A nonparametric change-point smooth test for dependent sequences

M Boutahar, Laurence Reboul

A nonparametric change-point smooth test for dependent sequences

M. BOUTAHAR[1] and L. REBOUL[1,*]

[1] *Institut de Mathematiques de Marseille, Campus de Luminy, 13288 Marseille cedex 9.*
E-mail: mohamed.boutahar@univ-amu.fr; *laurence.reboul@univ-amu.fr

Abstract In this paper we propose a change-point test for a strictly stationary short-memory sequences. Several cases of dependence are investigated among whose strong mixing and θ-dependence. A simulation study show the good performances of the method.

Keywords: Smooth test, Schwarz's rule, Strictly stationary process, Strong mixing, θ-dependence.

1. Introduction

Change-points are discontinuity times in the distribution of a time series. Change-point detection in time series, is a problem of great interest in many application areas. For instance, economic, financial or climatic time series are often affected by structural changes in forces induced by external or internal events (see e.g. [2], [25], [24]). Hence, since the pioneer paper [40], the problem of testing for changes points (structural changes) in the distribution of a time series have received a lot of attention in statistics. From a general point of view, change-point tests address the question whether a stochastic process is stationary during the entire observation period or not. More precisely, let us consider the observation $(X_1, \ldots, X_n)$ of a time phenomenon modeled by strictly stationary discrete time process X. We want to test:

$$H_0 : \exists\, f_1 \,(X_i)_{1\leqslant i\leqslant n} \text{ have common marginal density } f_1, \tag{1}$$

versus

$$H_1 : \exists\, (f_1, f_2),\ f_1 \neq f_2\ 0 < \theta < 1, \begin{cases} (X_i)_{1\leqslant i\leqslant [n\theta]} & \text{have marginal density} f_1 \\ (X_i)_{[n\theta]+1\leqslant i\leqslant n} & \text{have marginal density} f_2. \end{cases}$$

When $(X_1, \ldots, X_n)$ is a sequence of independent and identically distributed random variables, this testing problem has given rise to a huge literature (see [11] and [9] for accounts of various methods). Most of these works focus on the shift of a particular distribution parameter, such as mean or variance, neglecting more subtle changes. Moreover, most time series exhibit some kind of temporal dependence so that the tests developed in the iid context need suitable modifications to be adapted to the non iid case. The effect of dependence on change-point test procedures has been first addressed in [5] (see also Tang and MacNeill (1993), Antoch, Hušková, and Prášková (1997)). In the literature, the tests developed for uni- variate/multivariate time series include Horváth, Kokoszka, and Steinebach (1999), Vogelsang (1998, 1999) for a change in the mean; Inclan and Tiao (1994), Lee and Park (2001), Gom- bay, Horváth, and Hušková (1996) for a change in the mar- ginal variance; Giraitis, Leipus, and Surgailis (1996) and In- oue (2001) for a change in the marginal distribution function; Picard (1985), Giraitis and Leipus (1992), and Lavielle and Lu- dena (2000) for a change in the spectrum; Berkes, Gombay, and Horváth (2009), Galeano and Peña (2007) for a change in the autocovariance function at certain lags. For change-point detection in time series models or regression models with de- pendent errors, see Andrews (1993), Davis, Huang, and Yao (1995), Lee, Ha, and Na (2003), Ling (2007), Qu and Perron (2007), Aue et al. (2008), Gombay (2008), among others. We refer the interested readers to the excellent review articles by Kokoszka and Leipus (2002) and Perron (2006) for more refer- ences. It seems that the techniques developed for change-point detection are specific to the quantity/parameter of interest. For example, the test for a change in variance is quite different from the test for a change in spectrum. This brings considerable dif- ficulty to the practitioners, who want to use these tests routinely to check the stability of certain characteristics of a time series at hand. see Perron (2006) and Aue and Horva?th (2013) for recent reviews on change point tests for time series (for the mean)

Nonparametric tests for change points in the whole marginal distribution of strictly stationary processes under particular structures of temporal dependence have been proposed, among others by [34], [45], [23], [33], [35], [32]. Most of these tests are inspired from empirical distribution function based goodness of fit tests (such as Cramér-Von-Mises or Kolmogorov-Smirnov) or use kernel density estimates of the tested densities. In nonparametric goodness of fit tests theory, [39]'s type smooth tests have been shown to be very serious competitors of the above mentioned methods (see e.g. [20], [21], [28],[27]). Namely, such tests demonstrate better performance at detecting more subtle differences than constant location/scale shifts.

For the two-sample testing problem $H_0 : f_1 = f_2$, [18] recently proposed a data driven test for comparing the marginal densities of a bivariate strictly stationary dependent process X under different assumptions on the dependence structure of X, including long or short memory structures. Extending [29], [22], [26], and [37], the test statistics proposed by [18] is based on a coefficient's comparison of densities expansions along an appropriated orthogonal basis. The number of coefficients to compare is chosen by a data-driven method (see [30]). The authors derive the asymptotic distribution of their test statistic and show that the test based on the limiting quantile is consistent under contiguous alternatives. A simulation study enhances good performances in the short memory case, especially for strong-mixing X. In this paper, we use a generalization of [18]'s test strategy together with weak invariance principle theorems in view to deal with the change-point detection problem in a dependent time series. This test is proposed in a very general setting, covering several kind of dependence structure of the time sequence. As particular cases, we apply our results to independent, strong-mixing and θ-dependent sequences. The rest of the paper is organized as follows. In Section 2 we introduce a data-driven test statistic for (1). In Section 3, we derive its limiting distribution and study the consistency of the test based on the limiting quantile under suitable alternatives, in a general setting. In Section 4 we apply our general methodology to independent, strong-mixing and θ-dependent sequences. Section 6 shows simulations results and Section 7 is devoted to the proofs.

2. A data-driven smooth test

Let X be a discrete time process defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in $\mathbb{R}$. Let ν be a given probability measure with density h with respect to some reference measure λ (Lebesgue's or counting measure for instance). Hereafter, all the marginal densities involved in the test are taken with respect to λ and are assumed to belong to the space $\mathbb{L}_2(\nu)$ of square integrable functions with respect to ν.
In this setup, we wish to test (1) based on the observation $(X_1, \ldots, X_n)$ of X.

For that task, we consider the expansions of f_1 and f_2 along a dense family $(Q_j)_{j\in\mathbb{N}}$ of orthonormal functions in $\mathbb{L}^2(\nu)$:

$$f_1 = \sum_{j\geqslant 0} a_j(\theta)\, Q_j \quad \text{and} \quad f_2 = \sum_{j\geqslant 0} b_j(\theta)\, Q_j \ , \tag{2}$$

with

$$a_j(\theta) = \mathbb{E}(\widetilde{Q}_j(X_1)) = \int_{\mathbb{R}} Q_j(t) f_1(t) d\nu(t) \quad , \quad b_j(\theta) = \mathbb{E}(\widetilde{Q}_j(X_n)) = \int_{\mathbb{R}} Q_j(t) f_2(t) d\nu(t),$$

Here, $\mathbb{E}$ denotes the expectation and we set $\widetilde{Q}_j = hQ_j$ for all $j \in \mathbb{N}$ (recall that $\nu = h\lambda$). Then, (1) can be rewritten as

$$H_0 : \forall\, 0 \leqslant \theta \leqslant 1,\ \forall j > 0; a_j(\theta) = b_j(\theta) \quad \text{against} \quad H_1 : \exists\, 0 < \theta < 1,\ \exists\, j_\theta > 0\ a_{j_\theta}(\theta) \neq b_{j_\theta}(\theta). \tag{3}$$

In view to define a test strategy for (3), we shall consider for every $0 \leqslant \theta \leqslant 1$ a sequence $(U_n(k,\theta)_{k\geqslant 1}$ of Neyman's type test statistics. They are defined, up to a renormalization factor, as the sum of the squared differences between the empirical estimators of the k first coefficients in (2). More formally, setting for all $j \geqslant 1$

$$V_n^{(j)}(\theta) = \frac{[n\theta](n - [n\theta])}{n^{3/2}} (\widehat{a}_j(\theta) - \widehat{b}_j(\theta)),$$

$$\widehat{a}_j(\theta) = \begin{cases} 0 & \text{if } 0 \leqslant \theta < n^{-1} \\ \dfrac{1}{[n\theta]} \displaystyle\sum_{t=1}^{[n\theta]} \widetilde{Q}_j(X_t) & \text{if } \theta \in [n^{-1}, 1] \end{cases} \qquad \widehat{b}_j(\theta) = \begin{cases} \dfrac{1}{n-[n\theta]} \displaystyle\sum_{t=[n\theta]+1}^{n} \widetilde{Q}_j(X_t) & \text{if } \theta \in [0,1) \\ 0 & \text{if } \theta = 1 \end{cases}$$

we define for all $k \geqslant 1$

$$U_n(k,\theta) = \sum_{j=1}^{k} (V_n^{(j)}(\theta))^2. \tag{4}$$

For given $\theta \in (0,1)$, $U_n(k,\theta)$ is a proper statistic for testing that $a_j(\theta) = b_j(\theta), j = 1, \ldots, k$. More precisely, each component of $U_n(k,\theta)$ is used to compare a_j and b_j in order to detect H_1. Thus we need to let k tend to infinity to detect all possible alternatives. However, choosing too large k tends to power dilution of the test.
Hereafter, we follow the idea of [30] to make a data driven selection of k. In the case of independent observations, [30] first introduced such a rule which has been shown to be very successful further on (see among others [29], [27], [1], [31]). It consists of a modified version of Schwarz's Bayesian information criterion (see [44]), based on an expansion of the maximum likelihood function (for an heuristic justification of this rule in the short-range dependent case and the two-sample testing problem, see Remark 1 of [18].)

Namely, let us consider a strictly positive control sequence of integers d_n such that $\lim_{n\to\infty} d_n = \infty$. At first step, we select among all possible dimensions between 1 and d_n the minimum maximizer of the information criterion $K_{\theta,n}$, as follows:

$$K_{\theta,n} = \min\left\{k : 1 \leqslant k \leqslant d_n, U_n(k,\theta) - k\log n \geqslant U_n(j,\theta) - j\log n, 1 \leqslant j \leqslant d_n\right\}. \tag{5}$$

Once $K_{\theta,n}$ is determined, we define our main test statistic as

$$\widetilde{U}_n(\theta) = U_n(K_{\theta,n}, \theta). \tag{6}$$

Since θ is unknown in (3), we propose to use as a test statistic an obvious increasing function of the maximum of the $\widetilde{U}_n(\theta)$'s over $\theta \in [0,1]$. Here, we choose

$$\widetilde{U}_n = \sup_{0\leqslant\theta\leqslant 1} \sqrt{\widetilde{U}_n(\theta)}, \tag{7}$$

For practical purposes, notice that $\widetilde{U}_n(\theta)$ is a left hand continuous step function such that $\widetilde{U}_n(\theta) = 0$ for every $\theta < n^{-1}$, $\widetilde{U}_n(1) = 0$ with jumps at $\theta = i/n, i = 1, \ldots, n$ so that the computation of the optimum requires to compare at most n values.

3. Large sample properties

In this section, we first study the asymptotic distribution under H_0 of $\widetilde{U}_n$, under broad enough conditions to cover most of the usual short memory structures of dependence of the underlying process X. We then prove the consistency of our test against a large class of alternatives.

Our test strategy is then indicated in a last paragraph. We first give assumptions under which these results hold.

3.1. Assumptions

From now on, we set for all $k \geqslant 1$, $\widetilde{Q}_k^*(X) = \widetilde{Q}_k(X) - \mathbb{E}_0(\widetilde{Q}_k(X))$ where $\mathbb{E}_0$ is the expectation under H_0 and we denote by $(r_n(k))_{n>0}$ the partial sums of the absolute series of its auto-covariances :

$$r_n(k) = \sum_{t=0}^{n-1} \left|\mathbb{E}_0\big(\widetilde{Q}_k^*(X_0)\widetilde{Q}_k^*(X_t)\big)\right|, \ n = 1, 2, \ldots, \tag{8}$$

Moreover, we denote by $D[0,1]$ the Skorohod space of right continuous with left-hand limits functions from $[0,1]$ into $\mathbb{R}$ endowed with Skorohod's topology and we use $\overset{\mathcal{D}}{\Longrightarrow}$ to indicate weak convergence on $D[0,1]$. Moreover, we define $W_n : \Omega \to D[0,1]$ by

$$W_n(\theta) = \frac{[n\theta]}{n} a_1(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\theta]} \widetilde{Q}_1^*(X_t),\ \theta \in [0,1]$$

Hereafter, we make the following assumptions:

(A): $\dfrac{1}{d_n} \sum_{k=1}^{d_n} r_n(k) = O(1)$.

(B): $W_n(\theta) \overset{\mathcal{D}}{\Longrightarrow} \sigma W(\theta)$ under H_0, where $\{W(\theta), \theta \in [0,1]\}$ denotes a standard Brownian Motion and σ is a positive constant.

Let us describe further on the meaning and implications of these assumptions. First, notice that the behavior of $(r_n(k))_{n>0}$ describes the range of dependence of the process $\widetilde{Q}_k^*(X)$: it is short-range dependent when the sequence $(r_n(k))_{n>0}$ is bounded while unbounded sequences characterize long-range dependent processes (see e.g. [6]). Assumption **(A)** thus informs us about the asymptotic behavior of the partial sums $r_n(k)$ for a fixed k, such assumption generally holds if the processes $\widetilde{Q}_k^*(X)$, $k \geqslant 1$ have short memory as soon as their spectral densities are bounded at zero.

Assumption **(B)** amounts to say that the partial sums of $\widetilde{Q}_1^*(X)$ satisfy a Donsker's type invariance principle. Such a well known result in the i.i.d. setting generally still hold under suitable moments assumptions for certain classes of short range dependent processes. We will see in the next subsection that the asymptotic behavior under the null of $\widetilde{U}_n(\theta)$ is that of $\left(V_n^{(1)}(\theta)\right)^2$. To give an idea of this behavior, notice that simple calculations yield

$$V_n^{(1)}(\theta) \quad = \quad W_n(\theta) - \frac{[n\theta]}{n} W_n(1). \tag{9}$$

so that $V_n^{(1)}(\theta) \overset{\mathcal{D}}{\Longrightarrow} \sigma B(\theta)$, where B is a standard Brownian Bridge. Thus, when **(B)** holds, we have the convergence of $\widetilde{U}_n$ to the supremum of a Brownian Bridge.

3.2. Asymptotic distribution under the null

Hereafter, we derive the asymptotic distribution of $\widetilde{U}_n$ under the null. Theorem 1 below claims that under the null hypothesis, only the first component in $(U_n(k,\theta))_{1 \leqslant k \leqslant d(n)}$ is required to construct a nontrivial test.

Theorem 1. *Let $d_n = o(\log(n))$ and assume that **(A)** holds. Thus, under H_0, one has for every θ, $0 \leqslant \theta \leqslant 1$,*

$$K_{\theta,n} \xrightarrow{\mathcal{P}} 1. \tag{10}$$

*Moreover, if **(B)** holds*

$$\widetilde{U}_n \xrightarrow{\mathcal{L}} \sigma \widetilde{U}_\infty \tag{11}$$

where $\sigma > 0$, $\widetilde{U}_\infty = \sup_{\theta \in [0,1]} |B(\theta)|$ and $B(\theta)$ is a standard Brownian Bridge.

The cumulative distribution function of $\widetilde{U}_\infty$ is given by (see equation (11.39) of [7])

$$F_{\widetilde{U}_\infty}(z) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k \exp(-2k^2 z^2). \tag{12}$$

3.3. Convergence under contiguous alternatives

In the sequel, we show that for suitable alternatives, the test based on the limiting quantile is consistent. Namely, let us set $\delta_k(\theta) = a_k(\theta) - b_k(\theta)$ and consider the following two alternatives:

$$\begin{aligned} H_1^* &: \forall k \geqslant 1, \delta_k(\theta) = o(1/\sqrt{n}), \\ H_1^{**} &: \exists\, K > 1 \text{ such that } \forall\, k \neq K, \delta_k(\theta) = o(1/\sqrt{n}) \\ &\quad \text{and } \delta_K(\theta) = O(\sqrt{\log n}/n^{\beta-1/2}), \text{ with } 0 < \beta < 1. \end{aligned}$$

Denoting by $\mathbb{E}_1$ and $\mathbb{P}_1$ the expectation and probability under the corresponding alternative we consider the two following assumptions:

(A*): There exists some $C^* > 0$ such that

$$\frac{1}{d(n)} \sum_{s=1}^{d(n)} \mathbb{E}_1\big(|V_n^{(k)}(\theta) - \delta_k(\theta)|^2\big) < C^*,$$

(B*): $V_n^{(K)}(\theta) - \sqrt{n}\delta_K(\theta) = O_{\mathbb{P}_1}(\sqrt{\log(n)})$.

Theorem 2. *Let $d_n = o(\log(n))$ and assume that t **(A*)** holds. Then,*

- *Under H_1^*, $K_{\theta,n} \xrightarrow{\mathcal{P}} 1$. Moreover, if **(B)** holds $\widetilde{U}_n \xrightarrow{\mathcal{L}} \sigma\widetilde{U}_\infty$, where σ and $\widetilde{U}_\infty$ are defined as in Theorem 1.*
- *Under H_1^{**}, $K_{\theta,n} \xrightarrow{\mathcal{P}} K$. Moreover, if **(B*)** holds, $\widetilde{U}_n \xrightarrow{\mathcal{P}} +\infty$.*

Therefore, under H_1^*, the perturbation will not be detected by the test procedure while it will be detected under H_1^{**}.

3.4. Test strategy

In invariance principle results, the constant σ in **B** is generally such that $\sigma^2 = \lim_n W_n(1)$. Let $\widehat{\sigma}_n$ be a consistent estimator of σ, conveniently chosen according to the nature and the dependence structure of $\widetilde{Q}_1^*(X)$. Using the above mentioned results, we are able to construct a parameter free test based on the limiting distribution of $\widetilde{U}_n$.

In practice we will consider the following statistic, the asymptotic distribution of which is free of the parameter σ.

$$\widetilde{BR}_n = \sup_{0 \leqslant \theta \leqslant 1} \sqrt{\widetilde{BR}_n(\theta)}, \tag{13}$$

where

$$\widetilde{BR}_n(\theta) = BR_n(K_{n,\theta}, \theta), \text{ with } BR_n(k,\theta) = \frac{U_n(k,\theta)}{\widehat{\sigma}_n^2}$$

and $K_{n,\theta}$ is defined by (5). Thus, let q_α be the α-upper percentile point of the distribution of $\widetilde{U}_\infty$. Recall that its distribution is given by (12). Then, the test with critical region $\left\{\widetilde{BR}_n > q_\alpha\right\}$ has asymptotic level α.

4. Applications

In this section, we check the validity of Section 3's results on three types of short-memory structures of the process X. Subsection 4.1 deals with the independent case. In Subsections 4.2, and 4.3, we study

the strong mixing and θ-weak dependent cases. These cases are not exhaustives. The key point is that for short-memory processes X, processes $\widetilde{Q}_k(X)$ generally inherit the short-memory structure of X, for convenient classes of functions $\widetilde{Q}_k$ depending on the specific short-memory dependence structure of X (see Remark 1) so that **(B)** will generally be achieved by a weak invariance principle theorem and **(A)** will generally obtain under moment conditions controlling the order of magnitude of the partial sum $r_n(k)$ of the process $\widetilde{Q}_k(X)$.

The proofs are easily obtained by combining some results in [18] and known invariance principle theorems. So, they will only be sketched in the text. Extensions of these results are discussed in Subsection 4.4. Hereafter, we denote by $\|.\|_q$ denotes the $\mathbb{L}_q$ norm under H_0.

4.1. Independent sequences

We start with the basic case of independence. Let $(X_1, \ldots, X_n)$ be an i.i.d. sequence. Set

$$\sigma^2 = \|\widetilde{Q}_1^*(X_0)\|_2^2 = \mathbb{E}_0(\widetilde{Q}_1^*(X_0)^2), \tag{14}$$

and assume that $\sigma^2 > 0$. Then,

Corollary 1. *Let $d_n = o(\log n)$ and assume that there exist $C > 0$ and $n_0 > 0$ such that for all $n > n_0$,*

(i) $\dfrac{1}{d_n}\displaystyle\sum_{k=1}^{d_n} \|\widetilde{Q}_1^*(X_0)\|_2^2 < C.$

Then, under H_0

$$\widetilde{U}_n \xrightarrow{\mathcal{L}} \sigma\widetilde{U}_\infty,$$

with $\widetilde{U}_\infty$ defined as in Theorem 1.

Proof. Since for given $k > 0$, $\mathbb{E}_0(\widetilde{Q}_k(X_0)\widetilde{Q}_k(X_t)) = 0$ for all $t > 0$, we have $r_n(k) = \|\widetilde{Q}_1^*(X_0)\|_2^2$, with $r_n(k)$ defined by (8), so that **(A)** is equivalent to **(i)** and **(B)** straightforwardly arises from Donsker's theorem.

□

4.2. α-mixing sequences

We assume in the sequel that X satisfies a mixing condition. More specifically, we assume here strong or α-mixing in [43]'s sense. Formally, setting $\mathcal{P} = \sigma(X_t, t \leqslant 0)$, $\mathcal{F} = \sigma(X_t, t \geqslant m)$ and defining the decreasing sequence $(\alpha(m))_{m>0}$ of strong mixing coefficients of X by

$$\alpha(m) = \sup_{A\in\mathcal{P}, B\in\mathcal{F}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

we say that Z is α-mixing if

$$\lim_{m\to+\infty} \alpha(m) = 0.$$

A lot of classical models satisfy this condition. For instance, linear stochastic processes are α-mixing, as soon as the error process has a Lebesgue-integrable characteristic function (see [47]).

Let us set

$$\sigma^2 = \sum_{-\infty}^{\infty} \mathbb{E}_0\big(\widetilde{Q}_1^*(X_0)\widetilde{Q}_1^*(X_t)\big)$$

and assume that $\sigma^2 > 0$.Then,

Corollary 2. *Let $d_n = o(\log n)$ and assume that X is a strictly stationary α-mixing process with non-increasing coefficients sequence $(\alpha(m))_{m>0}$ and such that*

(i) *There exist* $C > 0$, $n_0 > 0$ *and* $\delta \geqslant 0$ *such that for all* $n > n_0$,

$$\frac{1}{d_n}\sum_{k=1}^{d_n} \|\widetilde{Q}_k^*(X_0)\|_{2+\delta}^2 < C.$$

(ii) $\sum_{m>0} m^{2/\delta}\alpha(m) < \infty$.

Then, under H_0

$$\widetilde{U}_n \xrightarrow{\mathcal{L}} \sigma\widetilde{U}_\infty,$$

with $\widetilde{U}_\infty$ *defined as in Theorem 1.*

Proof. In order to check **(A)** we first notice that $(\widetilde{Q}_k(X_t))$ is α-mixing since it can be expressed as a measurable function of X. Then, as in [18]'s proof of corollary 2, we use [42]'s covariance inequality given in Theorem 1.1 altogether with the former remark to show that

$$\lim_{n\to\infty} r_n(k) \leqslant K\|\widetilde{Q}_k^*(X_0)\|_{2+\delta}^2 \left(\sum_{t\geqslant 0}(t+1)^{2/\delta}\alpha(t)\right)^{\delta/(2+\delta)}, \tag{15}$$

with $K = 4\exp(2/(2+\delta))$. So, **(A)** obtains by **(i)**-**(ii)**.
Assumption **(B)** straightforwardly follows from [16]'s invariance principle Theorem 1 under **(i)** and **(ii)** (which imply condition (1.2) of the theorem) and $\sigma^2 > 0$. □

4.3. θ-dependent sequences

Mixing properties are satisfied by fairly general models, but they are not easy to check and sometimes do not hold. Namely, they do not cover the case of linear processes with discrete innovations (see e.g. [3]). A less restrictive and easier to prove condition is to assume that X admits one of the weak dependence conditions studied in [15]. We focus here on θ-dependence (see [13]). Let us define the Lipschitz modulus of a function g from $\mathbb{R}^d$ into $\mathbb{R}$ by

$$Lip(g) = \sup_{x\neq y}\frac{|g(x)-g(y)|}{\|x-y\|_1},$$

where $\|x\|_1 = \sum_{i=1}^d |x_i|$. For $(u,v) \in \mathbb{N}^*\times\mathbb{N}^*$, let $\mathcal{F}_u$ and $\mathcal{G}_v$ respectively denote the set of bounded by 1 measurable functions from $(\mathbb{R}^2)^u \doteq \mathbb{R}^2\times\ldots\times\mathbb{R}^2$ (u times) into $\mathbb{R}$ and the set of bounded by 1 functions from $(\mathbb{R}^2)^v$ into $\mathbb{R}$ with finite Lipschitz modulus. We say that X is θ-dependent if for all (u,v) and all sets of indices $i_1 \leqslant \ldots \leqslant i_u \leqslant j_1 - m \leqslant j_1 \leqslant \ldots \leqslant j_v$, we have for all $f \in \mathcal{F}_u$, $g \in \mathcal{G}_v$,

$$|cov(f(P), g(F))| \leqslant v Lip(g)\theta(m), \tag{16}$$

where $P = (X_{i_1},\ldots,X_{i_u})$, $F = (X_{j_1},\ldots,X_{j_v})$ and $(\theta(m))_{m>0}$ is a decreasing sequence such that

$$\lim_{m\to+\infty}\theta(m) = 0.$$

As previously, we set

$$\sigma^2 = \sum_{t=-\infty}^{+\infty}\mathbb{E}_0(\widetilde{Q}_1^*(X_0)\widetilde{Q}_1^*(X_t))$$

and we assume that $\sigma^2 > 0$. Then, we have

Corollary 3. *Let* $d_n = o(\log(n))$ *and assume that* X *is a strictly stationary* θ*-dependent process with non-increasing coefficients sequence* $(\theta(m))_{m>0}$ *and such that*

(i) *There exist* $C > 0$, $n_0 > 0$ *and* $\delta \geqslant 0$ *such that for all* $n > n_0$,

$$\frac{1}{d_n}\sum_{k=1}^{d_n}\|\widetilde{Q}_k^*(X_0)\|_{2+\delta}^{(2+\delta)/(1+\delta)} < C.$$

(ii) $\sum_{m>0} m^{1/\delta}\theta(m) < \infty$.

(iii) *The functions* $(Q_k^*)_{1\leqslant k}$ *are Lipschitz.*

Then, under H_0

$$\widetilde{U}_n \xrightarrow{\mathcal{L}} \sigma\widetilde{U}_\infty,$$

with $\widetilde{U}_\infty$ *defined as in Theorem 1.*

Proof. First note that **(iii)** implies the θ-dependence of $(\widetilde{Q}_k(X_t))$, with coefficients sequence $\theta_k(r) \leqslant \theta(r)$. Then, as in [18]'s proof of corollary 3, we use the covariance inequality of [12]'s Proposition 1 and follow the proof of [12]'s Lemma 2 to show that

$$\lim_{n\to\infty} r_n(k) \leqslant K\|\widetilde{Q}_k^*(X_0)\|_{2+\delta}^{(2+\delta)/(1+\delta)} \left(\sum_{t\geqslant 0}(t+1)^{1/\delta}\theta(t)\right)^{\delta/(1+\delta)}, \tag{17}$$

with $K = 2^{(4+\delta)/(2+\delta)}$, so that **(A)** obtains by **(i)**-**(ii)**.
Under conditions **(i)** and **(ii)** and since $\sigma^2 \neq 0$, Assumption **(B)** straightforwardly follows from the invariance principle theorem stated in [12]'s Theorem 2 (after application of Lemma 2 and Corollary 1 of this paper). □

Remark 1. *Notice that Corollaries 2 and 3 hold because the process* $\widetilde{Q}_k^*(X)$ *has the same dependence structure as* X*. But while in the mixing case, only measurability of the* $\widetilde{Q}_k^*$*'s is required to achieve this property, the* $\widetilde{Q}_k^*$*'s need to be Lipschitz in the* θ*-dependent case.*

4.4. Extensions

The aforementioned applications are not exhaustive.

- Firstly, $\widetilde{Q}_k^*(X)$ has the same dependence structure as X as soon as $\widetilde{Q}_k^*$ is a measurable function for mixing structure or satisfy a Lipschitz condition for weak-dependent structures so that invariance principles for X are easily transposed to $\widetilde{Q}_k^*(X)$.
- Secondly, the results obtained for α-mixing and θ-dependent X may be straightforwardly extended to several other weak dependence structures, by adapting the assumptions of the Corollaries and the covariances inequalities used in the proofs. Convariance inequalities and invariance principle theorems for various weak dependence conditions (such as τ, τ' λ and θ - dependences) can be found in [19], [12] and [15] while analogue results under mixing conditions (such as ρ-mixing, Φ-mixing, ...) are studied in [7], [17], [41]. See also [36] for a survey on invariance principle results for short memory processes.
- Thirdly, this study only focusses on short memory processes. Invariance principles for long-memory processes do not hold in general, except for particular classes (gaussian or linear processes, see e.g. [10, 14] and references therein). The $\sqrt{n}$- normalization and gaussian limit no more hold. Moreover in general, no heredity conditions ensures the dependence structure of $\widetilde{Q}_k^*(X)$.

5. Practical implementation of the test

In view to perform the test, many parameters have to be fixed in practice : the family $(Q_j)_{j\geqslant 0}$ and an associated measure ν, the sequence (d_n) and an estimator of the variance of the limiting distribution of the test statistic. These choices have already been discussed in details in [18], so we just recall them and only mention few modifications. We also recall alternative test statistic allowing to improve in some cases the performances of the test.

Choice of d_n**.** The computation of the selection rule (5) requires the choice of a sequence d_n. It can be seen on simulations that empirical levels and powers do not depend on d_n for sufficiently large values of this parameter. Detailed comments on this choice can be found in [18].

Choice of the basis $(Q_j)_{j\geqslant 0}$ and of the reference measure. The definition of the statistic relies on the choice of $(Q_j)_{j\geqslant 0}$ and of an absolute continuous measure ν such that $(Q_j)_{j\geqslant 0}$ is an orthogonal family of $\mathbb{L}_2(\nu)$ and f_1 and f_2 belong to $\mathbb{L}_2(\nu)$. This choice depends on the support of f_1. Relevant families that comply with the assumptions **(i)-(iii)** of Corollaries 1-3 are given in [18]. Namely, for X taking values in $\mathbb{R}$, we can choose the standard normal distribution for ν with its associated Hermite polynomials ; for X taking values values in $\mathbb{R}^+$, we can choose the exponential distribution with mean 1 for ν with its associated Laguerre polynomials ; for X taking bounded values, we can consider the uniform distribution for ν and its associated Legendre polynomials.

Choice of $\widehat{\sigma}_n$. The limiting distribution of the test statistic requires the estimation σ_n^2 of an unknown variance parameter σ^2. In the independent case, σ^2 is estimated by its empirical version $\widehat{\sigma}_n^2 = \dfrac{1}{n}\sum_{s=1}^{n}(V_s^{(1)})^2$. In the short-memory cases, classical estimators of the spectral density g of $\widetilde{Q}_1^*(X)$ may be used to estimate the long run variance σ^2. Usual choices are kernel based estimators (see e.g. [38], [4] and references therein). For linear processes, one can also consider the smoothed periodogram (see Brockwell and Davis (1991)). Several other estimators have been put forward for particular types of processes. In our simulations, we will use for instance the spectral estimator proposed in [8] for AR(1) processes.

However, in the change in mean case, it has been showed (see ...) that the power can decrease to zero as the magnitude of change ? gets very large, which is counterintuitive as we expect a test to be more powerful when the alternative gets far- ther away from the null. In the literature, this phenomenon was called nonmonotonic power (Vogelsang 1999) and its source has been identified through theoretical analysis and simulations by Vogelsang (1999), Deng and Perron (2008), and Juhl and Xiao (2009), among others. Heuristically, the decrease in power accompanied with larger shift is because the bandwidth ln is severely biased upward under the alternative, which leads to an inflation in the estimate of the scale s. When the scale estimate sn is too large, the KSn test statistic tends to be small, result- ing in a low rejection rate. Note that the fixed bandwidth (e.g., n1/3) is immune to the nonmonotonic power problem but is not adaptive to the magnitude of autocorrelations in the series and it could lead to severe size distortion; see Shao and Zhang (2010) for simulation evidence. Following the description of the SN idea in the previous section, a seemingly natural extension of the SN approach to the change point testing problem is to replace sn in KSn by Wn However, as shown in Shao and Zhang (2010), the above naive extension fails to attenuate the nonmonotonic power problem. The major problem with KSn is that it does not take into account the change-point alternative. To circumvent the prob- lem, Shao and Zhang (2010) proposed a new self-normalization process in order to estimate the long-run variance, taking into account the change point. Precisely, in order to test a break point in the time series Y_t, they used lobato(2001) to propose

$$V_n(l) = n^{-2}\left(\sum_{t=1}^{l}\left(S_{1,t} - \left(\frac{t}{l}\right)S_{1,l}\right)^2 + \sum_{t=l+1}^{n}\left(S_{t,n} - \left(\frac{n-t+1}{n-l}\right)S_{l+1,n}\right)^2\right)$$

where $S_{i,j} = \sum_{t=i}^{j} Y_t,\ i \leqslant j$. Here, we put $Y_t = Q_1^*(X_t)$ and use the test statistic

$$\widetilde{BR2}_n = \sup_{0\leqslant\theta\leqslant 1}\sqrt{\widetilde{BR}_n(\theta)} = \sup_{l=1,\dots n}\sqrt{\widetilde{BR}_n(l/n)}, \tag{18}$$

$$\widetilde{BR2}_n(\theta) = BR2_n(K_{n,\theta},\theta),\ \text{with}\ BR2_n(k,\theta) = \frac{U_n(k,\theta)}{V_n([n\theta])}$$

and $K_{n,\theta}$ is defined by (5). One has that

$$V_n([n\theta]) \Longrightarrow \sigma^2 V(\theta)$$

so that by the continuous mapping

$$\widetilde{BR2}_n \Longrightarrow \sup_{0\leqslant\theta\leqslant 1}\frac{|B(\theta)|}{\sqrt{V(\theta)}},$$

with

$$V(\theta)=\int_0^{\theta}\left(B(u)-\left(\frac{u}{\theta}\right)B(\theta)\right)^2du+\int_t heta^1\ B(1)-B(u)-\frac{1-u}{1-\theta}(B(1)-B(\theta))\Big)^2\,du$$

Thus, let q_α be the α-upper percentile point of the distribution of $\widetilde{U}_\infty$. Recall that its distribution is given by (12). Then, the test with critical region $\left\{\widetilde{BR}_n>q_\alpha\right\}$ has asymptotic level α.

Through extensive simulations, the SN-based test of KS has been shown to have superior size compared to some recently pro- posed alternatives by Crainiceanu and Vogelsang (2007) and Juhl and Xiao (2009), which have been developed to alleviate the nonmonotonic power problem. The power of the SN-based test was seen to be monotonic although there has been no rig- orous theory that justifies the nonmonotonic power property of the SN-based test. The extension to testing for a change point in other quantities, such as marginal quantiles and spectral distri- bution function at certain frequencies, have also been developed in Shao and Zhang (2010).

6. Simulation study

In order to evaluate the finite-sample performances of our test $\widetilde{BR}_n$ described in Subsection 3.4, we run Monte-Carlo simulations on order one gaussian autoregressive models and several sample sizes. For each model and sample size, we compute and compare the empirical levels and powers of $\widetilde{BR}_n$ with that of [34]'s test named $\widetilde{LN}_n$ in the sequel. The nominal level is fixed at $\alpha=5\%$. We present below the models and tests as well as the obtained results.

6.1. Models

The simulated examples are based on the observation of sequences of size $n\in\{50,100,200,500,1000\}$.

- In the first set of models, denoted by **Model 1** we consider α-mixing processes X:

$$X_t \;=\; \begin{cases} a_1X_{t-1}+\varepsilon_t & \text{if} \quad 1\leqslant t\leqslant[n/2] \\ b_0+b_1X_{t-1}+\sigma\varepsilon_t & \text{if} \quad [n/2]+1\leqslant t\leqslant n. \end{cases}$$

  Here, (ε_t) is a standard Gaussian white noise, (a_1,b_0,b_1) are real constants and $|a_1|<1$ and $|b_1|<1$. We get for $t\in\mathbb{Z}$ and $n\in\mathbb{N}^*$, $X_t\rightsquigarrow N\left(0,1/(1-a_1^2)\right)$ and $Cov(X_t,X_{t+n})=a_1^n/(1-a_1^2)$ before the change-point while $X_t\rightsquigarrow N\left(b_0/(1-b_1),\sigma^2/(1-b_1^2)\right)$ and $Cov(X_t,X_{t+n})=\sigma^2b_1^n/(1-b_1^2)$ after the change-point, so that this set of models allows us to investigate location-scale alterations of the null hypothesis by varying b_0 and σ as well as several degrees of within-sample dependence controlled by a_1 and b_1.
- In the first set of models, denoted by **Model 2**, we consider θ-dependent sequences. Namely, we simulate an order 1 Gaussian autoregressive (AR(1)) underlying processes X which allows us to investigate more subtle departures from the null than the classical location-scale shifts. Namely, X_t has a uniform distribution on $[-1,1]$ before the change-point while under the alternative, its density after the change-point satisfies:

$$f_2(x)=\frac{1}{2}+\frac{2x(\mu-|x|)}{\mu^2}1_{|x|<\mu}$$

  We vary μ in $\{0.33,0.67\}$. In order to get a θ-dependent structure, we generate a sequence of size n of the processes $X=2u-1$, $Y=F_Y^-1(V)$ where F_Y is the Cdf od Y and U and V are independent copies of the AR(1) process

$$U_t^*=\frac{1}{2}(U_{t-1}^*+\epsilon_t),\ \text{with}\ \ \epsilon_t\ \text{i.i.d}\ \rightsquigarrow\mathcal{B}(1/2)$$

  The process U^* has uniform marginals on $[0,1]$. In this case, the α-mixing condition fails to hold, but the process X is θ-dependent.

6.2. Tests

The computation of our test $\widetilde{BR}_n$ needs the calibration of the parameters described in Section 5. Firstly, the sequence d_n involved in the selection rule is set at $d_n = 10$. Secondly, according to the support $\mathbb{R}$ of the models considered in our simulation study, we have to choose an orthogonal family (Q_j) and an associated measure. For **Model 1**, we used for (Q_j) Hermite polynomials with associated Gaussian measure, so that

$$\widetilde{Q}_j(x) = Q_j(x)\exp(-x^2/2).$$

For **Model 2**, we used for (Q_j) Legendre polynomials with associated uniform measure.

Finally, in order to estimate σ^2, we use [8]'s estimator specifically adapted to AR(1) models. It is defined as follows.

$$\widehat{g}\,(0) = \frac{1}{2\pi}\sum_{|k|\leqslant q} w(\frac{k}{q+1})\widehat{\gamma}(k),\ q < n, \tag{19}$$

where $w(x) = 1 - x^2$ is the Parzen spectral window,

$$\widehat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-|k|}(Y_{t+k} - \bar{Y}_n)(Y_t - \bar{Y}_n), \quad \bar{Y}_n = \frac{1}{n}\sum_{t=1}^{n} Y_t$$

with $Y_t = \widetilde{Q}_1^*(X_t)$ and the truncation parameter q is given by

$$q = \left[\left(\frac{6\widehat{a}^2}{(1-\widehat{a}^2)^2}n\right)^{\frac{1}{5}}\right] \tag{20}$$

where $\widehat{a}$ is the least squares estimator of the autoregressive parameter in the model.

[34]'s competitor test is based on the test statistic :

$$\widetilde{LN}_n = \max_{1\leqslant k\leqslant m}\sup_{0\leqslant\theta\leqslant 1} d_n(\theta, x_k),$$

where $x_1, \ldots, x_m$ are distincts real numbers,

$$d_n(\theta, x) = \left(\frac{nh}{f_n(x)\|K\|^2}\right)^{1/2}\frac{[n\theta]}{n}\left(\frac{n-[n\theta]}{n}\right)\left((f_{[n\theta]}(x) - f^*_{n-[n\theta]}(x)\right),$$

$$f_{[n\theta]}(x) = \frac{1}{[n\theta]h}\sum_{j=1}^{[n\theta]} K\left(\frac{x-X_j}{h}\right),\ \ f^*_{n-[n\theta]}(x) = \frac{1}{(n-[n\theta])h}\sum_{j=[n\theta]+1}^{n} K\left(\frac{x-X_j}{h}\right),$$

K is an kernel function and h is a bandwidth satisfying $h = h_n \to 0$ and $nh \to \infty$ as n goes to infinity. Under suitable conditions, [34]'s theorem 2.3. states that

$$\widetilde{LN}_n \xrightarrow{\mathcal{L}} \widetilde{LN}_\infty = \max_{1\leqslant i\leqslant m}\sup_{\theta\in[0,1]}|B_i(\theta)|,$$

under the null, where the $B_i(\theta)$'s are independent Brownian Bridges, so that the distribution function of $\widetilde{LN}_\infty$ is $G(z) = F^m_{\widetilde{U}_\infty}(z)$, with $F_{\widetilde{U}_\infty}$ defined by (12).

For the practical computation of $\widetilde{LN}_n$, we use a gaussian Kernel. Bandwidth selection is done by the univariate plug-in selector (see [46]) and optimized for f_n.

6.3. Size of the test

The sizes in **Model 1** were obtained by setting $b_0 = 0$, $\sigma = 1$ and $a_1 = b_1$ so that

$$X_t = a_1 X_{t-1} + \varepsilon_t,$$

with $\varepsilon_t \rightsquigarrow N(0,1)$. We vary $a_1 \in \{0, 0.3, 0.5, 0.7\}$ in order to investigate several degrees of within-sample dependent. This degree increases with a_1 and $a_1 = 0$ corresponds to independence within sample.

In **Model 2**, X_t has a uniform distribution on $[-1, 1]$.

The empirical levels of the tests were defined as the percentage of rejection of the null hypothesis over 2000 replications of the test statistics and reported in Tables 6.4 and **??**.

Table 1. Empirical size of $\widetilde{BR}_n$ and $\widetilde{LN}_n$ (in %) in **Model 1**

Model α \ n		$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
$a_1 = -0.3$	$\widetilde{BR}_n$	$2.55 - 3.35$	$3.25 - 5.2$	$3.45 - 4.35$	$4.6 - 5.3$	
	$\widetilde{LN}_n$	0.05	0	0.05	0.15	
$a_1 = 0.0$	$\widetilde{BR}_n$	2.20	3.75	3.30	3.55	4.45
	$\widetilde{LN}_n$	0.05	0.05	0.05	0.35	0.30
$a_1 = 0.3$	$\widetilde{BR}_n$	2.45	4.1	4.85	5.15	4.95
	$\widetilde{LN}_n$	0.15	0.4	0.45	0.95	0.95
$a_1 = 0.5$	$\widetilde{BR}_n$	2.35	4.85	4.90	$5.4 - 7.33$	
	$\widetilde{LN}_n$	0.5	0.75	1.5	1.73	
$\mathbf{a_1} = 0.6$	$\widetilde{BR}_n$	3.20	5.20	7.25	6.55	4.95
	$\widetilde{LN}_n$	1.25	2.05	3.05	4.05	5.1
$a_1 = 0.7$	$\widetilde{BR}_n$	$3.9 - 4.55$	$6.8 - 5.95$	$7 - 8.35$	$6.93 - 10.26$	
	$\widetilde{LN}_n$	2.70	3.9	6.3	8.6	
$a_1 = 0.8$	$\widetilde{BR}_n$	4.50	9.15	12.45		
	$\widetilde{LN}_n$	6.65	13.30	18.3		

From Tables 1 and 2 we observe that our statistic $\widetilde{BR}_n$ has a good size i.e. the empirical size converges quickly to a nominal size α.
However the statistic $\widetilde{LN}_n$ of Le and Na has a size distortion, it underrejects the null hypothesis.

6.4. Power of the test

Table 2. Empirical size of $\widetilde{BR}_n$ and $\widetilde{LN}_n$ (in %) in **Model 1**

Model α \ n		$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
$b_0 = 1, a_1 = b_1 = 0, \sigma = 1$	$\widetilde{BR}_n$	$36.05(43.6ui)$	79.1	99.1	100	100
	$\widetilde{LN}_n$	3.85	32.3	84.65	100	100
$b_0 = a_1 = b_1 = 0, \sigma = 0.5$	$\widetilde{BR}_n$	$1.8 - 3.2(3.35)$	$2.9 - 3(3.85)$	$2.9 - 3.25(3.75)$	$4 - 3.8(3.4)$	
	$\widetilde{LN}_n$	0.4	8.2	46.35	97	
$b_0 = 0, a_1 = 0.3, b_1 = 0.4, \sigma = 1$	$\widetilde{BR}_n$	23.1	57.15	88.3	99.85	100
	$\widetilde{LN}_n$	6.85	32.05	70.45	99.30	100

Tables 3 and 4 show that $\widetilde{BR}_n$ outperforms the statistic $\widetilde{LN}_n$ for a moderate sample size (n=100, n=200). But for large size (n=500, n=1000) the two statistics $\widetilde{BR}_n$ and $\widetilde{LN}_n$ have the same empirical power. We observe also that as the sample size inreases the power of the two statistics becomes good.

7. Proofs

7.1. Proof of Theorem 1

In the following, we denote by $\mathbb{P}_0$ the probability under H_0 and we recall that. $U_n(k,\theta) = \sum_{j=1}^{k} (V_n^{(j)}(\theta))^2$.

To prove (10), let us show that $\mathbb{P}_0(K_{\theta,n} \geqslant 2)$ tends to zero as n tends to infinity.

By definition of $K_{\theta,n}$ and Markov's inequality, we have

$$\begin{aligned}
\mathbb{P}_0(K_{\theta,n} \geqslant 2) &= \mathbb{P}_0\left(\max_{2\leqslant k\leqslant d(n)}\{U_n(k,\theta) - k\log n\} \geqslant U_n(1,\theta) - \log n\right) \\
&= \mathbb{P}_0\left(\exists s, 2\leqslant k\leqslant d(n), U_n(k,\theta) - k\log n \geqslant U_n(1,\theta) - \log n\right) \\
&= \mathbb{P}_0\left(\exists s, 2\leqslant k\leqslant d(n), \sum_{j=2}^{k}(V_n^{(j)}(\theta))^2 \geqslant (k-1)\log n\right) \\
&\leqslant \mathbb{P}_0\left(\exists k, 2\leqslant k\leqslant d(n), (V_n^{(k)}(\theta))^2 \geqslant \log n\right) \\
&\leqslant \sum_{k=2}^{d(n)}\mathbb{P}_0\left((V_n^{(k)}(\theta))^2 \geqslant \log n\right) \\
&\leqslant \frac{1}{\log n}\sum_{k=2}^{d(n)}\mathbb{E}_0\left(|V_n^{(k)}(\theta)|^2\right),
\end{aligned}$$

Using similar calculations as in (9) yields

$$V_n^{(k)}(\theta) = W_n^{(k)}(\theta) - \frac{[n\theta]}{n}W_n^{(k)}(1),$$

with

$$W_n^{(k)}(\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[n\theta]}\widetilde{Q}_k^*(X_t),\ \theta\in[0,1]$$

so that

$$\begin{aligned}
\mathbb{E}_0\left(|V_n^{(k)}(\theta)|^2\right) &= \frac{1}{n}\sum_{l=1}^{[n\theta]}\sum_{t=1}^{[n\theta]}\mathbb{E}_0\big(\widetilde{Q}_k^*(X_t)\widetilde{Q}_k^*(X_l)\big) + \frac{[n\theta]^2}{n^3}\sum_{l=1}^{n}\sum_{t=1}^{n}\mathbb{E}_0\big(\widetilde{Q}_k^*(X_t)\widetilde{Q}_k^*(X_l)\big) \\
&- \frac{2[n\theta]}{n^2}\sum_{l=1}^{[n\theta]}\sum_{t=1}^{n}\mathbb{E}_0\big(\widetilde{Q}_k^*(X_t)\widetilde{Q}_k^*(X_l)\big)
\end{aligned}$$

Notice that since $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary, the same happens for $(\widetilde{Q}_k^*(X_t))_{t\in\mathbb{Z}}$. Therefore, classical calculations using the stationarity of the process yield

$$\sum_{l=1}^{n}\sum_{t=1}^{n}\mathbb{E}_0\big(\widetilde{Q}_k^*(X_t)\widetilde{Q}_k^*(X_l)\big) \leqslant 2n\sum_{t=1}^{n}\big(1-\frac{t}{n}\big)\mathbb{E}_0\left(\widetilde{Q}_k^*(X_1)\widetilde{Q}_k^*(X_t)\right)$$

Therefore,

$$\mathbb{E}_0\left(|V_n^{(k)}(\theta)|^2\right) \leqslant 6\sum_{t=1}^{n}\left|\mathbb{E}_0\big(\widetilde{Q}_k^*(X_1)\widetilde{Q}_k^*(X_t)\big)\right| = 6r_n(k)$$

so that

$$\mathbb{P}_0(K_{\theta,n}\geqslant 2) \leqslant 6C\,o(1)$$

by **(A)** and convergence (10) in Theorem 1 holds.

Let us now show (11). By (10), one has

$$\sqrt{\widetilde{U}_n(\theta)} = \left|V_n^{(1)}(\theta)\right| + o_{\mathbb{P}_0}(1)$$

with $V_n^{(1)}(\theta)$ given by (9). Thus, when **(B)** holds, one has

$$W_n(\theta) \overset{\mathcal{D}}{\Longrightarrow} \sigma W(\theta),\quad \frac{[n\theta]}{n}W_n(1) \overset{\mathcal{D}}{\Longrightarrow} \theta\sigma W(1)$$

where W is a standard Brownian motion so that

$$V_n^{(1)}(\theta) \overset{\mathcal{D}}{\Longrightarrow} \sigma B(\theta)$$

and (11) in Theorem 1 obtains by the continuous mapping theorem.

7.1.1. Proof of Theorem 2

Under H_1^*, following the proof of Theorem 1, then using standard power inequalities, we have

$$\begin{aligned}\mathbb{P}_1(K_{\theta,n} \geqslant 2) &\leqslant \frac{1}{\log n}\sum_{k=2}^{d(n)} \mathbb{E}_1\left(|V_n^{(k)}(\theta)|^2\right) \\ &\leqslant \frac{2d(n)}{\log(n)}\big(C^* + n|\delta_k(\theta)|^2\big).\end{aligned}$$

Therefore, $\mathbb{P}_1(K_{\theta,n} \geqslant 2) = o_{\mathbb{P}_1}(1)$ by Assumption **(A*)**.

Under H_1^{**}, one has with similar arguments $\mathbb{P}_1(K_{\theta,n} > K) = o_{\mathbb{P}_1}(1)$.
On the other hand,

$$\begin{aligned}\mathbb{P}_1(K_{\theta,n} < K) &= \mathbb{P}_1\left(\exists k < K, U_n(k,\theta) - k\log n \geqslant U_n(K,\theta) - K\log n\right) \\ &= \mathbb{P}_1\left(\exists k < K, \sum_{j=k}^{K}(V_n^{(j)}(\theta))^2 \leqslant (K-k)\log(n)\right) \\ &\leqslant \sum_{k=1}^{K-1}\mathbb{P}_1\left(\sum_{j=k}^{K}(V_n^{(j)}(\theta))^2 \leqslant (K-k)\log(n)\right) \\ &\leqslant K\mathbb{P}_1\left(\frac{|V_n^{(K)}(\theta))|}{\sqrt{\log(n)}} \leqslant \sqrt{K}\right).\end{aligned}$$

Moreover, one has

$$\frac{V_n^{(K)}(\theta)}{\sqrt{\log(n)}} = \frac{1}{\sqrt{\log(n)}}\left(V_n^{(K)}(\theta) - \frac{[n\theta](n-[n\theta])}{n^{3/2}}\delta_K(\theta)\right) + \frac{[n\theta](n-[n\theta])}{n^{3/2}}\frac{\delta_K(\theta)}{\sqrt{\log(n)}},$$

and

$$\frac{[n\theta](n-[n\theta])}{n^{3/2}} = O(\sqrt{n})$$

so that $\dfrac{V_n^{(K)}(\theta)}{\sqrt{\log(n)}} = O_{\mathbb{P}_1}(n^{1-\beta}) + O_{\mathbb{P}_1}(1)$ tends to infinity as n tends to infinity.

It follows that $K_{n,\theta} \to K$ and that the test statistic $\widetilde{U}_n$ converges to infinity as soon as **(B*)** holds.

References

[1] W. Albers, W.C.M. Kallenberg, and F. Martini. Data driven rank tests for classes of tail alternatives. *Journal of the American Statistical Association*, 96:685–696, 2001.

[2] E. Andreou, E. & Ghysels. *Handbook of Financial Time Series. Structural Breaks in Financial Time Series*. T. G. Andersen (ed) et al., Springer, Berlin, 2009.

[3] D. W. K. Andrews. Non strong mixing autoregressive processes. *Journal of Applied Probability*, 21:930–934, 1984.

[4] D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–856, 1991.

[5] M. Bagshaw and R.A. Johnson. The effect of serial correlation on the performance of cusum tests ii. *Technometrics*, pages 73–80, 1975.

[6] J. Beran. *Statistics for long-memory processes*. Chapman and Hall, New-York, 1994.

[7] P. Billingsley. *Convergence of probability measures*. Springer, 1968.

[8] M. Boutahar. Comparison of non-parametric and semi-parametric tests in detecting long memory. *Journal of Applied Statistics*, 36(9):945–972, 2009.

[9] B. E. Brodsky and B. S. Darkhovsky. Nonparametric methods in change-point problems. *Mathematics and Its Applications*, 243, 1993.

[10] S. Cohen and R. Marty. Invariance principle, multifractional gaussian processes and long-range dependence. *Ann. Inst. H. Poincaré Probab. Statist.*, 44(3):475–489, 2008.

[11] M. Csorgo and L. Horvath. *Limit Theorems in Change Point Analysis*. John Wiley & Sons, New York, 1997.
[12] J. Dedecker and P. Doukhan. A new covariance inequality and applications. *Stochastic Processes and their Applications*, 106:63–80, 2003.
[13] J. Dedecker, P. Doukhan, G. Lang, J.R. Leon, S. Louhichi, and C. Prieur. Weak dependence: With examples and applications. In *Lecture Notes in Statistics*, volume 190. Springer, 2007.
[14] J. Dedecker, F. Merlevède, and M. Peligrad. Invariance principles for linear processes with application to isotonic regression. *Bernoulli*, 17(1):88–113, 2011.
[15] P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84:313–342, 1999.
[16] P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. *Annales de l'I.H.P., section B*, 30(1):63–82, 1994.
[17] P. Doukhan, P. Massart, and E. Rio. Invariance principles for absolutely regular empirical processes. *Annales de l'I.H.P., section B*, 31(2):393–427, 1995.
[18] P. Doukhan, D. Pommeret, and L. Reboul. Data driven smooth test of comparison for dependent sequences. *Journal Of Multivariate Analysis*, 139:147–165, 2015.
[19] P. Doukhan and O. Wintenberger. An invariance principle for weakly dependent stationary general models. *Probab. Math. Statist*, 1:45–73, 2007.
[20] R. L. Eubank and V. N. LaRiccia. Asymptotic comparison of Cramér-Von-Mises and nonparametric function estimation techniques for testing goodness-of-fit. *The Annals of Statistics*, 20:2071–2086, 1992.
[21] J. Fan. Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association*, 91:674–688, 1996.
[22] B. Ghattas, D. Pommeret, L. Reboul, and A.F. Yao. Data driven smooth test for paired populations. *Journal of Statistical Planning and Inference*, 141(1):262–275, 2011.
[23] L. Giraitis, R. Leipus, and D. Surgailis. The change-point problem for dependent observations. *J. Stat. Plan. Inf.*, 53:297–310, 1996.
[24] L. Horvath, P. Kokoszka, and J. Steinebach. Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis*, 68(1):96–119, 1999.
[25] M. Huˇ sková and J. Antoch. Detection of structural changes in regression. *Tatra Mt. Math. Publ.*, 26:201–215, 2003.
[26] R. Ignaccolo. Goodness-of-fit tests for dependent data. *Nonparametric Statistics*, 16:19–38, 2004.
[27] T. Inglot, W.C.M. Kallenberg, and T. Ledwina. Data-driven smooth tests for composite hypotheses. *The Annals of Statistics*, 25:1222–1250, 1997.
[28] T. Inglot and T. Ledwina. Asymptotic optimality of data driven neyman's tests. *The Annals of Statistics*, 24:1982–2019, 1996.
[29] J.A. Janic-Wróblewska and T. Ledwina. Data driven rank test for two-sample problem. *The Scandinavian Journal of Statistics*, 27:281–297, 2000.
[30] W.C.M. Kallenberg and T. Ledwina. Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *The Annals of Statistics*, 23:1594–1608, 1995a.
[31] W.C.M. Kallenberg and T. Ledwina. On data-driven Neyman's tests. *Probability and Mathematical Statistics*, 15:409–426, 1995b.
[32] P. Kokoszka and R. Leipus. *Detection and Estimation of Changes in Regime*, pages 325–337. SBoston: Birkhauser, 2002.
[33] L. Lajos Horváth and G. Rice. Extensions of some classical methods in change point analysis. In *Test*, 2014.
[34] S. Lee and S. Na. A nonparametric test for the change in the density function in strong mixing processes. *Statist. Prob. Letters*, 66:1–25, 2004.
[35] Degui Li and Zhengyan Lin. A nonparametric test for the change of the density function under association. *Journal of Nonparametric Statistics*, 19:1–12, 2007.
[36] F. Merlevède, M. Peligrad, and U. Sergey. Recent advances in invariance principles for stationary sequences. *Probab. Surveys*, 3:1–36, 2006.
[37] A. Munk, J.P. Stockis, J. Valeinis, and G. Giese. Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data. *Annals of the Institute of Statistical Mathematics*, 63:939–959, 2009.

[38] W. K. Newey and K. D. West. Automatic lag selection in covariance matrix estimation. *Review of Economic studies*, 61:631–653, 1994.
[39] J. Neyman. Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift*, 20:149–199, 1937.
[40] E.S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42:523–526, 1955.
[41] M. Peligrad. Invariance principles for mixing sequences of random variables. *The annals of Probability*, 10(4):968–981, 1982.
[42] E. Rio. Théorie asymptotique des processus aléatoires faiblement dépendants. In *Mathématiques et Applications*, volume 31, Berlin, 2000. Springer.
[43] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Set. USA*, 42:43–47, 1956.
[44] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6.2:461–464, 1978.
[45] L. Selk and N. Neumeyer. Testing for a change of the innovation distribution in nonparametric autoregression : the sequential empirical process approach. *Scandinavian Journal of Statistics*, 2012.
[46] M.P. Wand and M.C. Jones. *Kernel Smoothing*. CRC Press, 1994.
[47] C.S. Withers. Conditions for linear processes to be strong mixing. *Probability Theory and Related Fields*, 57:477–480, 1981.