Statistical Papers (2022) 63:287–316
https://doi.org/10.1007/s00362-021-01242-3

REGULAR ARTICLE

On change-points tests based on two-samples U-Statistics for weakly dependent observations

Joseph Ngatchou-Wandji[1] · **Echarif Elharfaoui**[2] · **Michel Harel**[3,4]

Received: 12 August 2020 / Revised: 17 April 2021 / Accepted: 18 May 2021 / Published online: 28 May 2021

Abstract
We study change-points tests based on U-statistics for absolutely regular observations. Our method avoids some technical assumptions on the data and the kernel. The asymptotic properties of the U-statistics are studied under the null hypothesis, under fixed alternatives and under a sequence of local alternatives. The asymptotic distributions of the test statistics under the null hypothesis and under the local alternatives are given explicitly and the tests are shown to be consistent. A small set of simulations is done for evaluating the performance of the tests in detecting changes in the mean, variance and autocorrelation of some simple time series.

Mathematics Subject Classification 60F17 · 62F03 · 62M10

1 Introduction

We are interested in detecting possible differences between the distributions of real-valued random variables $X_1, X_2, \ldots, X_n$. In practice, this issue can be of primary importance for data from industrial quality control, financial markets, medical diagnostics, hydrology, climatology etc.

This statistical matter is known as change-points problem whose theory, well developed for independent data, has considerably been studied in the literature both from parametric and non-parametric point of view. Since Page (1954), so many tests have been proposed for testing changes in the distribution of iid data. Among others, Chernoff and Zacks (1964) propose test statistics for detecting shifts in the mean of a normal

✉ Joseph Ngatchou-Wandji
joseph.ngatchou-wandji@univ-lorraine.fr

1 EHESP Sorbonne Paris Cité & Institut Élie Cartan de Lorraine, 54506 Vandoeuvre-lès-Nancy Cedex, France

2 Faculté des Sciences, Université Chouaîb Doukkali, BP. 20, El Jadida 24000, Marocco

3 INSPÉ Limoges, 87036 Limoges Cedex, France

4 IMT, UMR 5219 UPS, 31062 Cedex 9 Toulouse, France

distribution function. Their results are generalized to exponential family by Kander and Zacks (1966), Gardner (1969) and MacNeill (1974). Matthews et al. (1985) study maximal score statistics to test for constant hazard against a change-point alternative. Haccou et al. (1988) propose a likelihood ratio test for a change-point in a sequence of independent exponentially distributed random variables. They prove their test is optimal in the sense of Bahadur. Yao and Davis (1986) consider the asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent Gaussian random variables. Csörgő and Horváth (1987) propose statistics based on processes of linear rank statistics with quantile scores. A review on non-parametric procedures is given by Wolfe and Schechtman (1984), who summarize among others, pioneer works on change-points as Page (1954, 1955), Bhattacharya and Zhou (2017), Sen and Srivastava (1975) and Pettitt (1979).

These last years there is a growing interest in change-points study in time series data. Most of the techniques and approaches used are mainly based either on testing for the existence of changes or for their locations, or on estimating the locations. Some relevant references on change-points estimation are Härdle and Tsybakov (1997), Härdle et al. (1998), Bardet and Wintenberger (2009), Döring (2010, 2011), Ciuperca (2011), Bardet and Kengne (2014), Amano (2012), Horváth and Hušková (2005), Yang and Song (2014), Mohr and Selk (2020) and Yang et al. (2020). A non-exhaustive list of references on testing approaches are, among others, Kengne (2012), Chen et al. (2011), Dehling et al. (2013), Dehling et al. (2015), Wang and Phillips (2012), Francq and Zakoïan (2012), Bardet et al. (2012), Zhou (2014), Fotopoulos et al. (2009), Huh (2010), Enikeeva et al. (2018), Gombay (2008), Gombay and Serban (2009), Hlávka et al. (2020) and Ma et al. (2020) which uses both estimation and testing. Meintanis (2016) gives an interesting survey of testing procedures based on the empirical characteristic function. We would also like to mention the interesting and related work of Rackauskas and Wendler (2020) who deal with a robust test based on the Wilcoxon statistic for detecting epidemic changes. While the asymptotic behavior of the test statistic is studied under the null hypothesis, with techniques close to ours, its consistency is only discussed.

In almost all the existing testing papers, except perhaps Fotopoulos et al. (2009), Khakhubia (1987), Bhattacharyya and Johnson (1968), Dehling et al. (2013), Dehling et al. (2017a) and Dehling et al. (2017b), the local power is rarely studied. This issue is considered in this paper where the tests studied are derived from basic processes of the general form

$$Z_n^*(\lambda) = n^{-3/2} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} h(X_i, X_j),\ 0 \leq \lambda \leq 1,$$

with $h : \mathbb{R}^2 \to \mathbb{R}$ a kernel function.

The asymptotic distribution of a related process has been studied in a Hölder space by Rackauskas and Wendler (2020) for stationary mixing data and antisymmetric h, while this process has been studied in a Skorohod space for instance by Csörgő and Horváth (1988) for iid data, and by Dehling et al. (2015) for data assumed to be functions of mixing random variables satisfying some other conditions such as *1-*

approximating functional, 1-sided functional, bounded density function, with kernels h being *1-continuous*. These conditions are avoided here by our mixing assumption on the data, and our study is done in a Skorohod space. Furthermore, besides the Kolmogorov-Smirnov (KS) type test usually studied in the literature, we study a Cramer-von Mises (CM) version which has the advantage that its theoretical limiting distribution under the null hypothesis can be approximated for any kernel h. We restrict our study to the classical case of one change-point detection. But our results can be generalized to multi-change-points detection which we postpone to a future paper.

The paper is organized as follows. In Sect. 2, we define useful quantities such as the test statistics, and we list some assumptions. In Sect. 3 we study the asymptotic properties of our tests statistics under the null hypothesis, under a sequence of local alternatives and under fixed alternatives. Practical considerations are presented and discussed in Sect. 4, while the last section contains the proofs of the main results.

2 General definitions and assumptions

For cumulative distribution functions Q and R, denote by $\theta(Q, R)$ the following real number

$$\theta(Q, R) = \int\int h(x, y)dQ(x)dR(y).$$

For any $i = 1, 2, \ldots, n$, let F_i be the cumulative distribution function of X_i. We aim to check possible differences between the F_i's. We restrict ourselves to checking if there exists only one index i_0 for which F_{i_0} and F_{i_0+1} are different. We study this problem by testing the hypothesis $\mathcal{H}_0$ against the alternative $\mathcal{H}_1$, defined respectively by

$$\begin{aligned}
&\mathcal{H}_0 : F_1(x) = F_2(x) = \ldots = F_n(x),\ x \in \mathbb{R}\\
&\mathcal{H}_1 : \exists\, \lambda_0 \in (0, 1) \text{ such that } F_1(x) = F_2(x) = \ldots = F_{[n\lambda_0]}(x) = F(x), x \in \mathbb{R} \text{ and}\\
&\quad F_{[n\lambda_0]+1}(x) = \ldots = F_n(x) = G(x), x \in \mathbb{R},\ \text{and } \theta(F, F) \neq \theta(F, G).
\end{aligned}$$

Figure 1 exhibits the chronograms of some time series each of size 200, owning a change-point at $t = 100$. The first graphic in the first row shows a change in the mean of a shifted Gaussian white noise, while the one at its right shows a change in its variance. The first graphic in the second row shows a change in both the mean and the variance of a shifted Gaussian white noise, and the second shows a change in the autocorrelation of a Gaussian AR(1) model.
In order to evaluate the capacity of the tests to detect weak changes, we also consider the local alternatives $\mathcal{H}_{1,n}$ of the form

$$\begin{aligned}
&\mathcal{H}_{1,n} : \exists\, \lambda_0 \in (0, 1) \text{ such that } F_1(x) = F_2(x) = \ldots = F_{[n\lambda_0]}(x) = F(x),\ \text{and}\\
&\quad F_{[n\lambda_0]+1}(x) = \ldots = F_n(x) = G^{(n)}(x), x \in \mathbb{R},\ \text{and}\\
&\quad \theta(F, G^{(n)}) = \theta(F, F) + n^{-1/2}[A + o(1)],\ \text{for some } \gamma \in \mathbb{R}^*.
\end{aligned}$$

Remark 1 Particular examples of local alternatives $\mathcal{H}_{1,n}$ are those for which there exists a constant γ such that $G^{(n)}(x) = F(x + n^{-1/2}\gamma)$ and the kernel function h is twice differentiable with finite integral $\int\int(\partial h(x, y)/\partial y)dF(x)dF(y)$ and bounded second-order derivatives $\partial^2 h(x, y)/\partial^2 y$. This can be checked easily by a suitable application of the Taylor-Young formula. With this example, one finds

$$A = -\gamma \int\int \frac{\partial h}{\partial u}(x, u)dF(x)dF(u).$$

In the purpose of solving our testing problem, the tests we are going to use are based on the following KS and CM statistics

$$T_{1,n} = \max_{1\le k\le n-1} \left| n^{-3/2} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left\{h(X_i, X_j) - \theta(\widehat{F}, \widehat{F})\right\} \right| \tag{1}$$

$$T_{2,n} = \frac{1}{n} \sum_{1\le k\le n-1} \left\{ n^{-3/2} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left[h(X_i, X_j) - \theta(\widehat{F}, \widehat{F})\right] \right\}^2, \tag{2}$$

where $\widehat{F}$ stands for any consistent estimator of F, a simple example being the empirical cumulative distribution function.
Denote by $[x]$ the integer part of any real number x. Noting that for any $k \in \{1, \ldots, n-1\}$, there exists $\lambda_* \in [0, 1]$ such that $k = [\lambda_* n]$, one can write, asymptotically,

$$T_{1,n} = \sup_{\lambda\in[0,1]} |Z_n(\lambda)|$$
$$T_{2,n} = \int_0^1 Z_n^2(\lambda)d\lambda,$$

where Z_n stands for the following stochastic process

$$Z_n(\lambda) = n^{-3/2} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \left[h(X_i, X_j) - \theta(F, F)\right],\ 0 \le \lambda \le 1. \tag{3}$$

Define the following U-statistic U_n with kernel h, and the following functions

$$U_n = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} h(X_i, X_j)$$
$$h_{F,1}(x) = \int h(x, y)dF(y) - \theta(F, F)$$
$$h_{F,2}(y) = \int h(x, y)dF(x) - \theta(F, F)$$

$$h_{G,1}(x) = \int h(x, y) dG(y) - \theta(F, G)$$
$$h_{G,2}(y) = \int h(x, y) dF(x) - \theta(F, G)$$
$$g_F(x, y) = h(x, y) - h_{F,1}(x) - h_{F,2}(y) + \theta(F, F)$$
$$g_G(x, y) = h(x, y) - h_{G,1}(x) - h_{G,2}(y) + \theta(F, G).$$

Consider the Hoeffding decomposition of U_n under $\mathcal{H}_0$

$$U_n = \theta(F, F) + U_{n,1}^{(1)} + U_{n,2}^{(1)} + U_n^{(2)}, \tag{4}$$

where

$$U_{n,1}^{(1)} = n^{-1} \sum_{i=1}^{n} h_{F,1}(X_i)$$
$$U_{n,2}^{(1)} = n^{-1} \sum_{i=1}^{n} h_{F,2}(X_i)$$
$$U_n^{(2)} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \left[h(X_i, X_j) - h_{F,1}(X_i) - h_{F,2}(X_j) \right] - \theta(F, F).$$

Also, define the following numbers

$$\sigma_{kl} = \mathbb{E}\left[h_{F,k}(X_1) h_{F,l}(X_1) \right] + 2 \sum_{j=1}^{\infty} \mathbb{C}\text{ov}\left(h_{F,k}(X_1), h_{F,l}(X_{1+j}) \right),\ k, l = 1, 2.$$

We make the following assumptions :

(A1) The sequence $\{X_i\}_{i \in \mathbb{N}}$ is absolutely regular with the rate

$$\beta(k) = \mathcal{O}(\tau^k), \quad 0 < \tau < 1. \tag{5}$$

(A2) $\{X_i\}_{i \in \mathbb{N}}$ is stationary.

(A3) We consider $(Y_i)_{1 \le i \le n}$ a sequence of stationary and absolute regular random variables with rate (5). We assume the cumulative distribution function of the Y_i's is G.

(A4) We consider $(Y_{ni})_{1 \le i \le n, n \ge 1}$ a sequence of stationary and absolute regular random variables with cumulative distribution function $G^{(n)}(x) = F(x + \eta n^{-\frac{1}{2}})$. We assume the cumulative distribution functions $G_{ij}^{(n)}$ and $G_{ij}^{*(n)}$ of the (Y_{ni}, Y_{nj})'s and (X_i, Y_{nj})'s respectively satisfy

$$\lim_{n \to \infty} G_{ij}^{(n)}(x, y) = F_{ij}(x, y) \text{ and } \lim_{n \to \infty} G_{ij}^{*(n)}(x, y) = F_{ij}(x, y),\ 1 \le i < j \le n,$$

where F_{ij} is the cumulative distribution function of (X_i, X_j).

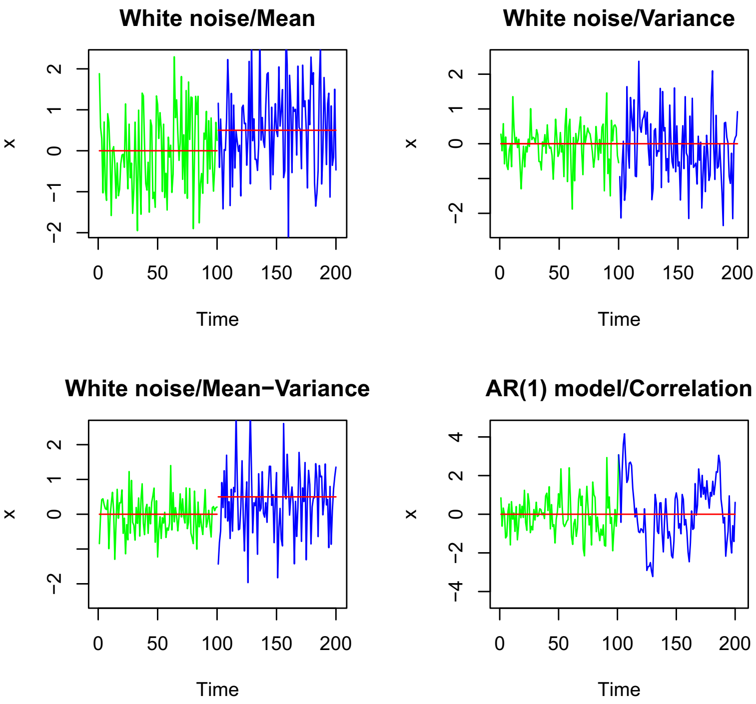

Fig. 1 First row : change in the mean and change in the variance of a shifted white noise. Second row : change in both the mean and the variance of a shifted white noise, and change in the autocorrelation of an AR(1) model. The red lines represent the mean. The green, is the chronogram before the change occurred, and the blue is that after the change occurred

- We recall from Harel and Puri (1994) that a non-necessarily stationary triangular sequence $\{\mathcal{V}_{ni}, 1 \leq i \leq n, n \geq 1\}$ is absolutely regular if $\beta(k) \longrightarrow 0$, as $k \to \infty$, where

$$\beta(k) = \sup_{n\in\mathbb{N}} \sup_{k\leq n} \max_{1\leq j\leq n-k} \mathbb{E}\left(\sup_{A\in\mathcal{A}^{\infty}_{n,j+k}} \left|\mathbb{P}(A \mid \mathcal{A}^{j}_{n,0}) - \mathbb{P}(A)\right| \right),$$

with $\mathcal{A}^{j}_{n,i}$ standing for the σ-algebra generated by $\mathcal{V}_{ni}, \ldots, \mathcal{V}_{nj},\ i, j \in \mathbb{N} \cup \{\infty\}$. It will be said to be strong mixing or α-mixing if $\alpha(k) \longrightarrow 0$ as $k \to \infty$, where

$$\alpha(k) = \sup_{n\in\mathbb{N}} \max_{1\leq j\leq n-k} \sup_{A\in\mathcal{A}^{\infty}_{n,j+k}, B\in\mathcal{A}^{j}_{n,0}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| .$$

- Note that a β-mixing sequence of random variables is also α-mixing.

Remark 2 We assume a geometrical mixing rate by convenience. We believe our results can be established as well for arithmetical mixing rates to be found.

3 Asymptotics

In this section, we state all our theoretical results. Most of the proofs are postponed to the last section.

Theorem 1 *Assume that the assumptions* (A1)–(A3) *hold. Then, under* $\mathcal{H}_0$*, if* $\max\{\sup_{i,j} \mathbb{E}\left[|h(X_i, X_j)|^{2+\delta}\right], \int\int_{\mathbb{R}^2} |h(x,y)|^{2+\delta} dF(x)dF(y)\} < \infty$ *for some* $\delta > 0$ *and the absolute regularity condition* (5) *is satisfied, then for any* $k, l = 1, 2$*,* $\sigma_{kl} < \infty$.

If in addition $\sigma_{kl} > 0$*,* $1 \le k, l \le 2$*, then the sequence of processes of* $\{Z_n(\lambda); 0 \le \lambda \le 1\}_{n\in\mathbb{N}}$ *converges in distribution towards a zero-mean Gaussian process with representation*

$$Z(\lambda) = (1-\lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)),\ 0 \le \lambda \le 1,$$

where $\{W_1(\lambda), W_2(\lambda)\}_{0\le\lambda\le1}$ *is a two-dimensional zero-mean Brownian motion with covariance kernel matrix with entries* $\mathbb{C}ov(W_k(s), W_l(t)) = \min(s,t)\sigma_{kl},\ k, l = 1, 2.$

Proof See Appendix.

Remark 3 The covariance kernel of the Gaussian process Z defined in Theorem 1 is given for all $s, t \in [0, 1]$ by

$$\begin{aligned}\Delta(s,t) = \mathbb{C}\mathrm{ov}(Z(s), Z(t)) \\ &= \sigma_{11}(1-s)(1-t)\min(s,t) + \sigma_{22}st[1-s-t+\min(s,t)] \\ &+\sigma_{12}\{t(1-s)[s-\min(s,t)] + s(1-t)[t-\min(s,t)]\}.\end{aligned} \tag{6}$$

Theorem 2 *Assume* (A1)*,* (A2) *and* (A4) *hold, h is twice differentiable with bounded second-order derivatives* $\partial^2 h(x,y)/\partial x\partial y$*, and the integral* $\int\int(\partial h(x,y)/\partial y)dF(x)dF(y)$ *is finite. Then, under* $\mathcal{H}_{1,n}$*, if*

$$\begin{aligned}&\sup_{1\le i,j\le n} \mathbb{E}\left[|h(X_i, X_j)|^{2+\delta}\right],\ \sup_{n\ge1}\sup_{i,j} \mathbb{E}\left[|h(Y_{ni}, Y_{nj})|^{2+\delta}\right], \\ &\sup_{n\ge1}\sup_{1\le i,j\le n} \mathbb{E}\left[|h(X_i, Y_{nj})|^{2+\delta}\right],\ \int\int_{\mathbb{R}^2} |h(x,y)|^{2+\delta} dF(x)dF(y), \\ &\sup_{n\ge1}\int\int_{\mathbb{R}^2} |h(x,y)|^{2+\delta} dG^{(n)}(x)dG^{(n)}(y),\ \sup_{n\ge1}\int\int_{\mathbb{R}^2} |h(x,y)|^{2+\delta} dF(x)dG^{(n)}(y)\end{aligned}$$

are finite for some $\delta > 0$*, if for any* $k, l = 1, 2$*,* $\sigma_{kl} > 0$*, then the sequence of processes* $\{Z_n(\lambda); 0 \le \lambda \le 1\}_{n\in\mathbb{N}}$ *converges in distribution towards a Gaussian process* $\widetilde{Z}$ *with mean* $(1-\lambda)\lambda A$ *and representation*

$$\widetilde{Z}(\lambda) = (1-\lambda)\lambda A + Z(\lambda),\quad 0 \le \lambda \le 1,$$

where $\{Z(\lambda)\}_{0\le\lambda\le1}$ *is the zero-mean Gaussian process defined in Theorem* 1.

Proof See Appendix. □

Theorem 3 *Assume* (A1)–(A3) *hold and that under* $\mathcal{H}_1$*, the integrability conditions in Theorem* 2 *hold. Then,*

$$n^{-1/2}Z_n^*(t) \xrightarrow{a.s.}_{n\to\infty} \begin{cases} \theta(F,F)t(\lambda_0 - t) + \theta(F,G)t(1-\lambda_0), & 0 \le t \le \lambda_0 \\ \theta(G,G)(t-\lambda_0)(1-t) + \theta(F,G)\lambda_0(1-t), & \lambda_0 \le t < 1. \end{cases} \quad (7)$$

Proof See Appendix.

Theorem 4 *Assume that the assumptions of Theorem* 2 *hold. Let* $(Z(\lambda) : 0 \le \lambda \le 1)$ *be the limiting process defined in Theorems* 1 *and* 2 *, and* Δ *its covariance kernel. Then*

(i) *Under* $\mathcal{H}_0$*, as n tends to infinity, one has the following convergence in distribution,*

$$T_{1,n} \longrightarrow \sup_{\lambda\in[0,1]} |Z(\lambda)|$$
$$T_{2,n} \longrightarrow \sum_{j\ge 1} \zeta_j \chi_j^2,$$

where the χ_j^2*'s are iid chi-square random variables with one degree of freedom and the* ζ_j*'s are standing for the eigenvalues of the linear integral operator* ∇ *defined for any square integrable function* τ *on* $[0,1]$ *by*

$$\nabla[\tau(\cdot)] = \int_0^1 \Delta(\cdot, s)\tau(s)ds. \quad (8)$$

(ii) *Under* $\mathcal{H}_{1,n}$*, as n tends to infinity, one has the following convergence in distribution,*

$$T_{1,n} \longrightarrow \sup_{\lambda\in[0,1]} |(1-\lambda)\lambda A + Z(\lambda)|$$
$$T_{2,n} \longrightarrow \sum_{j\ge 1} \zeta_j \chi_j^{*2},$$

where the χ_j^{*2}*'s are iid non-central chi-square random variables with one degree of freedom and non-centrality parameters* $\rho_j^2\zeta_j^{-1}$ *with the* e_j*'s standing for the eigenvectors of the integral operator* ∇*, associated with the eigen-value* ζ_j*, and*

$$\rho_j = A\int_0^1 \lambda(1-\lambda)e_j(\lambda)d\lambda.$$

(iii) *Under* $\mathcal{H}_1$*, as n tends to infinity, one has the following convergence in probability,*

$$T_{1,n} \longrightarrow \infty, \quad T_{2,n} \longrightarrow \infty.$$

Proof (i) From Theorem 1 and the continuous mapping theorem, $T_{1,n}$ and $T_{2,n}$ converge in distribution respectively to $\sup_{\lambda\in[0,1]}|Z(\lambda)|$ and $\int_0^1 Z^2(\lambda)d\lambda$.

Now, we show that $\int_0^1 Z^2(\lambda)d\lambda$ has the same distribution as a sum of weighted iid chi-square distribution with one degree of freedom. Noting that Δ is a Mercer kernel (it is easy to prove), it follows from Riesz and Nagy (1972) that the integral operator defined by (8) admits eigenvalues $\zeta_1 \geq \zeta_2 \geq \ldots \geq 0$ with associated eigenfunctions $e_1, e_2, \ldots$ forming an orthonormal basis of $L^2[0, 1]$, the set of square integrable functions on $[0, 1]$. From this result, it is an easy matter that the zero-mean Gaussian process Z as a function in $L^2[0, 1]$, has the Karhunen-Loève representation

$$Z(\lambda) = \sum_{j\geq 1} N_j e_j(\lambda), \lambda \in [0, 1],$$

with the independent random variables' N_j's defined as $N_j = \int_{[0,1]} Z(\lambda)e_j(\lambda)d\lambda \sim \mathcal{N}(0, \zeta_j)$. One easily deduces from this that in distribution

$$\int_0^1 Z^2(\lambda)d\lambda = \sum_{j\geq 1} \zeta_j \chi_j^2,$$

where the χ_j^2's are iid chi-square random variables with one degree of freedom.

(ii) From Theorem 2 and the continuous the mapping theorem, $T_{1,n}$ and $T_{2,n}$ converge in distribution respectively to $\sup_{\lambda\in[0,1]}|(1-\lambda)\lambda A + Z(\lambda)|$ and $\int_0^1 |(1-\lambda)\lambda A + Z(\lambda)|^2 d\lambda$.

For the same reasons as above, one has the decomposition

$$(1-\lambda)\lambda A + Z(\lambda) = \sum_{j\geq 1} \widetilde{N}_j e_j(\lambda), \lambda \in [0, 1],$$

with the independent random variables $\widetilde{N}_j$'s defined as $\widetilde{N}_j = \int_0^1 Z(\lambda)e_j(\lambda)d\lambda \sim \mathcal{N}(\rho_j, \zeta_j)$.

It follows from this that, in distribution,

$$\int_0^1 |(1-\lambda)\lambda A + Z(\lambda)|^2 d\lambda = \sum_{j\geq 1} \zeta_j \chi_j^{*2},$$

where the χ_j^{*2}'s are non-central iid chi-square random variables with one degree of freedom and non-centrality parameter $\rho_j^2 \zeta_j^{-1}$.

(iii) The last part is an easy consequence of Theorem 3.

□

Define σ^2 by

$$\sigma^2 = \mathbb{V}\mathrm{ar}(h_{F,1}(X_1)) + 2\sum_{j=1} \mathbb{C}\mathrm{ov}(h_{F,1}(X_1), h_{F,1}(X_{1+j})).$$

Corollary 1 *Assume that the assumptions of Theorem 2 hold, and that h is such that its associated $h_{F,1}$ and $h_{F,2}$ satisfy $h_{F,1}(x) = -h_{F,2}(x)$. Then*

(i) Under H_0, as n tends to infinity, one has the following convergences in distribution

$$T_{1,n} \longrightarrow \sigma \sup_{\lambda\in[0,1]} \left|W^0(\lambda)\right|$$
$$T_{2,n} \longrightarrow \sigma^2 \sum_{j\geq 1} \frac{1}{j^2\pi^2}\chi_j^2.$$

(ii) Under $H_{1,n}$, as n tends to infinity, one has the following convergences in distribution

$$T_{1,n} \longrightarrow \sup_{\lambda\in[0,1]} \left|(1-\lambda)\lambda A + \sigma W^0(\lambda)\right|$$
$$T_{2,n} \longrightarrow \sum_{j\geq 1} \frac{1}{j^2\pi^2}\chi_j^{*2},$$

where W^0 is the Brownian bridge on $[0, 1]$, the χ_j^2's and χ_j^{*2}'s are as in Theorem 4 but the non-centrality parameters are $2A^2\left\{2[1-(-1)^j]/j\pi\right\}^2\sigma^{-2}$.

Proof i- If h is such that its associated $h_{F,1}$ and $h_{F,2}$ satisfy $h_{F,1}(x) = -h_{F,2}(x)$, then from the proof of Theorem 1 one sees that $W_1(\lambda) = -W_2(\lambda)$. Whence, the representation of the limit process Z in that theorem reduces to

$$Z(\lambda) = W_1(\lambda) - \lambda W_1(1),\ \lambda \in [0, 1],$$

where for any $\lambda \in [0, 1]$, $W_1(\lambda) = \sigma W^0(\lambda)$ with $W^0(\lambda)$ standing for the Brownian bridge on $[0, 1]$. Thus, again, from the continuous mapping theorem, one has that $T_{1,n}$ and $T_{2,n}$ converge in distribution respectively to $\sup_{\lambda\in[0,1]} |\sigma W^0(\lambda)|$ and $\sigma^2 \int_0^1 |W^0(\lambda)|^2 d\lambda$. Using the Karhunen-Loève expansion of the Brownian bridge given e.g. in Shorack and Wellner (1986) or Pycke (2001), one has $\zeta_j = \frac{1}{j^2\pi^2}$ and $e_j(\lambda) = \sqrt{2}\sin(j\pi\lambda)$, $j \geq 1$, and the convergence in distribution of $T_{2,n}$ stated in the theorem follows by elementary computations.
ii- From Theorem 4, $T_{1,n}$ and $T_{2,n}$ converge in distribution respectively to $\sup_{\lambda\in[0,1]} |(1-\lambda)\lambda A + \sigma W^0(\lambda)|$ and $\int_0^1 |(1-\lambda)\lambda A + \sigma W^0(\lambda)|^2 d\lambda$.

The convergence result stated for $T_{2,n}$ in the corollary results from the Karhunen-Loève expansion of the Gaussian process $(1-\lambda)\lambda A + \sigma W^0(\lambda)$, as sketched in Sect. 4, and for which more details can be found in Ngatchou-Wandji (2009).

Remark 4 It is easy to check that anti-symmetric kernels h are such that their associated $h_{F,1}$ and $h_{F,2}(x)$ satisfy the property $h_{F,1}(x) = -h_{F,2}(x)$.

Remark 5 In the context of Corollary 1, under $\mathcal{H}_0$, the asymptotic distributions of $T_{1,n}$ and $T_{2,n}$ can be approximated, as done in Sect. 4. Under $\mathcal{H}_{1,n}$ it is not easy to do this for the asymptotic distribution of $T_{1,n}$. This is a great disadvantage over $T_{2,n}$ whose asymptotic distribution under $\mathcal{H}_0$ as well as under $\mathcal{H}_{1,n}$ can be approximated, even for more general kernel h, by using Theorem 4, and proceeding in a way fully described in Ngatchou-Wandji (2009).

4 Practical considerations

4.1 Numerical simulations

Here, we apply our results to detecting a change in the mean and/or in the variance and/or in the correlation of data from some simple models. Concretely, we check the difference between the distributions of the observations only by checking the differences between their means and/or variances and/or autocorrelations. In this purpose, we consider the kernels $h(x, y) = 1\!\!1(x < y)$ and $h(x, y) = x - y$. For $h(x, y) = 1\!\!1(x < y)$, from simple computations one shows that $\theta(F, F) = 1/2$ and $h_{F,1}(x) = 1/2 - F(x) = -h_{F,2}(x)$. For $h(x, y) = x - y$, it is a trivial matter that $\theta(F, F) = 0$ and that by Remark 4, $h_{F,1}(x) = -h_{F,2}(x)$ as h is anti-symmetric. Consequently, Corollary 1 holds for these two kernels whose corresponding σ^2 are respectively

$$\sigma_1^2 = \sigma^2 = \mathbb{V}\text{ar}\,[F(X_1)] + 2\sum_{j\geq 1}\mathbb{C}\text{ov}(F(X_1), F(X_{1+j}))$$

and

$$\sigma_2^2 = \sigma^2 = \mathbb{V}\text{ar}\,(X_1) + 2\sum_{j\geq 1}\mathbb{C}\text{ov}(X_1, X_{1+j}).$$

We sampled 1000 sets of $n = 200$ data $X_1, X_2, \ldots, X_n$ from the model

$$X_i = \begin{cases} \varepsilon_i & i = 1, \ldots, 100 \\ \mu + \rho X_{i-1} + \omega\varepsilon_i & i = 101, \ldots, 200, \end{cases} \tag{9}$$

where μ is a real number, ω is a positive number, the ε_i's are iid and for all $i = 1, \ldots, 200$, $\varepsilon_i \sim \mathcal{N}(0, 1)$, or $\varepsilon_i \sim \mathcal{T}(3)$ (Student distribution with 3 degrees of freedom), or $\varepsilon_i = \mathcal{E}_i - 1$ with $\mathcal{E}_i \sim \mathcal{E}(1)$ ($\mathcal{E}(1)$ exponential distribution with parameter 1).

We first apply our Kolmogorov-Smirnov and Cramér-von Mises type tests to testing $\mu = 0$ against $\mu \neq 0$ for $\omega = 1$ and $\rho = 0$ (testing a change in the mean of a shifted white noise). Next, we apply the two tests to testing $\omega = 1$ against $\omega \neq 1$ for $\mu = 0$ and $\rho = 0$ (testing change in the variance of a white noise). Finally, we consider testing $\rho = 0$ against $\rho \neq 0$ for $\mu = 0$ and $\omega = 1$ (testing a change in the autocorrelation of an AR(1) model).

For $j = 1, 2$ let $c_{\alpha j}$ be the critical value at level of significance $\alpha \in [0, 1]$, of the test based on $T_{j,n}$. Then the empirical power of the test can be computed as the ratio of the number of samples for which $T_{j,n} > c_{\alpha j}$ over the number of replications (taken here to be 1000). However, the critical values of our tests are difficult to compute. For this reason, as in Ngatchou-Wandji (2009), we use the p-value method as follows : instead of counting the number of samples for which $T_{j,n} > c_{\alpha j}$, we rather count the number of samples for which the p-value is less than α. Denoting by $\widehat{\sigma}_j$ an estimator of σ_j when this number is unknown, from Corollary 1, the p-values of our tests are given respectively by

$$p_1 = \mathbb{P}\left(\sup_{\lambda\in[0,1]} |W^0(\lambda)| > \frac{T_{1,n}}{\widehat{\sigma}_1}\right)$$

and

$$p_2 = \mathbb{P}\left(\sum_{j\geq 1} \frac{1}{j^2\pi^2}\chi_j^2 > \frac{T_{2,n}}{\widehat{\sigma}_2^2}\right),$$

where we recall that W^0 is the Brownian bridge on [0, 1] and the χ_j^2's are iid chi-square random variables with one degree of freedom. Note that for iid X_i's the covariance terms in the expressions of σ^2 vanish, so that under $\mathcal{H}_0$, for $h(x, y) = \mathbb{1}(x < y)$, $\sigma^2 = \mathbb{V}\mathrm{ar}[F(X_1)] = 1/12$ and needs not be estimated, while for $h(x, y) = x - y$, $\sigma^2 = \mathbb{V}\mathrm{ar}(X_1)$ which can be estimated by the sample variance as done in the trials. From Billingsley (1999) p. 103, one has

$$p_1 = -2\sum_{|j|=0}^{\infty} (-1)^j \exp\left[-2j^2\left(\frac{T_{1,n}}{\widehat{\sigma}_1}\right)^2\right].$$

Truncating the sum to the most significant terms yields an approximation for p_1. Also, truncating the sum in the expression of p_2, many well known results can be used for approximating p_2. Here, we use those of Imhof (1961).

On any of the graphics below, the green color represents the level of significance of the test. On Fig. 2, the blue color represents the empirical power of the test based on $T_{1,n}$ while the red one represents that of the test based on $T_{2,n}$. The upper graphics in Fig. 2 display the empirical power functions of both tests as functions of $\mu \in \{0, 0.1, 0.2, \ldots, 1.1\}$, for observations from (9) with $\rho = 0$, $\omega = 1$ and standard Gaussian noises. The first graphic corresponds to the kernel $h(x, y) = \mathbb{1}(x < y)$, while the second corresponds to the kernel $h(x, y) = x - y$. On both graphics, one can observe that at $\mu = 0$ the test based on $T_{2,n}$ estimates the nominal level of the test more accurately than the one based on $T_{1,n}$. More over, it has a larger power for smaller values of the mean.

The lower graphics on Fig. 2 are respectively the power functions for the same kernels, for data from (9) with $\rho = 0$, $\omega = 1$ with Student noises with 3 degrees of freedom, and centered exponential noises (from exponential distribution with parameter 1). One can see on these graphics that the powers grow a bit more slowly than those in the upper graphics.

The upper graphics in Fig. 3 present the power functions of both tests as functions of $\omega^2 \in \{1, 1.2, 1.4, \ldots, 2, 2.2, \ldots, 2.8\}$, for observations from (9) with $\mu = 0$ and $\rho = 0$. From these graphics, one sees that for the kernel $h(x, y) = \mathbb{1}(x < y)$ both tests are not sensitive to any change in the variance, while they are for $h(x, y) = x - y$. Here, in contrast to the previous situations, the test based on $T_{1,n}$ does better than the one based on $T_{2,n}$.

The lower graphics in Fig. 3 exhibit the power functions of both tests as functions of $\rho \in \{0, .05, .1, .2, .3, .4, .5, .6, .7, .8, .85, .9, .95\}$, for observations from (9) with $\mu = 0$ and $\omega = 1$. One can see that for the kernels $h(x, y) = \mathbb{1}(x < y)$ and $h(x, y) = x - y$, both tests are able to detect change in the autocorrelation. In the vicinity of the null hypothesis the test based on $T_{2,n}$ does better than the one based on $T_{1,n}$, while far from this hypothesis, the test based on $T_{1,n}$ does better than that based on $T_{2,n}$.

We did a lot of trials. In particular, we studied the detection of a change in the mean of a shifted Student white noise with 3, 4 and 5 degrees of freedom ($\mathcal{T}(j)$, $j = 3, 4, 5$) associated with $h(x, y) = x - y$. Our tests did not detect a mean change in the $\mathcal{T}(3)$ case. This is likely due to the fact that with this kernel, the theoretical results assume the existence of moment of order larger than 2, which is not the case for $\mathcal{T}(3)$. For the $\mathcal{T}(4)$ and $\mathcal{T}(5)$ cases, a mean change was detected. We do not present the results as they were very similar to those presented.

4.2 Concluding remarks

The main theoretical results in this paper are the same as those of Dehling et al. (2015). But they are established for more general kernels. In addition to the traditional Kolmogorov-Smirnov (KS) type test used for change-point detection, a Cramér-von Mises (CM) type test is studied. For the kernels and the data considered, the CM test seems to have better power properties than the KS test for detecting small changes in the mean of a shifted Gaussian white noise. For the kernel $h(x, y) = \mathbb{1}(x < y)$, both test are not sensitive to change in the variance of the observations studied. This may be explained by the fact that this function involves the rank of the observations rather than the observations themselves. Furthermore, the corresponding tests are associated with uniform random variables ($F(X_j) \sim \mathcal{U}[0, 1]$) through σ^2. This is not the case for the tests based on $h(x, y) = x - y$ which involves directly the observation and are related to their variances through σ^2 and for which change in the variance can be detected by the tests studied here (with a favor to KS). Our study also shows that for each of these kernels, our tests are able to detect changes in the autocorrelation of an AR(1) model. The KS test does better far from the null hypothesis, while the CM does better in the vicinity of the null hypothesis. However, the results seem to show that they are more adapted to detect changes in the mean than changes in the variance or autocorrelation. Indeed, in testing the mean, the power of the tests grow quickly to 1, as one moves away from the null hypothesis.

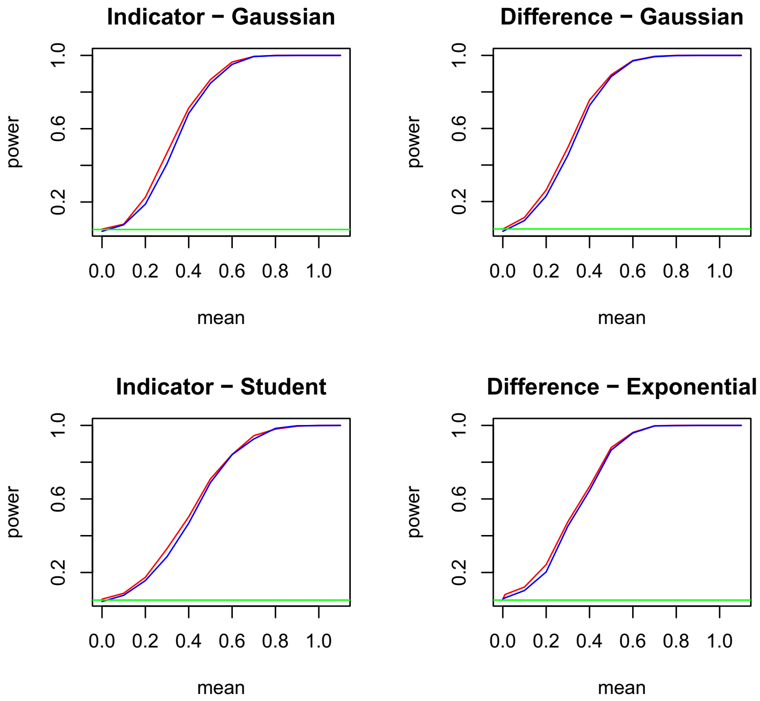

Fig. 2 Empirical power of CM test (red color). Empirical power of KS test blue color. Nominal level 5% green color. First row : change in the mean of a shifted Gaussian white noise respectively with the "indicator" and "difference" kernels. Second row : change in the mean of a shifted Student white noise with the "indicator" kernel, and change in the mean of a shifted centered exponential white noise with the "difference" kernel

5 Appendix: proofs of the results

5.1 Preliminary results

In this subsection, we prove some preliminary results necessary to the proofs of Theorems 1 and 2 .

Proposition 1 *Under the conditions of Theorem 1, we have, in probability*

$$n^{-3/2} \sup_{0\leq\lambda\leq 1} \left| \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} g_F(X_i, X_j) \right| \longrightarrow_{n\to\infty} 0.$$

Under the conditions of Theorem 2, we have, in probability

$$n^{-3/2} \sup_{0\leq\lambda\leq 1} \left| \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} g_{G^{(n)}}(X_i, Y_{nj}) \right| \longrightarrow_{n\to\infty} 0.$$

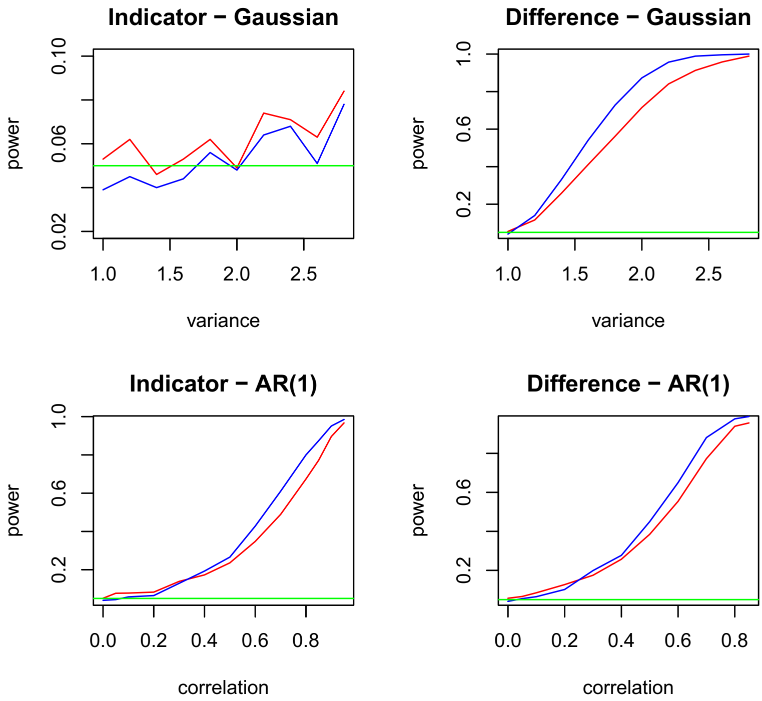

Fig. 3 Empirical power of CM test (red color). Empirical power of KS test blue color. Nominal level 5% green color. First row : change in the variance of a shifted Gaussian white noise respectively with the "indicator" and the "difference" kernels. Second row : change in the correlation of an AR(1) model respectively with the "indicator" and the "difference" kernels

Proof We only prove the first part. This needs two lemmas that we first state and prove.

Lemma 1 *Under the conditions of Theorem 1, there exists a Constant* $Cst > 0$ *such that*

$$\mathbb{E}\left\{\left[\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} g_F(X_i, X_j)\right]^2\right\} \leq Cst[n\lambda](n - [n\lambda]).$$

Proof of Lemma 1 We can write

$$\begin{aligned}\mathbb{E}\left\{\left[\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} g_F(X_i, X_j)\right]^2\right\} &\leq \sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} \mathbb{E}\left\{\left[g_F(X_i, X_j)\right]^2\right\} \\ &\quad + 2\sum_{1\leq i_1<i_2\leq[n\lambda]}\sum_{[n\lambda]+1\leq j_1<j_2\leq n} H\left[(i_1, j_1), (i_2, j_2)\right] \\ &= A_n(\lambda) + 2B_n(\lambda).\end{aligned}$$

where

$$H\left[(i_1, j_1), (i_2, j_2)\right] = \mathbb{E}\Big\{\left[g_F(X_{i_1,}X_{j_1}) - h_{F,1}(X_{i_1}) - h_{F,2}(X_{j_1}) + \theta(F,F)\right] \\ \times \left[g_F(X_{i_2}X_{j_2}) - h_{F,1}(X_{i_2}) - h_{F,2}(X_{j_2}) + \theta(F,F)\right]\Big\}.$$

From the integrability condition, we have

$$\sup_{i,j\in\mathbb{N}} \mathbb{E}\left\{\left[g_F(X_i, X_j)\right]^2\right\} \leq Cst,$$

then

$$A_n(\lambda) \leq Cst[n\lambda](n - [n\lambda]).$$

Since

$$\int_{\mathbb{R}} \left[g_F(x, y) - h_{F,1}(x) - h_{F,2}(y) + \theta(F,F)\right] dF(x) = 0$$

so from Lemma 1 of Yoshihara (1976), we have the following inequalities:

(a) If $1 \leq i_1 < j_1 \leq [n\lambda],\ [n\lambda]+1 \leq i_2 < j_2 \leq n$ and $i_2 - i_1 \geq j_2 - j_1$, then

$$H\left[(i_1, j_1), (i_2, j_2)\right] \leq Cst\beta^{\frac{\delta}{2+\delta}}(i_2 - i_1).$$

Then we deduce

$$\sum_{1\leq i_1<i_2\leq[n\lambda]} \sum_{[n\lambda]+1\leq j_1<j_2\leq n} H\left[(i_1, j_1), (i_2, j_2)\right] \leq Cst[n\lambda](n-[n\lambda]) \sum_{k=1}^{n} k\beta^{\frac{\delta}{2+\delta}}(k)$$
$$Cst[n\lambda](n-[n\lambda]) \sum_{k=1}^{n} k\beta^{\frac{\delta}{2+\delta}}(k) \leq Cst[n\lambda](n-[n\lambda]),$$

where $k = j_2 - i_2$.

Suppose k fixed, we have $[n\lambda]$ ways to choose i_1, once i_1 is chosen we have one way to choose $i_2 = i_1 + k$. For j_1 we have $n - [n\lambda]$ ways to choose j_1 and then for each j_1, j_2 need to be in the interval $[j_1, j_1 + k]$ and there are exactly k integers in such interval.

(b) Similarly, if $1 \leq i_1 < j_1 \leq [n\lambda],\ [n\lambda]+1 \leq i_2 < j_2 \leq n$ and $i_2 - i_1 \leq j_2 - j_1$, then

$$H\left[(i_1, j_1), (i_2, j_2)\right] \leq M_0\beta^{\frac{\delta}{2+\delta}}(j_2 - j_1).$$

Thus, we deduce that

$$B_n(\lambda) \leq Cst[n\lambda](n - [n\lambda])$$

and Lemma 1 is proved. □

We now define the process $\mathcal{G}_n(\lambda)$, $0 \le \lambda \le 1$ by

$$\mathcal{G}_n(\lambda) = n^{-3/2} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} g_F(X_i, X_j), 0 \le \lambda \le 1.$$

Lemma 2 *Under the conditions of Theorem 1, we have*

$$\mathbb{E}\Big(\Big|\mathcal{G}_n(\lambda) - \mathcal{G}_n(\lambda')\Big|^2\Big) \le \frac{Cst}{n}(\lambda - \lambda'), \text{ for all } 0 \le \lambda' \le \lambda \le 1.$$

Proof of Lemma 2 We can write

$$\mathbb{E}\big(\big|\mathcal{G}_n(\lambda) - \mathcal{G}_n(\lambda')\big|^2 \big) \le 2n^{-3}\mathbb{E}\left\{ \left[\sum_{i=1}^{[n\lambda']} \sum_{j=[n\lambda']+1}^{[n\lambda]} g_F(X_i, X_j) \right]^2 \right\}$$
$$+2n^{-3}\mathbb{E}\left\{ \left[\sum_{i=1+[n\lambda']}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} g_F(X_i, X_j) \right]^2 \right\}.$$

From Lemma 1, we deduce that

$$\mathbb{E}\left(\Big|\mathcal{G}_n(\lambda) - \mathcal{G}_n(\lambda')\Big|^2\right) \le \frac{Cst}{n^3}\{[n\lambda']([n\lambda] - [n\lambda']) + ([n\lambda] - [n\lambda'])(n - [n\lambda'])\}$$
$$\times \frac{Cst}{n}(\lambda - \lambda')$$

and Lemma 2 is proved. □

From Lemma 2, we deduce that

$$\mathbb{P}\left(\big|\mathcal{G}_n(\lambda) - \mathcal{G}_n(\lambda')\big| \ge \epsilon\right) \le \frac{Cst}{\epsilon^2 n}(\lambda - \lambda')$$

for all $\epsilon > 0$. It implies for $0 \le l_1 \le l_2 \le n$ with $l_1, l_2, n \in \mathbb{N}$,

$$\mathbb{P}\left(\left|\mathcal{G}_n\left(\frac{l_2}{n}\right) - \mathcal{G}_n\left(\frac{l_1}{n}\right)\right| \ge \epsilon\right) \le \frac{Cst}{\epsilon^2 n}(l_2 - l_1) \le \frac{Cst}{\epsilon^2 n^{\frac{5}{3}}}(l_2 - l_1)^{\frac{4}{3}}.$$

Consider the partial sum process defined by $S_0 = 0$ and $S_i = \sum_{j=1}^{i} A_j$ where $A_j = \mathcal{G}_n(\frac{j}{n}) - \mathcal{G}_n(\frac{j-1}{n})$ if $1 \le j \le n-1$ and 0 otherwise. It results that $S_i = \mathcal{G}_n(\frac{i}{n})$. The last inequality is equivalent to

$$\mathbb{P}\big(\big|S_{l_2} - S_{l_1}\big| \ge \epsilon\big) \le \frac{Cst}{\epsilon^2}\left(\frac{l_2 - l_1}{n^{\frac{5}{4}}}\right)^{\frac{4}{3}}.$$

From Theorem 10.2 of Billingsley (1999), we easily deduce that

$$\mathbb{P}\Big(\max_{1 \le i \le n-1} |S_i| \ge \epsilon \Big) \le \frac{Cst}{\epsilon^2} \left(\frac{l_2 - l_1}{n^{\frac{5}{4}}} \right)^{\frac{4}{3}} \le \frac{Cst}{\epsilon^2 n^{\frac{1}{3}}}$$

which implies that, in probability,

$$n^{-3/2} \sup_{0 \le \lambda \le 1} \left| \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} g_F(X_i, X_j) \right| \longrightarrow_{n\to\infty} 0.$$

This completes the proof of Proposition 1. □

We need the following result proved by Oodaira and Yoshihara (1972).

Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be a strictly stationary sequence of zero-mean random variables, and let

$$\sigma_*^2 = \mathbb{E}(\xi_1^2) + 2 \sum_{i=1}^{\infty} \mathbb{E}\left(\xi_1 \xi_{i+1}\right).$$

Proposition 2 *Assume* $\mathbb{E}\left(|\xi_i|^{2+\delta}\right) < \infty$ *for some positive* δ *and* $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ *is* α*-mixing with* α*-rate satisfying*

$$\sum_{i=1}^{\infty} [\alpha(i)]^{\frac{\delta}{2+\delta}} < \infty.$$

Then $\sigma_*^2 < \infty$.

If $\sigma_* > 0$*, then the sequence of processes*

$$S_n(\lambda) = \frac{1}{\sigma_* \sqrt{n}} \sum_{i=1}^{[n\lambda]} \xi_i, \lambda \in [0, 1]$$

converges weakly to a Wiener measure on $(D, \mathcal{D})$*, where* $\mathcal{D}$ *is the* σ*-fields of Borel sets for the Skorohod topology.*

Proof See the proof of Theorem 2 of Oodaira and Yoshihara (1972).

Proposition 3 *Under the conditions of Theorem 1, we have*

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} \begin{pmatrix} h_{F,1}(X_i) \\ h_{F,2}(X_i) \end{pmatrix} \right\}_{0 \le \lambda \le 1} \longrightarrow_{n\to\infty} \left\{ \begin{pmatrix} W_1(\lambda) \\ W_2(\lambda) \end{pmatrix} \right\}_{0 \le \lambda \le 1}. \tag{10}$$

Under the conditions of Theorem 2, we have

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} \begin{pmatrix} h_{F,1}(X_i) \\ h_{G^{(n)},2}(Y_{ni}) \end{pmatrix} \right\}_{0 \le \lambda \le 1} \longrightarrow_{n\to\infty} \left\{ \begin{pmatrix} W_1(\lambda) \\ W_2(\lambda) \end{pmatrix} \right\}_{0 \le \lambda \le 1}. \tag{11}$$

Proof We only prove (11). The part relating to (10) involves a series which is not a triangular array. It is more easier to handle than (11).

For establishing (11), we need to establish a finite-dimensional convergence and a tightness results.

Starting by the finite-dimensional convergence, by the Cramér-Wold device it suffices to show that for any $k \in \mathbb{N}^*$, any $a_j, b_j, \lambda_j \in \mathbb{R}$, $a_1 < \ldots < a_k$, $b_1 < \ldots < b_k$, $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_k = 1$

$$\sum_{j=1}^{k} \frac{1}{\sqrt{n}} \sum_{i=[n\lambda_{j-1}]+1}^{[n\lambda_j]} \left[a_j h_{F,1}(X_i) + b_j h_{G^{(n)},2}(Y_{ni})\right]$$

converges in distribution to a Gaussian random variable.

For that, we need the following lemma.

Lemma 3 *(Harel and Puri 1989) Let $\{X_{ni}\}$ be a sequence of zero-mean absolutely regular random variables (rv)'s with rates satisfying*

$$\sum_{n\geq} \left[\beta(n)\right]^{\delta/(2+\delta)} < \infty \text{ for some } \delta > 0. \tag{12}$$

Suppose that for any κ, there exists a sequence $\{Y^{\kappa}_{ni}\}$ of rv's satisfying (12) such that

$$\sup_{n\in\mathbb{N}} \max_{0\leq i\leq n} |Y^{\kappa}_{ni}| \leq B_{\kappa} < \infty, \tag{13}$$

where B_{κ} is some positive constant

$$\sup_{n\in\mathbb{N}} \max_{0\leq i\leq n} \mathbb{E}\left(|X^*_{ni} - Y^{\kappa}_{ni}|^{2+\delta}\right) \longrightarrow 0 \text{ as } \kappa \to \infty \tag{14}$$

$$\frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^{n} X^*_{ni}\right)^2\right] \longrightarrow c \text{ as } n \to \infty, \tag{15}$$

where c is some positive constant

$$\frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^{n} Y^{\kappa}_{ni} - \mathbb{E}(Y^{\kappa}_{ni})\right)^2\right] \longrightarrow c_{\kappa} \text{ as } n \to \infty, \tag{16}$$

where c_{κ} is some constant > 0

$$c_{\kappa} \longrightarrow c \text{ as } \kappa \to \infty. \tag{17}$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} X^*_{ni}$$

converges in distribution to the normal distribution with mean 0 and variance c.

Without loss of generality, we take $k = 2$ and $0 = \lambda_0 < \lambda_1 < \lambda_2 = 1$, $a_1 < a_2$, $b_1 < b_2$.
The assumption (12) readily holds from (5).

Define, for $j = 1, 2$,

$$\psi_{ni}^{(j)} = a_j h_{F,1}(X_i) + b_j h_{G^{(n)},2}(Y_{ni}).$$

For establishing (15), we need proving that, as n tends to infinity,

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{[n\lambda_1]}\psi_{ni}^{(1)} + \sum_{i=[n\lambda_1]+1}^{n}\psi_{ni}^{(2)}\right\}\right)^2\right]$$

tends to some positive constant c.
We have

$$\mathbb{E}\left\{\left[\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{[n\lambda_1]}\psi_{ni}^{(1)} + \sum_{i=[n\lambda_1]+1}^{n}\psi_{ni}^{(2)}\right)\right]^2\right\} = \frac{1}{n}\left\{\mathbb{E}\left[\left(\sum_{i=1}^{[n\lambda_1]}\psi_{ni}^{(1)}\right)^2\right]\right.$$
$$\left.+\mathbb{E}\left[\left(\sum_{i=1}^{[n\lambda_1]}\psi_{ni}^{(1)}\right)\left(\sum_{i=[n\lambda_1]+1}^{n}\psi_{ni}^{(2)}\right)\right] + \mathbb{E}\left[\left(\sum_{i=[n\lambda_1]+1}^{n}\psi_{ni}^{(2)}\right)^2\right)\right\}.$$

Since the random variables $\psi_{ni}^{(1)}$ and $\psi_{ni}^{(2)}$ are centered, we obtain

$$\mathbb{E}\left[\left(\sum_{i=1}^{[n\lambda_1]}\psi_{ni}^{(1)}\right)^2\right] = [n\lambda_1]\mathbb{E}\left[\left(\psi_{n1}^{(1)}\right)^2\right] + 2\sum_{i=1}^{[n\lambda_1]}\sum_{j=1}^{[n\lambda_1]-i}\mathbb{E}\left[\left(\psi_{ni}^{(1)}\psi_{n,i+j}^{(1)}\right)\right].$$

From the condition of Theorem 2, we deduce that $\mathbb{E}\left[\left(\psi_{n1}^{(1)}\right)^{2+\delta}\right] < \infty$, which implies that

$$\sup_{n,i,j\geq 1}\left|\mathbb{E}\left(\psi_{ni}^{(1)}\psi_{n,i+j}^{(1)}\right)\right| \leq \beta^{\frac{\delta}{2+\delta}}(j)\left\{\mathbb{E}\left(\left|\psi_{ni}^{(1)}\right|^{2+\delta}\right)\right\}^{\frac{1}{2+\delta}}\left\{\mathbb{E}\left(\left|\psi_{n,i+j}^{(1)}\right|^{2+\delta}\right)\right\}^{\frac{1}{2+\delta}}.$$

We get

$$\mathbb{E}\left[\left(\sum_{i=1}^{[n\lambda_1]}\psi_{ni}^{(1)}\right)^2\right] \leq [n\lambda_1]\mathbb{E}\left[\left(\psi_{n1}^{(1)}\right)^2\right] + 2[n\lambda_1]\sum_{j=1}^{[n\lambda_1]}\beta^{\frac{\delta}{2+\delta}}(j)M^2,$$

where $M = \sup_{n\geq 1}\left\{\mathbb{E}\left[\left(\psi_{n1}^{(1)}\right)^{2+\delta}\right]\right\}^{\frac{1}{2+\delta}}$.
It results that

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\left\{[n\lambda_1]\mathbb{E}\left[\left(\psi_{n1}^{(1)}\right)^2\right]+2\sum_{i=1}^{[n\lambda_1]}\sum_{j=1}^{[n\lambda_1]-i}\left|\mathbb{E}\left(\psi_{ni}^{(1)}\psi_{n,i+j}^{(1)}\right)\right|\right\}\\
&\quad\leq\lambda_1\left\{\mathbb{E}\left[\left(\psi_{n1}^{(1)}\right)^2\right]+2\sum_{j=1}^{\infty}\beta^{\frac{\delta}{2+\delta}}(j)M^2\right\}. \qquad (18)
\end{aligned}$$

We also have

$$\mathbb{E}\left[\left(\sum_{i=1}^{[n\lambda_1]}\psi_{ni}^{(1)}\right)\left(\sum_{i=[n\lambda_1]+1}^{n}\psi_{ni}^{(2)}\right)\right]=\sum_{i=1}^{[n\lambda_1]}\sum_{j=[n\lambda_1]+1}^{n}\mathbb{E}\left(\psi_{ni}^{(1)}\psi_{nj}^{(2)}\right).$$

From

$$\sup_{n\geq 1}\sup_{i,j\geq 1}\left|\mathbb{E}\left(\psi_{ni}^{(1)}\psi_{nj}^{(2)}\right)\right|\leq\beta^{\frac{\delta}{2+\delta}}(j-i)MM^*,$$

where $M^* = \sup_{n\geq 1}\left\{\mathbb{E}\left[\left(\psi_{n1}^{(2)}\right)^{2+\delta}\right]\right\}^{\frac{1}{2+\delta}}$, it results that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{[n\lambda_1]}\sum_{j=[n\lambda_1]+1}^{n}\left|\mathbb{E}\left(\psi_{ni}^{(1)}\psi_{nj}^{(2)}\right)\right|\leq\lambda_1\sum_{j=1}^{\infty}\beta^{\frac{\delta}{2+\delta}}(j)MM^*. \qquad (19)$$

Similarly, we get

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\left(\sum_{i=[n\lambda_1]+1}^{n}\psi_{ni}^{(2)}\right)^2\right]\leq(1-\lambda_1)\left\{\mathbb{E}\left[\left(\psi_{n1}^{(1)}\right)^2\right]+2\sum_{j=1}^{\infty}\beta^{\frac{\delta}{2+\delta}}(j)(M^*)^2\right\}. \qquad (20)$$

From (18)-(20), we deduce (15).
Now, we turn to proving (14). For all $i\geq 1$, and for any $\kappa>0$, define

$$\psi_{ni}^{(j),\kappa}=\begin{cases}\psi_{ni}^{(j)} & \text{if } \left|\psi_{ni}^{(j)}\right|\leq\kappa\\ 0 & \text{if } \left|\psi_{ni}^{(j)}\right|\geq\kappa,\ j=1,2.\end{cases}$$

It is immediate that

$$\sup_{n\geq 1}\sup_{i\geq 1}\left|\psi_{ni}^{(j)}\right|\leq\kappa<\infty.$$

It results from the integrability condition in Theorem 2 that the sequences $\{\psi_{ni}^{(j)};\ i \geq 1,\ j = 1, 2\}$ are uniformly integrable.
Whence

$$\sup_{i\geq 1} \mathbb{E}\left(\left|\psi_{ni}^{(j)} - \psi_{ni}^{(j),\kappa}\right|^{2+\delta}\right) \longrightarrow 0 \text{ as } \kappa \to \infty,\ j = 1, 2$$

and (14) is proved.
The proof of (16), that is

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{[n\lambda_1]}\left[\psi_{ni}^{(1),\kappa} - \mathbb{E}\left(\psi_{ni}^{(1),\kappa}\right)\right] + \sum_{i=[n\lambda_1]+1}^{n}\left[\psi_{ni}^{(2),\kappa} - \mathbb{E}\left(\psi_{ni}^{(2),\kappa}\right)\right]\right\}\right)^2\right] = c_\kappa,$$

where c_κ is some positive constant, is similar to that of (15).
It remains to prove (17).

For any i, $j = 1, 2$, denote by $\psi_i^{(j),\kappa}$ the counterpart of $\psi_{ni}^{(j),\kappa}$ obtained by substituting the Y_{ni}'s for the X_i's.
We have

$$\begin{aligned}
c_\kappa = {} & \lambda_1 \mathbb{E}\left\{\left[\psi_1^{(1),\kappa} - \mathbb{E}(\psi_1^{(1),\kappa})\right]^2\right\} + 2\lambda_1 \sum_{i=1}^{\infty} \mathbb{E}\left\{\left[\psi_1^{(1),\kappa} - \mathbb{E}(\psi_1^{(1),\kappa})\right]\left[\psi_{i+1}^{(1),\kappa} - \mathbb{E}(\psi_{i+1}^{(1),\kappa})\right]\right\} \\
& + \lambda_1 \sum_{i=1}^{\infty} \mathbb{E}\left\{\left[\psi_1^{(1),\kappa} - \mathbb{E}(\psi_1^{(1),\kappa})\right]\left[\psi_{i+1}^{(2),\kappa} - \mathbb{E}(\psi_{i+1}^{(2),\kappa})\right]\right\} \\
& + (1-\lambda_1)\mathbb{E}\left\{\left[\psi_1^{(2),\kappa} - \mathbb{E}(\psi_1^{(2),\kappa})\right]^2\right\} \\
& + 2(1-\lambda_1) \sum_{i=1}^{\infty} \mathbb{E}\left\{\left[\psi_1^{(2),\kappa} - \mathbb{E}(\psi_1^{(2),\kappa})\right]\left[\psi_{i+1}^{(1),\kappa} - \mathbb{E}(\psi_{i+1}^{(2),\kappa})\right]\right\}.
\end{aligned}$$

By the Lebesgue dominated convergence theorem, one obtains

$$\begin{aligned}
& \mathbb{E}\left\{\left[\psi_1^{(1),\kappa} - \mathbb{E}\left(\psi_1^{(1),\kappa}\right)\right]^2\right\} \longrightarrow \mathbb{E}\left[\left(\psi_1^{(1)}\right)^2\right] \text{ as } \kappa \to \infty, \\
& \mathbb{E}\left\{\left[\psi_1^{(1),\kappa} - \mathbb{E}(\psi_1^{(1),\kappa}\right]\left[(\psi_{i+1}^{(1),\kappa} - \mathbb{E}(\psi_{i+1}^{(1),\kappa})\right]\right\} \longrightarrow \mathbb{E}\left(\psi_1^{(1)}\psi_{i+1}^{(1)}\right) \text{ as } \kappa \to \infty, \\
& \mathbb{E}\left\{\left[\psi_1^{(1),\kappa} - \mathbb{E}(\psi_1^{(1),\kappa})\right]\left[\psi_{i+1}^{(2),\kappa} - \mathbb{E}(\psi_{i+1}^{(2),\kappa})\right]\right\} \longrightarrow \mathbb{E}\left(\psi_1^{(2)}\psi_{i+1}^{(2)}\right) \text{ as } \kappa \to \infty, \\
& \mathbb{E}\left\{\left[\psi_1^{(2),\kappa} - \mathbb{E}(\psi_1^{(2),\kappa})\right]^2\right\} \longrightarrow \mathbb{E}\left[\left(\psi_1^{(2)}\right)^2\right] \text{ as } \kappa \to \infty
\end{aligned}$$

and

$$\mathbb{E}\left\{\left[\psi_1^{(2),\kappa} - \mathbb{E}(\psi_1^{(2),\kappa})\right]\left[\psi_{i+1}^{(2),\kappa} - \mathbb{E}(\psi_{i+1}^{(2),\kappa})\right]\right\} \longrightarrow \mathbb{E}\left(\psi_1^{(2)}\psi_{i+1}^{(2)}\right) \text{ as } \kappa \to \infty.$$

Therefore

$$\lim_{\kappa\to\infty} c_\kappa = c$$

and (17) is proved. Whence, the finite dimensional convergence is established.

For proving the tightness, we need the following Lemma.

Lemma 4 *(Phillips and Durlauf 1986) Probability measures on a product space are tight iff all the marginal probability measures are tight on the component spaces.*

It results from this lemma that it suffices to prove the tightness of each component of the sequence of processes in (11). It is immediate from Proposition 2 that the first is tight. For the second, define

$$\mathcal{M}_n(\lambda) = \sigma_{22}^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} h_{G^{(n)},2}(Y_{ni}).$$

If $\lambda_1 \le \lambda \le \lambda_2$, from the integral conditions and condition (5), there exists a constant C such that

$$\begin{aligned}
\mathbb{E}\left(|\mathcal{M}_n(\lambda) - \mathcal{M}_n(\lambda_1)|^2 |\mathcal{M}_n(\lambda_2) - \mathcal{M}_n(\lambda)|^2\right) &\le C\frac{1}{n^2}([n\lambda] - [n\lambda_1])([n\lambda_2] - [n\lambda]) \\
&\le C\frac{1}{n^2}([n\lambda_1] - [n\lambda])([n\lambda] - [n\lambda_2]) \\
&\le C\frac{1}{n^2}([n\lambda_2] - [n\lambda_1])^2 \\
&\le C(\lambda_2 - \lambda_1)^2.
\end{aligned}$$

If $\lambda_2 - \lambda_1 \ge 1/n$ the last inequality follows and if $\lambda_2 - \lambda_1 < 1/n$, then either λ_1 and λ lie in the same subinterval $[(i-1)/n, i/n]$ or else λ and λ_2 do. In either of these cases the left hand of last inequality vanishes. From Theorem 13.5 of Billingsley (1999), the process $\mathcal{M}_n$ is tight. This ends the proof of Proposition 3. □

5.2 Proof of Theorem 1

Using the Hoeffding decomposition, we can write $Z_n(\lambda)$ as

$$\begin{aligned}
Z_n(\lambda) &= n^{-3/2} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \left[h_{F,1}(X_i) + h_{F,2}(X_j) + g_F(X_i, X_j)\right] \\
&= n^{-3/2} \left[(n - [n\lambda]) \sum_{i=1}^{[n\lambda]} h_{F,1}(X_i) + [n\lambda] \sum_{j=[n\lambda]+1}^{n} h_{F,2}(X_j)\right] \\
&\quad + n^{-3/2} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} g_F(X_i, X_j). \qquad (21)
\end{aligned}$$

From Proposition 1, we have

$$n^{-3/2} \sup_{0 \le \lambda \le 1} \left| \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} g_F(X_i, X_j) \right| \longrightarrow_{n\to\infty} 0$$

in probability.
Thus, by Slutsky's lemma, it suffices to show that the sum of the first two terms

$$\left\{ n^{-3/2}(n - [n\lambda]) \sum_{i=1}^{[n\lambda]} h_{F,1}(X_i) + n^{-3/2}[n\lambda] \sum_{j=[n\lambda]+1}^{n} h_{F,2}(X_j) \right\}_{0 \le \lambda \le 1}$$

converges in distribution to the desired limit process.
It results from Proposition 2 that the process

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} h_{F,1}(X_i) \right\}_{0 \le \lambda \le 1}$$

converges weakly to a Brownian motion $\{W(\lambda)\}_{0 \le \lambda \le 1}$.
Proposition 3 yields

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} \begin{pmatrix} h_{F,1}(X_i) \\ h_{F,2}(X_i) \end{pmatrix} \right\}_{0 \le \lambda \le 1} \longrightarrow_{n\to\infty} \left\{ \begin{pmatrix} W_1(\lambda) \\ W_2(\lambda) \end{pmatrix} \right\}_{0 \le \lambda \le 1}$$

in distribution on the space $(D[0, 1])^2$ to $(D[0, 1])^2$.
Now, we consider the mapping defined by

$$\begin{pmatrix} x_1(\lambda) \\ x_2(\lambda) \end{pmatrix} \mapsto (1 - \lambda)x_1(\lambda) + \lambda(x_2(1) - x_2(\lambda)),\ 0 \le \lambda \le 1.$$

This is a continuous mapping from $(D[0, 1])^2$ to $D[0, 1]$. Whence,

$$\left\{ n^{-3/2}(n - [n\lambda]) \sum_{i=1}^{[n\lambda]} h_{F,1}(X_i) + n^{-3/2}[n\lambda] \sum_{j=[n\lambda]+1}^{n} h_{F,2}(X_j) \right\}_{0 \le \lambda \le 1} \longrightarrow_{n\to\infty} \{Z(\lambda)\}_{0 \le \lambda \le 1},$$

where for any $\lambda \in [0, 1]$,

$$Z(\lambda) = (1 - \lambda)W_1(\lambda) + \lambda[W_2(1) - W_2(\lambda)].$$

Whence, Theorem 1 is proved. □

5.3 Proof of Theorem 2

Now we prove Theorem 2. Under the conditions of Theorem 2, we have the following equality

$$\begin{aligned}Z_n(\lambda) &= Z_n^*(\lambda) - n^{-3/2}[n\lambda](n-[n\lambda])\theta(F,F)\\ &= n^{-3/2}[n\lambda](n-[n\lambda])\theta(F,G^{(n)}) - n^{-3/2}[n\lambda](n-[n\lambda])\theta(F,F)\\ &\quad + \frac{[n\lambda](n-[n\lambda])}{n^{3/2}}\frac{1}{[n\lambda]}\sum_{i=1}^{[n\lambda]} h_{F,1}(X_i)\\ &\quad + \frac{[n\lambda](n-[n\lambda])}{n^{3/2}}\frac{1}{(n-[n\lambda])}\sum_{j=[n\lambda]+1}^{n} h_{G^{(n)},2}(Y_{nj})\\ &\quad + \frac{1}{n^{3/2}}\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} g_F(X_i, Y_{nj}).\end{aligned}$$

From Proposition 1, we deduce that

$$n^{-3/2}\sup_{0\le\lambda\le1}\left|\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} g_F(X_i,Y_{nj})\right| \longrightarrow_{n\to\infty} 0$$

in probability.
From Proposition 2, we deduce that

$$n^{-1/2}\sum_{i=1}^{[n\lambda]} h_{F,1}(X_i)$$

converges weakly to the Brownian process $\{W_1(\lambda)\}_{0\le\lambda\le1}$ and

$$n^{-1/2}\sum_{j=[n\lambda]+1}^{n} h_{G^{(n)},2}(Y_{nj})$$

converges weakly to the Brownian process $\{W_2(1)-W_2(\lambda)\}_{0\le\lambda\le1}$.
We have also from the $\mathcal{H}_{1,n}$

$$\lim_{n\to\infty} n^{-3/2}[n\lambda](n-[n\lambda])\theta(F,G^{(n)}) - n^{-3/2}[n\lambda](n-[n\lambda])\theta(F,F) = \lambda(1-\lambda)A.$$

From Proposition 3, we obtain that

$$\left\{n^{-3/2}(n-[n\lambda])\sum_{i=1}^{[n\lambda]} h_{F,1}(X_i) + n^{-3/2}[n\lambda]\sum_{j=[n\lambda]+1}^{n} h_{G^{(n)},2}(Y_{nj})\right\}_{0\le\lambda\le1}$$

$$\longrightarrow_{n\to\infty} \left\{\widetilde{Z}(\lambda)\right\}_{0\leq\lambda\leq 1},$$

where for any $\lambda \in [0, 1]$,

$$\widetilde{Z}(\lambda) = (1-\lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)).$$

This establishes Theorem 2.

5.4 Proof of Theorem 3

Let $1 \leq [(n+1)t] \leq [n\lambda_0]$, then

$$\begin{aligned}
Z_n^*(t) &= n^{-3/2} \sum_{1\leq i<j\leq[n\lambda_0]} h(X_i, X_j) + n^{-3/2} \sum_{i=1}^{[n\lambda_0]} \sum_{j=[n\lambda_0]+1}^{n} h(X_i, X_j) \\
&\quad -n^{-3/2}\Bigg[\sum_{1\leq i<j\leq[(n+1)t]} h(X_i, X_j) + \sum_{[(n+1)t]+1\leq i<j\leq[n\lambda_0]} h(X_i, X_j) \\
&\quad + \sum_{[(n+1)t]+1\leq i\leq[n\lambda_0]} \sum_{[n\lambda_0]+1\leq j\leq n} h(X_i, X_j)\Bigg] \\
&= R_n^{(1)} + R_n^{(2)} - \left(R_n^{(3)} + R_n^{(4)} + R_n^{(5)}\right).
\end{aligned}$$

First we prove that

$$n^{-1/2} R_n^{(1)} \xrightarrow{a.s.}_{n\to\infty} \lambda_0^2\theta(F, F)/2.$$

From the Hoeffding decomposition (4), we have

$$\begin{aligned}
n^{-1/2} R_n^{(1)} &= \frac{1}{2n^2} U_{[n\lambda_0]} \\
&= \frac{[n\lambda_0]([n\lambda_0]-1)}{2n^2}\theta(F, F) + \frac{[n\lambda_0]([n\lambda_0]-1)}{n^2} \frac{2}{[n\lambda_0]} \sum_{i=1}^{[n\lambda_0]} h_{F,1}(X_i) \\
&\quad + \frac{[n\lambda_0]([n\lambda_0]-1)}{2n^2} U_{[n\lambda_0]}^{(2)}. \qquad (22)
\end{aligned}$$

As $(h_1^{(1)}(X_i))_{1\leq i\leq n}$ is stationary and ergodic, we have

$$\frac{1}{[n\lambda_0]} \sum_{i=1}^{[n\lambda_0]} h_{F,1}(X_i) \xrightarrow{a.s.}_{n\to\infty} 0.$$

For any $\varepsilon > 0$, put

$$A_{[n\lambda_0]} = \mathbb{P}\left(\left|U_{[n\lambda_0]}^{(2)}\right| > \varepsilon\right).$$

One has from Markov inequality and Lemma 2 of Yoshihara (1976)

$$\mathbb{P}\left(\left|U^{(2)}_{[n\lambda_0]}\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\mathbb{E}\left[\left(U^{(2)}_{[n\lambda_0]}\right)^2\right]$$
$$= \mathcal{O}([n\lambda_0]^{-1-\gamma}),\ \gamma > 0,$$

which implies

$$\sum_{i=1}^{[n\lambda_0]} A_{[i\lambda_0]} < \infty.$$

Then from Borel–Cantelli lemma

$$\frac{[n\lambda_0]([n\lambda_0]-1)}{2n^2}U^{(2)}_{[n\lambda_0]} \xrightarrow{a.s.}_{n\to\infty} 0.$$

Then from (22), we have

$$n^{-1/2}R_n^{(1)} \xrightarrow{a.s.}_{n\to\infty} \lambda_0^2\theta(F,F)/2.$$

Similarly, we prove

$$n^{-1/2}R_n^{(3)} \xrightarrow{a.s.}_{n\to\infty} t^2\theta(F,F)/2$$

and

$$n^{-1/2}R_n^{(4)} \overset{\mathcal{D}}{=} \sum_{1\le i<j\le[n\lambda_0]-[(n+1)t]} h(X_i,X_j) \xrightarrow{a.s.}_{n\to\infty} (t-\lambda_0)^2\theta(F,F)/2.$$

Now, we establish that

$$n^{-1/2}R_n^{(2)} \xrightarrow{a.s.}_{n\to\infty} \lambda_0(1-\lambda_0)\theta(F,G).$$

From (21), we have

$$n^{-1/2}R_n^{(2)} = \frac{1}{n^2}\sum_{i=1}^{[n\lambda_0]}\sum_{j=[n\lambda_0]+1}^{n}\{\theta(F,G)+h_{G,1}(X_i)+h_{G,2}(Y_j)+g_G(X_i,Y_j)\}$$
$$= \frac{[n\lambda_0](n-[n\lambda_0])}{n^2}\theta(F,G) + \frac{[n\lambda_0](n-[n\lambda_0])}{n^2}\frac{1}{[n\lambda_0]}\sum_{i=1}^{[n\lambda_0]} h_{F,1}(X_i)$$
$$+\frac{[n\lambda_0](n-[n\lambda_0])}{n^2}\frac{1}{(n-[n\lambda_0])}\sum_{j=[n\lambda_0]+1}^{n} h_{G,2}(Y_j)$$

$$+\frac{1}{n^2}\sum_{i=1}^{[n\lambda_0]}\sum_{j=[n\lambda_0]+1}^{n} g_F(X_i, Y_j), \tag{23}$$

where we recall that the Y_j's are random variables with cumulative distribution function G and satisfy (5).
From the ergodic theorem, we have that

$$\frac{1}{[n\lambda_0]}\sum_{i=1}^{[n\lambda_0]} h_{F,1}(X_i) \xrightarrow{a.s.}_{n\to\infty} 0.$$

and

$$\frac{1}{(n-[n\lambda_0])}\sum_{j=[n\lambda_0]+1}^{n} h_{G,2}(Y_j) \xrightarrow{a.s.}_{n\to\infty} 0.$$

From Lemma 2, we deduce that

$$\mathbb{E}\left\{\left[\frac{1}{n^2}\sum_{i=1}^{[n\lambda_0]}\sum_{j=[n\lambda_0]+1}^{n} g_F(X_i, Y_j)\right]^2\right\} \leq Cst[n\lambda_0](n-[n\lambda_0])n^{-4}.$$

From Markov inequality, we deduce for any $\epsilon > 0$ that

$$\mathbb{P}\left(\left|\frac{1}{n^2}\sum_{i=1}^{[n\lambda_0]}\sum_{j=[n\lambda_0]+1}^{n} g_F(X_i, Y_j)\right| > \epsilon\right) = \mathcal{O}(n^{-2}).$$

Also, by Borel–Cantelli Lemma one has

$$\frac{1}{n^2}\sum_{i=1}^{[n\lambda_0]}\sum_{j=[n\lambda_0]+1}^{n} g_F(X_i, Y_j) \xrightarrow{a.s.}_{n\to\infty} 0.$$

Similarly, we prove that

$$n^{-1/2}R_n^{(5)} \xrightarrow{\mathbb{P}} (\lambda_0 - t)(1-\lambda_0)\theta(F, G).$$

These observations clearly imply the first part of (7). The proof of its second part is similar. □

References

Amano T (2012) Asymptotic optimality of estimating function estimator for Charn model. Adv Decis Sci

Bardet J-M, Kengne W (2014) Monitoring procedure for parameter change in causal time series. J Multivar Anal 125:204–221
Bardet J-M, Wintenberger O (2009) Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. Ann Stat 37(5B):2730–2759
Bardet J-M, Kengne W, Wintenberger O (2012) Multiple breaks detection in general causal time series using penalized quasi-likelihood. Electr J Stat 6:435–477
Bhattacharyya GK, Johnson R (1968) Nonparametric tests for shifts at an unknown time point. J Multvar Anal 102(39):1731–1743
Bhattacharya P, Zhou H (2017) Nonparametric stopping rules for detecting small changes in location and scale families. From statistics to mathematical finance. Springer, Cham
Billingsley P (1999) Convergence of probability measures. Wiley, New York
Chen KM, Cohen A, Sackrowitz H (2011) Consistent multiple testing for change points. J Multvar Anal 102:1339–1343
Chernoff H, Zacks S (1964) Estimating the current mean of a normal distribution which is subjected to changes in time. Ann Math Stat 35:999–1018
Ciuperca G (2011) A general criterion to determine the number of change-points. Stat Probab Lett 81(8):1267–1275
Csörgő M, Horváth L (1987) Nonparametric tests for the changepoint problem. Stat Probab Lett 17:1–9
Csörgő M, Horváth L (1988) Invariance principales for changepoint problems. J Multivar Anal 17:151–168
Dehling H, Fried R, Garcia I, Wendler M (2015) Change-point detection under dependence based on two-sample **U**-statistics. asymptotic laws and methods in stochastics. Springer, New York
Dehling H, Rooch A, Taqqu M (2013) Non-parametric change-point tests for long-range dependent data. Scand J Stat 40:153–173
Dehling H, Franke B, Woerner J (2017a) Estimating drift parameters in a fractional ornstein uhlenbeck process with periodic mean. Stat Inference Stoch Process 20:1–14
Dehling H, Rooch A, Taqqu M (2017b) Power of change-point tests for long-range dependent data. Elect J Stat 11:2168–2198
Döring M (2010) Multiple change-point estimation with **U**-statistics. J Stat Plann Inference 104(7):2003–2017
Döring M (2011) Convergence in distribution of multiple change point estimators. J Stat Plann Inference 141(7):2238–2248
Enikeeva F, Munk A, Werner F (2018) Bump detection in heterogeneous gaussian regression. Bernoulli 42(2):1266–1306
Fotopoulos SB, Jandhyala VK, Tan L (2009) Asymptotic study of the change-point MLE in multivariate gaussian families under contiguous alternatives. J Statist Plann Inference 139(3):1190–1202
Francq C, Zakoïan JM (2012) Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. Econometrica 80(2):821–861
Gardner LA (1969) On detecting changes in the mean of normal variates. Ann Math Stat 116–126
Gombay E (2008) Change detection in autoregressive time series. J Multvar Anal 99(3):451–464
Gombay E, Serban D (2009) Monitoring parameter change in $ar(p)$ time series models. J Multvar Anal 100(4):715–725
Haccou P, Meelis E, van de Geer S (1988) The likelihood ratio test for a change point in a sequence of independent exponentially distributed random variables. Stoch Process Appl 30:121–139
Härdle W, Tsybakov A (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. J Econometrics 81(1):223–242
Härdle W, Tsybakov A, Yang L (1998) Local polynomial estimators of the volatility function in nonparametric autoregression. J Stat Plann Inference 68(2):221–245
Harel M, Puri ML (1989) Limiting behavior of **U**-statistics, **V**-statistics and one-sample rank order statistics for nonstationary absolutely regular processes. J Multvar Anal 30:180–204
Harel M, Puri M (1994) Law of the iterated logarithm for perturbed empirical distribution functions evaluated at a random point for nonstationary random variables. J Theor Probab 4:831–855
Hlávka Z, Hušková M, Meintanis S (2020) Change-point methods for multivariate time-series: paired vectorial observations. Stat Pap. 61:1351–1383
Horváth L, Hušková M (2005) Testing for changes using permutations of **U**-statistics. J Stat Plann Infer 128:351–371
Huh J (2010) Detection of a change point based on local-likelihood. Statistics 101:1–17
Imhof JP (1961) Computing the distribution of quadratic forms in normal variables. Biometrika 48:419–426

Kander Z, Zacks S (1966) Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. Ann Math Stat 37:1196–1210

Kengne WC (2012) Testing for parameter constancy in general causal time-series models. J Time Ser Anal 33(3):503–518

Khakhubia TG (1987) A limit theorem for a maximum likelihood estimate of the disorder time. Theor Probab Appl. 31:141–144

Ma L, Grant J, Sofronov G (2020) Multiple change point detection and validation in autoregressive time series data. Stat Pap 61:1507–1528

MacNeill I (1974) Tests for change of parameter at unknown times and distributions of some related functionals on brownian motion. Ann Stat 31(2):950–962

Matthews DA, Farewell VT, Pyke R (1985) Asymptotic score-statistic processes and tests for constant hazard against a changepoint alternative. Ann Stat 31(13):583–591

Meintanis SG (2016) A review of testing procedures based on the empirical characteristic function. S Afr Stat J 50:1–14

Mohr M, Selk L (2020) Estimating change points in nonparametric time series regression models. Stat Pap 61:1437–1463

Ngatchou-Wandji J (2009) Testing for symmetry in multivariate distributions. Stat Methodol 6:230–250

Oodaira H, Yoshihara K (1972) Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition. Kodai Math Semin Rep 24:259–269

Page ES (1954) Continuous inspection schemes. Biometrika 41:100–115

Page ES (1955) A test for a change in a parameter occurring at an unknown point. Biometrika 42:523–526

Pettitt AN (1979) A non-parametric approach to the change-point problem. Appl Stat 28:126–135

Phillips P, Durlauf S (1986) Multiple time series regression with integrated processes. Rev Econom Stud 53(4):473–495

Pycke J (2001) Une généralisation du développement de **K**arhunen-**L**oève du pont brownien(french) [a generalization of the **K**arhunen-**L**oève expansion of the brownian bridge]. C R Acad Sci Ser I 333(7):685–688

Rackauskas A, Wendler M (2020) Convergence of u-processes in hölder spaces with application to robust detection of a changed segment. Stat Pap 61:1409–1435

Riesz F, Nagy B (1972) Leçons d'analyse fonctionnelle, 6th edn. Gauthier-Villars, Paris

Sen A, Srivastava MS (1975) On tests for detecting changes in mean. Ann Stat 3:98–108

Shorack G, Wellner J (1986) Empirical processes with applications to statistics. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley, New York

Wang Q, Phillips PC (2012) A specification test for nonlinear nonstationary models. Ann Stat 40:727–758

Wolfe DA, Schechtman E (1984) Nonparametric statistical procedures for the changepoint problem. J Stat Plann Inference 9:389–396

Yang Y, Song Q (2014) Jump detection in time series nonparametric regression models: a polynomial spline approach. Ann Inst Stat Math 66:325–344

Yang Q, Li Y-N, Zang Y (2020) Change point detection for nonparametric regression under strongly mixing process. Stat Pap 61:1465–1506

Yao YC, Davis RA (1986) The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. Sankhya 48:339–353

Yoshihara Y (1976) Limiting behavior of **U**-statistics for stationary absolutely regular processes. Z Wahrscheinlichkeitstheorie Verw Gebiete 35:237–252

Zhou Z (2014) Nonparametric specification for non-stationary time series regression. Bernoulli 20(1):78–108