	24	18	24
	1st Qu.	8	9	10	9	12	12	11	14	13	30	24	31
1	Median	11	13	13	16	20	19	13	18	15	34	27	34
	Mean	12	15	14	21	24	23	14	18	16	35	28	35
	3rd Qu.	15	18	17	28	31	29	16	22	18	39	32	39
	Max.	52	54	46	126	124	100	33	42	30	81	64	67
	Min.	5	5	6	4	4	4	5	4	7	25	18	27
	1st Qu.	14	19	18	18	25	22	35	52	32	44	36	45
0.5	Median	22	31	26	38	48	38	59	85	44	55	47	53
	Mean	27	36	29	60	69	46	73	99	45	60	52	55
	3rd Qu.	34	47	37	77	91	64	96	131	57	71	62	64
	Max.	153	197	110	500	500	250	464	500	128	205	214	124

TABLE 3
Summary of the stopping time for errors forming AR(1) process. Upper part – change in mean of +1 occurs, lower part-change in distribution of innovations, $k^* = 0$ for both.

		$\rho = 0$			$\rho = 0.2$			$\rho = 0.4$		
distribution	$\backslash m$	50	100	500	50	100	500	50	100	500
	Min.	5	5	5	2	4	11	3	6	10
	1st Qu.	8	9	10	37	34	32	56	47	42
t_4	Median	11	13	13	63	50	42	141	78	60
	Mean	12	15	14	120	61	45	234	125	66
	3rd Qu.	15	18	17	123	71	56	500	140	83
	Max.	52	54	46	500	500	168	500	500	365
	Min.	1	2	4	3	5	8	4	8	9
	1st Qu.	9	10	10	19	19	19	38	33	32
LN(0,1)-$e^{-1/2}$	Median	14	13	13	34	27	24	113	58	45
	Mean	23	15	13	90	38	26	230	116	51
	3rd Qu.	23	18	16	76	41	31	500	122	62
	Max.	500	75	30	500	500	67	500	500	324
	Min.	1	2	6	2	5	7	4	6	7
t_4	1st Qu.	59	49	43	201	106	70	500	311	109
$\downarrow$	Median	500	159	83	500	500	145	500	500	262
LN(0,1)-$e^{-1/2}$	Mean	328	247	106	383	339	193	423	395	283
	3rd Qu.	500	500	141	500	500	276	500	500	500
	Max.	500	500	500	500	500	500	500	500	500

4. Proofs

We focus on the proofs for independent observations and give modifications needed for dependent ones. The line of both proofs is the same, however for dependent observations it is more technical.

Proof of Theorem 1.
(I) The detector $Q(m, k)$ can be decomposed into two summands:

$$\widehat{\sigma}_m \sqrt{m} Q(m, k) = J_1(m, k) + J_2(m, k),$$

Table 4
Summary of the stopping time for independent observations with different distributions when a change in a standard deviation (multiplied by κ^0) occurs, $\widehat{\sigma}_m^2 = 1/12$ and $k^* = 0$.

	LN(0,1)-$e^{-1/2}$						$\chi_2^2 - 2$					
$\backslash\kappa^0$	2			$\sqrt{2}$			2			$\sqrt{2}$		
$\backslash m$	50	100	500	50	100	500	50	100	500	50	100	500
Min.	1	1	3	1	2	3	1	1	3	1	2	3
1st Qu.	14	14	14	56	35	33	25	25	24	500	119	60
Median	50	39	31	500	158	75	500	93	59	500	500	173
Mean	151	72	44	333	239	113	282	193	91	397	360	223
3rd Qu.	221	89	61	500	500	153	500	440	125	500	500	396
Max.	500	500	365	500	500	500	500	500	500	500	500	500

where

$$J_1(m,k) = \frac{1}{m} \sum_{i=m+1}^{m+k} \sum_{j=1}^{m} h(X_j, X_i),$$

$$J_2(m,k) = \sum_{i=m+1}^{m+k} (F_0(X_i) - 1/2) - k/m \sum_{i=1}^{m} (F_0(X_i) - 1/2)$$

with

$$h(X_j, X_i) = I\{X_j \le X_i\} - E(I\{X_j \le X_i\}|X_i) - E(I\{X_j \le X_i\}|X_j) + EI\{X_j \le X_i\}.$$

Since given $X_1, \dots, X_m$ term $J_1(m,k)$ can be expressed as the sum of independent random variables with zero mean and since for $i \ne j$

$$E(h(X_j, X_i)|X_i) = E(h(X_j, X_i)|X_j) = Eh(X_j, X_i) = 0$$

we get by the Hájek -Rényi inequality for any $q > 0$:

$$E\Big(P(\max_{1\le k\le N} \frac{|J_1(m,k)|}{\sqrt{m}(1+k/m)(k/(m+k))^{\gamma}} \ge q|X_1, \dots, X_m)\Big)$$

$$\le q^{-2} \sum_{k=1}^{N} \frac{E\Big(\sum_{j=1}^{m} h(X_j, X_i)\Big)^2}{m^3(1+k/m)^2(k/(m+k))^{2\gamma}}$$

$$\le q^{-2} D\Big(\sum_{k=1}^{m} m^{-2+2\gamma} k^{-2\gamma} + \sum_{k=m+1}^{N} k^{-2}\Big) = q^{-2} O(m^{-1})$$

for some $D > 0$. The last relation holds true for any N integer and therefore the limit behavior of $\max_{1\le k\le N} \frac{|Q(m,k)|}{q_\gamma(k/m)}$ is the same as $\max_{1\le k\le N} \frac{|J_2(m,k)|}{\sqrt{m} q_\gamma(k/m)}$. The proof can be finished along the line of Theorem 2.1 in [8].

(II) The proof follows the same line as above but due to dependence modifications are needed. Notice that α-mixing of $\{X_i\}_i$ implies α-mixing of $\{\phi(X_i)\}_i$ for any measurable function ϕ with the same mixing coefficient as the original sequence. Then by Lemma 3.3 in [4] we get that there is a positive constant D such that for $h(\cdot,\cdot)$ defined above

$$|E(h(X_{i_1}, X_{i_2}) h(X_{i_3}, X_{i_4})| \le D(\alpha(i))^{2/3-\xi}$$

for any $\xi > 0$, where $i = \min(i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)})$ with $i_{(1)} \le i_{(2)} \le i_{(3)} \le i_{(4)}$. Then after some standard calculations we get that

$$EJ_1(m,k)^2 \le Dmk$$

for some $D > 0$ and hence by Theorem B.4 in [9] we get that also under present assumptions

$$P(\max_{1\le k\le N} \frac{|J_1(m,k)|}{(1+k/m)(k/(m+k))^\gamma} \ge q) \le q^{-2} O(m^{-1}(\log N)^2).$$

The proof is then again finished along the line of Theorem 2.1 in [8] but instead of Komlós-Major-Tusnády results we use Theorem 4 in [10]. □

Proof of Theorem 2 Going through the proof of Theorem 1(I) we find that if in $J_2(m,k)$ we replace 1/2 by $EF(X_i)$ and denote this by $J_2^A(m,k)$ then even under our alternative

$$\max_{1\le k\le N} \frac{|J_2^A(m,k)|}{\sqrt{m} q_\gamma(k/m)} = O_P(1), \quad \max_{1\le k\le N} \frac{|J_1(m,k)|}{\sqrt{m} q_\gamma(k/m)} = o_P(1).$$

Moreover,

$$\max_{1\le k\le N} \frac{|\max(0, k-k^*)|}{\sqrt{m} q_\gamma(k/m)} \to \infty.$$

To prove part (II) we proceed similarly. □

References

[1] Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-point monitoring in linear models. *Econometrics Journal* **9** 373–403.

[2] Bandyopadhyay, U. and Mukherjee, A. (2007). Nonparametric partial sequential test for location shift at an unknown time point. *Sequential Analysis* **26** 99-113.

[3] Chu, C.-S., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica* **64** 1045–1065.

[4] Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *Journal of Multivariate Analysis* **101** 126–137.

[5] Gombay, E. (1995). Nonparametric truncated sequential change-point detection. *Statistics & Decisions* **13** 71–82.

[6] Gombay, E. (2004). U-statistics in sequential tests and change detection. *Sequential Analysis* **23** 254–274.

[7] Gombay, E. (2008). Weighted logrank statistics in sequential tests. *Sequential Analysis* **27** 97–104.

[8] Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference* **126** 225–251.

[9] Kirch C. (2006). Resampling Methods for the change analysis of dependent data. *PhD. Thesis, University of Cologne, Cologne.* http://kups.ub.uni-koeln.de/volltexte/2006/1795/.

[10] Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *Ann. Probab.* **8** 6 1003–1036.

[11] Lee S., Lee Y. and Na O. (2009). Monitoring distributional changes in autoregressive models. *Commun. Statist. Theor. Meth.* **38** 2969-2982.

[12] Lee S., Lee Y. and Na O. (2009). Monitoring parameter change in time series. *Journal of Multivariate Analysis* **100** 715–725.

[13] Mukherjee, A. (2009). Some rank-based two-phase procedures in sequential monitoring of exchange rate. *Sequential Analysis* **28** 137-162.