



**HAL**  
open science

# Testing the equality of the laws of two strictly stationary processes

D. Pommeret, L. Reboul, L. Yao

► **To cite this version:**

D. Pommeret, L. Reboul, L. Yao. Testing the equality of the laws of two strictly stationary processes. 2020. hal-02490333

**HAL Id: hal-02490333**

**<https://hal.science/hal-02490333v1>**

Preprint submitted on 25 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Testing the equality of the laws of two strictly stationary processes

D. Pommeret<sup>a,b</sup>, L. Reboul<sup>b</sup>, L. Yao<sup>c</sup>

<sup>a</sup> *Univ Lyon 1, Lab. SAF EA2429, ISFA, France*

<sup>b</sup> *Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France*

<sup>c</sup> *Clermont Auvergne Univ., LMBP - UMR 6620, CNRS, France*

---

## Abstract

In this paper we propose to generalize the recent works of [4] and [3] on the comparison of the marginal distributions of two strictly stationary processes. Our aim is to test the equality the whole distributions of two such processes. For that task, we compare all possible  $d$  dimensional joint distributions of both processes. Our procedure consist in expanding their densities in a multivariate orthogonal basis and comparing their  $k$  first coefficients. The number  $d$  of dimensions to consider and the number  $k$  of coefficients to compare in view to perform the test can growth with the sample size and are automatically selected by a two step data driven procedure. The method works for possibly paired, short or long range dependent processes. A simulation study shows the good behavior of the test procedure. In particular we apply our method to compare ARFIMA processes. Real data sets also illustrate this approach.

*Keywords:* Goodness of fit test, Orthogonal polynomials Stationary process, Smooth test, Short memory, Smooth test, Two step data driven procedure

---

## 1. Introduction

There exists a recent literature in statistics concerned by the comparison of two stationary processes with short or long range dependence. Usually the null hypothesis is restricted to the equality of the processes margins as in [4] and [3]. Comparison of two joint distributions has been considered in [10] or [2]. But these authors consider fixed dimension of the joint distribution and independent samples. Then the criticism about all these approaches are essentially that the full law of the processes are not compared, but only a marginal or a fixed order finite-dimensional distribution. In this paper, we consider the more general problem of testing whether two strictly stationary processes have the same distribution. More precisely, let  $(X_t)_{t \geq 0}$  and  $(Y_t)_{t \geq 0}$  be two strictly stationary

---

*Email addresses:* `denys.pommeret@univ-amu.fr` (D. Pommeret),  
`laurence.reboul@univ-amu.fr` (L. Reboul), `Anne-francoise.Yao@math.univ-bpclermont.fr`  
(L. Yao)

processes taking values in a real space  $E$ . Note that  $E$  can be continuous as well as discrete. The aim of this paper is to test the null hypothesis:

$$H_0 : \mathcal{L}(X) = \mathcal{L}(Y),$$

where  $\mathcal{L}$  denotes the law of the process, wholly characterized by its finite-dimensional distributions. This testing problem is of interest in finance when we want to compare different time series. For instance [4] compared three financial index time series: the Dow Jones Composite Average, the NASDAQ Composite and the NYSE International 100 Index. But these authors only considered the restricted null hypothesis  $f_X = f_Y$ , where  $f_X$  and  $f_Y$  stands for the marginal densities of the stationary processes. We propose here to extend the previous works by considering the general null hypothesis  $H_0$ . Since the processes are stationary, the hypothesis  $H_0$  may be rewritten as

$$H_0 : \mathcal{L}(X_1, \dots, X_d) = \mathcal{L}(Y_1, \dots, Y_d), \quad \text{for all integer } d > 0.$$

Hereafter, we propose a smooth test allowing to compare the coefficients of suitable expansions of the  $d$ -dimensional density functions of  $X$  and  $Y$ . For that task, we use a sequence of test statistics indexed by  $d$  and for each we associate a  $k(d)$ -th order expansion of the density in the  $d$ -dimensional basis. Our approach consists in selecting an optimal couple  $(d, k(d))$  by a two-step data-driven penalization procedure. The dimension  $d$  and the number of coefficients  $k(d)$  can growth with the sample size.

Under the null, we show that this rule selects simultaneously the first dimension  $d = 1$ , and the first component  $k(1) = 1$ . Therefore, the null distribution of the resulting test statistic coincides with that of a comparison of margins, as in [4] or [3], and we can use their previous results to obtain the limit distribution of the test statistic in different contexts of Short Range Dependence (SRD) and Long Range Dependence (LRD) of the processes. The detection of alternatives is more refined, because there are two directions to be detected:

- the dimension  $d$ , meaning that the  $d$ -dimensional joint distributions of both processes may differ.
- the order  $k(d)$  of the coefficients to take into account in the expansions, meaning that the difference between the two processes can be related to the  $k(d)$ -th coefficients of the expansions of their  $d$ -dimensional joint densities.

Under the alternative the test procedure selects such directions and the test statistic is asymptotically divergent. A simulation study on SRD and LRD processes is displayed in order to investigate the finite sample performances of the test. We studied in particular Autoregressive Fractionally Integrated Moving Average (ARFIMA) models. Moreover, our test is applied on the data set of financial assets of the US economic sector considered in [4], allowing to complete their study. A second illustration of the behavior our test is displayed on a data set from Kepler campaigns, obtained from the Center for Astrophysics

Harvard & Smithsonian (CfA), where light curves are compared to determinate which galaxies companion of the Milky Way could have significant different flux.

The rest of the paper is organized as follows: in Section 2 we present the general methodology of the test. Section 3 describes the two step data driven procedure and studies the asymptotic distribution of our test statistic under the null. Section 4 shows the convergence of the test under alternatives. Section 5 contains simulation results for short and long range dependent processes and Section 6 deals with the real data set. A discussion closes the paper in Section 7. In Appendix 1 we recall previous results on short and long memory processes and we show how our test procedure can be adapted in this context. All proofs are relegated in Appendix 2.

## 2. Description of the method

Let  $X$  and  $Y$  be two possibly correlated strictly stationary discrete time processes defined on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  and taking values on a real space  $E$ . We wish to test

$$H_0 : \mathcal{L}(X) = \mathcal{L}(Y) \quad \text{again} \quad H_1 : \mathcal{L}(X) \neq \mathcal{L}(Y) \quad (1)$$

based on sample paths  $(X_1, \dots, X_{n_X})$  and  $(Y_1, \dots, Y_{n_Y})$  of  $X$  and  $Y$ . For the sake of simplicity, we assume throughout that  $n_X = n_Y = n$  (the general case may be handled similarly as soon as we assume that  $n_X/n_Y \rightarrow a \neq 0$ ). To compare  $\mathcal{L}(X)$  and  $\mathcal{L}(Y)$  we propose to compare their  $d$  dimensional joint distributions. For that task, let us consider for all integer  $d > 0$ , a reference probability measure  $\mu_d$  on  $E^{\otimes d}$  such that there exists  $\mathcal{Q}_d = \{Q_{k,d}; k \in \mathbb{N}^d\}$ , a dense  $\mu_d$ -orthogonal basis of  $d$ -dimensional functions of  $L^2(\mu_d)$ , the space of square integrable functions with respect to  $\mu_d$ , satisfying  $Q_{0,d}(x) = 1$  and

$$\int_{E^d} Q_{j,d}(x) Q_{k,d}(x) \mu_d(dx) = \delta_{jk},$$

where  $\delta_{jk} = 1$  if  $j = k$  and 0 otherwise (see Remark 1 for an example of construction of  $\mu_d$  and  $\mathcal{Q}_d$ ).

We denote by  $f_{X,d}$  and  $f_{Y,d}$  the joint densities of  $(X_1, \dots, X_d)$  and  $(Y_1, \dots, Y_d)$  with respect to the reference measure  $\mu_d$ . We assume that for all  $d \in \mathbb{N}^*$ ,  $f_{X,d}$  and  $f_{Y,d}$  belong to  $L^2(\mu_d)$  so that for all  $x \in E^{\otimes d} = E \otimes \dots \otimes E$ , we can write

$$f_{X,d}(x) = \sum_{k \in \mathbb{N}^d} a(k,d) Q_{k,d}(x) \quad \text{and} \quad f_{Y,d}(x) = \sum_{k \in \mathbb{N}^d} b(k,d) Q_{k,d}(x),$$

with

$$a(k,d) = \int_{\mathcal{X}^d} Q_{k,d}(x) f_{X,d}(x) \mu_d(dx) = \mathbb{E}(Q_{k,d}(X_1, \dots, X_d)),$$

$$b(k,d) = \int_{\mathcal{X}^d} Q_{k,d}(x) f_{Y,d}(x) \mu_d(dx) = \mathbb{E}(Q_{k,d}(X_1, \dots, X_d)).$$

Notice that since  $Q_{0,d}(x) = 1$ , we have  $a_{0,d} = b_{0,d} = 1$  for all  $d \in \mathbb{N}^*$ . Therefore, in this setting,  $H_0$  can be rewritten as

$$H_0 : a(k, d) = b(k, d), \quad \text{for all } d \in \mathbb{N}^* \text{ and for all } k \in \mathbb{N}^{*d}. \quad (2)$$

In order to construct a suitable test statistic for (2), we can compare the empirical estimates of  $a(k, d)$  and  $b(k, d)$ . They are defined for  $k > 0$  and  $d < n$  by

$$a_n(k, d) := \frac{1}{u_n} \sum_{i=1}^{n-d+1} Q_{k,d}(X_{i,d}), \quad (3)$$

$$b_n(k, d) := \frac{1}{u_n} \sum_{i=1}^{n-d+1} Q_{k,d}(Y_{i,d}), \quad (4)$$

where  $X_{i,d} = (X_i, \dots, X_{i+d-1})$  (res.  $Y_{i,d} = (Y_i, \dots, Y_{i+d-1})$ ) is a  $d$ -dimensional vector with joint distribution  $f_{X,d}$  (resp.  $f_{Y,d}$ ) and  $u_n \rightarrow \infty$  is a conveniently chosen sequence, depending on the nature of the dependence range of  $X$  and  $Y$  (SRD or LRD) to obtain their consistency. So, our test procedure will be based on the sequences of vectors of differences

$$r_n(k, d) := a_n(k, d) - b_n(k, d) = \frac{1}{u_n} \sum_{i=1}^{n-d+1} V_{k,d}^{(i)}$$

for  $d < n$ , with

$$V_{k,d}^{(i)} = (Q_{k,d}(X_{i,d}) - Q_{k,d}(Y_{i,d})). \quad (5)$$

In order to select automatically the number of polynomial coefficients we introduce the notion of *total order* defined for any vector  $j = (j_1, \dots, j_d)$  by

$$|j| = j_1 + \dots + j_d.$$

Then, we define the sequence of test statistics

$$T_n(k, d) = \sum_{j \in \mathbb{N}^d; 0 < |j| \leq |k|} \tilde{r}_n(j, d)^2 := \sum_{j \in \mathbb{N}^d; 0 < |j| \leq k} \frac{r_n(j, d)^2}{\max(\hat{\sigma}_n^2, e_n)}, \quad (6)$$

for  $d < n$ . Here  $\hat{\sigma}_n^2$  is a convergent estimator of the asymptotic variance of  $r_n(1, 1)$ . It will be seen in Section 7 that such a normalization by the variance allows to get a parameter free distribution for the statistic. The sequence  $e_n$  is a sequence of trimming terms satisfying  $e_n \rightarrow 0$ , allowing to avoid instability of the test statistic.

Notice that  $T_n(k, d)$  is a proper statistic for testing that  $a(j, d) = b(j, d)$ ,  $j \in \mathbb{N}^{*d}$ ,  $|j| = 1, \dots, k$ . More precisely, each component  $\tilde{r}_n(j, d)$  is used to compare the coefficients  $a(j, d)$  and  $b(j, d)$  in the expansion of the  $d$ -dimensional distributions

in order to detect the alternative. Thus we need to let  $|k|$  tend to infinity to detect all possible alternatives. However, choosing too large parameters tends to power dilution of the test. Hereafter, we propose to construct a penalized selection of these parameters by a two step data driven procedure.

**Remark 1.** A simple construction of  $\mu_d$  is to fix  $\mu_d = \mu \otimes \dots \otimes \mu$  ( $d$  times), where  $\mu$  is a probability measure on  $E$  with associated orthonormal basis  $\{Q_k; k \in \mathbb{N}\}$ . Their existence is related to moments conditions (see for instance [1]) and is satisfied in particular if there exists  $a > 0$  such that  $\int e^{a|x|} \mu(dx) < \infty$ . An associated  $\mu_d$ -orthogonal basis is then given by  $\mathcal{Q}_d = \{Q_{k,d}; k = (k_1, \dots, k_d) \in \mathbb{N}^d\}$ , with  $Q_{k,d}(x_1, \dots, x_d) = Q_{k_1}(x_1) \otimes \dots \otimes Q_{k_d}(x_d)$ .

### 3. A two step data driven procedure

#### 3.1. Selecting $k$ for fixed dimension $d$

Following [9] and [8], we suggest a first data driven procedure to select automatically the number of coefficients when the dimension  $d$  is fixed. For that task, we introduce a penalized rule allowing to pick the best total order  $|k|$  to test the hypothesis  $f_{X,d} = f_{Y,d}$ . Namely, we set

$$k_n(d) := \min \left\{ \operatorname{argmax}_{1 \leq |k| \leq K_n} (T_n(k, d) - [k]_d \log(n)) \right\}, \quad (7)$$

where

$$[k]_d = \sum_{j=1}^k \frac{(j+d-1)!}{j!(d-1)!}$$

and  $K_n \rightarrow +\infty$  as  $n \rightarrow +\infty$ . Clearly  $[k]_d$  denotes the number of coefficients with total order lower or equal to  $|k|$ , or in other words, the number of terms in  $T_n(k, d)$ . For instance, for  $|k| = 1$  and  $d = 1$  we must have  $k = 1$ . But if  $d = 2$  we have two possibilities for  $|k| = 1$ :  $k = (1, 0)$  and  $k = (0, 1)$ .

Then, setting

$$\sigma_n(k, d) = \sum_{i=0}^{n-1} |\mathbb{E}_0 (V_{k,d}^{(0)} V_{k,d}^{(i)})|,$$

the partial sums of the absolute series of the auto-covariances of the process defined in (5) and assuming

**(A1)**  $K_n = o(\log(n)e_n)$ .

**(A2)** There exists  $M > 0$  such that for all  $n \in \mathbb{N}$  and for all integers  $d < n$ ,

$$\frac{1}{K_n} \sum_{k \in \mathbb{N}^d, |k| \leq K_n} \sigma_n(k, d) \frac{(n-d)}{u_n^2} < M,$$

we have

**Theorem 1.** *If (A1-A2) hold, then, under  $H_0$ , for all integer  $0 < d < n$ ,*

$$k_n(d) \xrightarrow{P} 1$$

as  $n \rightarrow +\infty$ .

Theorem 1 says that, whatever the dimension of the joint distribution, under the null the penalized coefficient retained is always asymptotically the first.

**Remark 2.** It has been mentioned earlier and it will be made clearer in Section 7 that  $u_n$  is the rate at which a (non) central limit theorem holds for the partial sums of the processes  $V_{k,d}$ . In cases where such processes are SRD (or i.i.d.),  $u_n = \sqrt{n}$  and the absolute series of the auto-covariances converges so that  $\sigma_n(k, d)$  is bounded ( $\sigma_n(k, d) = \mathbb{E}_0(V_{k,d}^{(0)})^2$  in the i.i.d. case) and Assumption (A2) is automatically satisfied. When the  $V_{k,d}$ 's are LRD, a non central limit theorems holds in some cases (see Section 7) at rates  $u_n$  such that  $\sqrt{n}/u_n \rightarrow 0$ , while  $\sigma_n(k, d)$  diverge. In such cases, (A2) will be satisfied as soon as  $\sigma_n(k, d)$  do not diverge too fast.

### 3.2. Selecting the dimension $d$

We use a second data driven procedure to select the best dimension  $d$  among the  $d$ -dimensional distributions to detect alternative. From (3) and (4) we must have  $d < n$ . Then for any integer  $\delta < n$  we define

$$d_n(\delta) := \min_{1 \leq d \leq \delta} \{ \operatorname{argmax}(T_n(k_n(d), d) - [k_n(d)]_d \log(n)) \}. \quad (8)$$

Thus, one has

**Theorem 2.** *If (A1-A2) hold, then, under  $H_0$ , for any integer  $\delta$ ,*

$$d_n(\delta) \xrightarrow{P} 1$$

as  $n \rightarrow +\infty$ .

Theorem 2 says that under the null the penalized dimension retained is always asymptotically the first.

### 3.3. A data-driven test

Let us define for our testing problem the test statistic for any integer  $\delta < n$ ,

$$\tilde{T}_n(\delta) = T_n(k_n(d_n(\delta)), d_n(\delta)),$$

where  $d_n$  and  $k_n$  are defined by (7) and (8). In the following section, we will study the test based on its asymptotic distribution under the null. With this in mind, we give below a preliminary result. Let us assume

(B) There exists  $\sigma > 0$  such that under  $H_0$ ,  $r_n(1, 1) \xrightarrow{\mathcal{L}} U$ , where  $U$  follows a  $\mathcal{N}(0, \sigma^2)$ .

It is important to note that  $T_n(1, 1) = \tilde{r}_n^2(1, 1) = r_n^2(1, 1) / \max(\hat{\sigma}_n^2, e_n)$ , where  $e_n \rightarrow 0$  and where  $\hat{\sigma}_n^2$  is a consistent estimator of the asymptotic variance of  $r_n(1, 1)$ . Thus, combining Theorems 1 and 2 we have

**Theorem 3.** *If (A1–A2) and (B) hold, then, under  $H_0$ , for any integer  $\delta$ ,*

$$\tilde{T}_n(\delta) \xrightarrow{\mathcal{L}} Z$$

as  $n \rightarrow +\infty$ , where  $Z$  follows a chi-squared distribution with one degree of freedom.

Thus, in order to test (1) at a nominal level  $\alpha$ , we can use  $\tilde{T}_n(\delta)$  and the  $(1 - \alpha)^{th}$ -percentile of the chi-squared distribution with one degree of freedom.

Theorem 3 shows that under the null, the test statistic selects simultaneously the first dimension, that is  $d = 1$ , and the first order, that is  $k_n(1) \rightarrow 1$ , asymptotically. Finally, under the null, we only have to consider the distribution of the first statistic related to the margins of  $X$  and  $Y$ . Then we can adapt Doukhan et al (2015, 2019)'s to obtain the limit distribution of  $\tilde{T}_n(\delta)$  in the various situations of dependence they consider, including SRD and LRD cases. These adaptations are described in Appendix 1 where we focus on  $\alpha$ -mixing processes,  $\theta$  dependent processes, function of Gaussian processes, and linear processes.

**Remark 3.** Assumption (B) amounts to say that there exists a (non) central limit theorem for the partial sums of the process  $V_{1,1}$ . Such a result essentially rely on the dependence range of this process, itself depending on that of  $X$  and  $Y$  and on the nature of  $\mathcal{Q}_{1,1}$  as it will be detailed in the next sections.

#### 4. Behavior under alternatives

Finally, we study the behaviour of the test procedure under alternatives. First recall that  $H_0$  can be rewritten as  $a(k, d) = b(k, d)$  for all  $d \in \mathbb{N}^*$  and for all  $k \in \mathbb{N}^{*d}$ . For any positive integers  $k_1$  and  $d_1$  we consider alternatives of the form

$$\begin{aligned} H_1(k_1, d_1) : \quad & a(k, d) = b(k, d), \quad \forall d < d_1 \text{ and } \forall k \in \mathbb{N}^{*d} \\ & a(k, d_1) = b(k, d_1), \quad \forall |k| < |k_1| \text{ with } k \in \mathbb{N}^{*d_1} \\ & a(k_1, d_1) \neq b(k_1, d_1). \end{aligned}$$

If  $d_1 = 1$ ,  $H_1(k_1, d_1)$  means that  $X$  and  $Y$  do not have the same marginal density and this difference appears in their  $k_1$ -th order expansion coefficients. If  $d_1 > 1$ ,  $H_1(k_1, d_1)$  means that for all  $d < d_1$  the first  $d$  joint distributions of  $X$  and  $Y$  coincide, but their  $d_1$  dimensional margins differ, and this difference can be detected from their  $k_1$ -th order expansion coefficients.

We assume here that the multivariate bases are embedded, that is,

$$\mathcal{Q}_d = \{Q_{k,d}, k \in \mathbb{N}^d\} \subset \mathcal{Q}_{d_1} = \{Q_{k,d_1}, k \in \mathbb{N}_{1}^{d_1}\} \quad \forall d < d_1. \quad (9)$$



This condition is satisfied as soon as we use the formal construction evoked in Remark 1, which is very natural and which is used in our numerical studies. Combining (9) and the construction given in (6), we can see that for  $d > 1$ :

$$T_n(k, d) = T_n(k, d-1) + \sum_{j \in \mathbb{N}_*^d; 0 < |j| \leq |k|} \tilde{r}_n(j, d)^2,$$

where  $\mathbb{N}_* = \{1, 2, \dots\}$ .

We need the following assumption that we check for LRD processes in Appendix 3.

- (C) Write  $\delta_1 = a(k_1, d_1) - b(k_1, d_1)$ . There exists  $\sigma > 0$  such that under  $H_1(k_1, d_1)$ ,  $\tilde{r}_n(k_1, d_1) - \frac{n}{u_n} \delta_1$  converges as  $n$  tends to infinity to a random variable  $Z$  such that  $\mathbb{P}(Z < +\infty) = 1$ .

**Proposition 1.** *Assume that (C) holds. Under  $H_1(k_1, d_1)$ , for all  $\delta > 0$ ,  $\tilde{T}_n(\delta) \rightarrow +\infty$  when  $n$  tends to infinity.*

## 5. Simulation results

Hereafter, we run Monte-Carlo simulations with 1000 replications for two types of memory processes:  $\alpha$ -mixing processes and LRD Autoregressive Fractionally Integrated Moving Average (ARFIMA) processes. The nominal level is fixed at  $\alpha = 5\%$ . We present below the models and tests as well as the obtained results. According to the support  $\mathbb{R}$  of the processes considered in our simulation study, we used here the standard Gaussian distribution and its associated Hermite polynomials. More precisely, we consider  $\mu_d = \mu \otimes \dots \otimes \mu$  ( $d$  times), where  $\mu$  is the standard Gaussian measure on  $\mathbb{R}$ , with  $Q_{k,d}(x_1, \dots, x_d) = H_{k_1}(x_1) \otimes \dots \otimes H_{k_d}(x_d)$ , where  $H_k$  is the  $k$ th normalized Hermite polynomial.

We fixed  $e_n = 1/(10n)$ ,  $K_n = 6$  and  $\delta = 4$ . The choice  $e_n = 1/(10n)$  permits to avoid too small values for the variance estimator in (6). The choice of  $K_n$  is dictated by the previous numerical studies done in [3], in which it has been observed that coefficients of order greater than 6 are rarely chosen under alternative hypotheses and this choice does not modify the level of the test. The choice  $\delta = 4$  means that we are not able to detect the difference in law of processes having equal common joint density up to dimension 4, but with a difference for a greater dimension. However,  $\delta$  has to be small enough to have an efficient and faster program. To give an order of idea, the number  $[j]_d$  gives 209 coefficients to be evaluated for  $j = 6$  and  $d = 4$ . For  $d = 5$  we obtain  $[j]_d = 461$  which leads to an important number of estimations. In our simulation study, we only study departures from the null up to the dimension 4.

### 5.1. $\alpha$ -mixing processes

*Model.* The simulated examples are based on the observation of sequences of size  $n \in \{100, 200, 500, 1000\}$  of the following  $\theta$  dependent processes  $Z = (X, Y)$ :

$$Z_t = C + \Theta Z_{t-1} + \epsilon_t, t \in \mathbb{Z}, \quad (10)$$

where  $C = (0, c)'$ ,  $\Theta$  is a diagonal matrix with main diagonal vector  $\theta = (\theta_X, \theta_Y)'$  such that  $|\theta_X| < 1$  and  $|\theta_Y| < 1$  and  $(\epsilon_t)_{t \in \mathbb{Z}}$  is a bivariate white noise with mean zero and auto-covariance structure  $\mathbb{E}(\epsilon_t \epsilon_{t+h}') = \Sigma - \Theta \Sigma \Theta'$  for  $h = 0$  and zero otherwise, where  $\Sigma$  is a symmetric square matrix of order 2 with main diagonal vector  $(1, \sigma_Y^2)'$  and cross term  $v$ .

Therefore,  $Z_t$  has a mean vector  $(0, \mu)$ , with  $\mu = c/(1 - \theta_Y)'$  and an auto-covariance matrix  $\mathbb{E}(Z_t Z_{t+h}') = \Theta^h \Sigma$ .

The underlying process  $Z$  is  $\alpha$ -mixing, defined as a stationary bivariate vector autoregressive process of order one,  $X$  has a standard Gaussian marginal distribution while  $Y$  has a Gaussian marginal distribution with mean  $\mu$  and variance  $\sigma_Y^2$ . Then:

- The null hypothesis corresponds to  $\mu = 0$ ,  $\sigma_Y^2 = 1$  and  $\theta_X = \theta_Y$ .
- The null distribution of our test statistic  $\tilde{T}_n(\delta)$  is given by Theorem 2, with  $u_n = \sqrt{n}$  and  $\tilde{\sigma}_n^2$  the Kernel HAC estimator defined as in [4].

This model allows us to investigate several degrees of between and within sample dependence:

- within sample dependence is controlled by  $\theta_X$  and  $\theta_Y$  and increases with their absolute values. The case  $(\theta_X, \theta_Y) = (0, 0)$  corresponds to the within sample independence while other values of the pair  $\theta$  correspond to either positive or negative within dependence.
- Dependence between samples is controlled by the cross-term  $v$  of  $\Sigma$  and increases with its absolute value. We chose  $v \in \{0, 0.5, 0.9\}$ . The case  $v = 0$  corresponds to independent sequences.
- The null hypothesis is denoted by  $H_0(\theta, v)$ . It corresponds to the case where  $\mu = 0$ ,  $\sigma_Y^2 = 1$  and  $\theta_X = \theta_Y = \theta$ .
- Alternative hypotheses related to a change of mean (mean deviation) or variance (variance deviation) impact directly the marginal distribution. Such alternatives have been studied in Doukhan et al. (2015) and are well detected. Here we focus our study on more refined alternative hypothesis, when  $\mu = 0$ ,  $\sigma_Y^2 = 1$  and  $\theta_Y \neq \theta_X$ . In that case both marginal distributions of  $X$  and  $Y$  are equal. But the two processes are different and their joint distribution are not the same. We write  $H_1(\theta_X, \theta_Y, v)$  this alternative.

*Empirical levels.* Figure 1 shows the empirical levels based on 1000 replications under  $H_0(\theta, v)$  for  $\theta = \theta_X = \theta_Y \in \{0, 0.5\}$ ,  $\sigma_Y^2 = 1$ ,  $\mu = 0$  and  $v \in \{0, 0.5\}$ . We observe a fast convergence to the theoretical level of 5%.

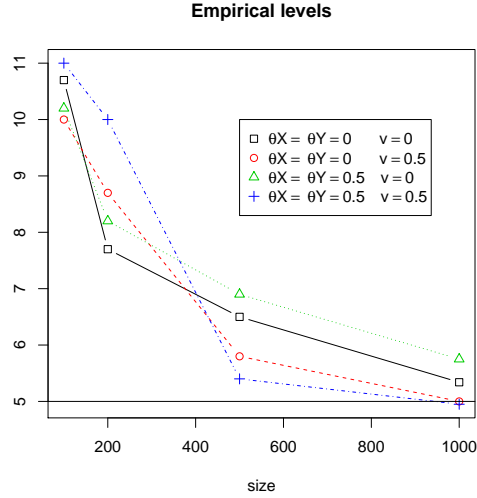


Figure 1: Empirical levels (in % of rejections) under  $H_0(\theta, v)$  for  $\theta = \theta_X = \theta_Y \in \{0, 0.5\}$ ,  $\sigma_Y^2 = 1$ ,  $\mu = 0$  and  $v \in \{0, 0.5\}$ .

*Empirical powers.* The case of independent processes is represented in Figure 2, where  $v = 0$  and  $\theta_X \neq \theta_Y$ . Here both marginal densities are equal, but the dependence and then joint densities differ. Then the test should detect a difference for  $d = 2$ , that is for the dimension where the joint densities differ. As expected, Figure 2 shows that the empirical power is greater for larger difference between the two memory parameters. In more than 90% of the rejected cases the choice of the dimension  $d$  is equal to 2, as expected. It means that the test detects differences between the bivariate densities of the two processes.

Figures 3-4 represent the case where  $X$  and  $Y$  are dependent processes, with  $v \neq 0$ . Both processes have the marginal distribution. In Figure 3,  $v = 0.5$  and in Figure 4,  $v = 0.9$ . We observe a larger empirical power than in the independent case. The power is greater for larger value of  $v$ , that is, for greater dependence. Again the larger the difference between  $\theta_X$  and  $\theta_Y$ , the greater the power is. The selected dimension in case of rejection is essentially 2, as expected, that is the test detects a difference between the bivariate densities.

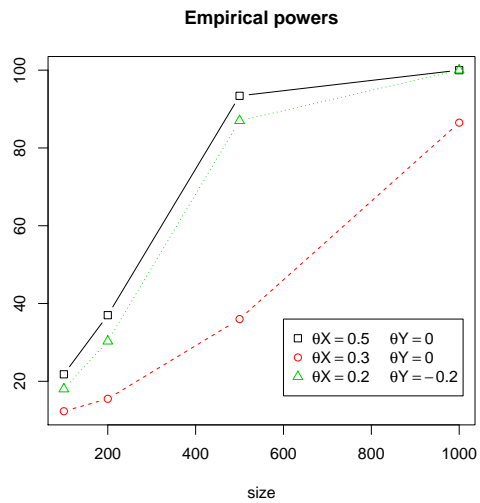


Figure 2: Empirical powers (in % of rejections) under  $H_1(\theta_X, \theta_Y, v)$  for  $\theta_X \neq \theta_Y$  and  $v = 0$ .

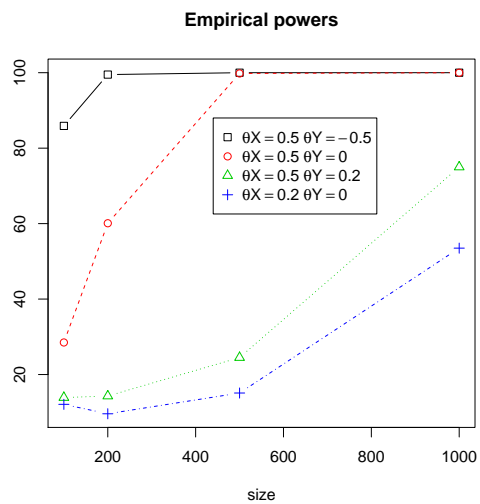


Figure 3: Empirical powers (in % of rejections) under  $H_1(\theta_X, \theta_Y, v)$  for  $\theta_X \neq \theta_Y$  and  $v = 0.5$ .

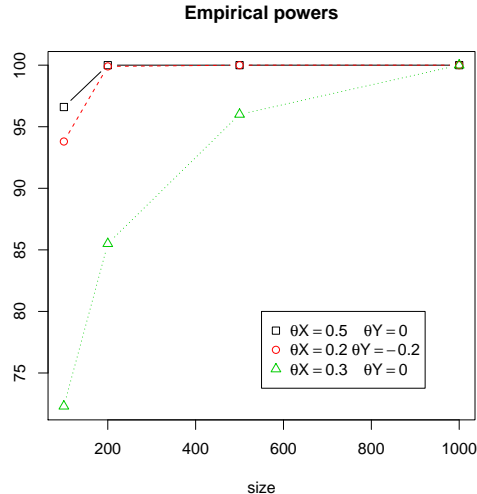


Figure 4: Empirical powers (in % of rejects) under  $H_1(\theta_X, \theta_Y, v)$  for  $\theta_X \neq \theta_Y$  and  $v = 0.9$

### 5.2. ARFIMA processes

In this section, we consider Autoregressive Fractionally Integrated Moving Average (ARFIMA) models, with different long memory parameters. The ARFIMA processes are one of the best-known classes of long-memory models used in finance and our method gives a way to compare such processes.

*Models.* The simulations consist of sequences with respective sizes  $n \in \{100, 200, 500, 1000\}$  of independent ARFIMA processes  $X$  and  $Y$  that we briefly describe here (see [5] and [7] for more details). Setting  $B^k x_t = x_{t-k}$ , recall that  $X$  is an ARFIMA( $p, \theta, q$ ) with mean  $\mu$  if

$$\Phi(B)(1 - B)^\theta(X_t - \mu) = \Theta(B)\epsilon_t, \quad (11)$$

with

$$\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p, \quad \Theta(B) = 1 + \Theta_1 B + \dots + \Theta_q B^q,$$

$$(1 - B)^\theta = \sum_{k=0}^{\infty} \frac{\Gamma(k - \theta)}{\Gamma(-\theta)\Gamma(k + 1)} B^k,$$

and the  $\epsilon_i$ 's are a white noise sequence with mean zero and variance  $\sigma_\epsilon^2$ . We fix  $\Phi = 0$  (resp.  $\Theta = 0$ ) if  $p = 0$  (resp.  $q = 0$ ). Parameters  $p$  and  $q$  model short term dependency effects. The memory parameter  $\theta$  is allowed to assume every real value. The restriction of  $\theta$  to integer values gives rise to classical ARIMA processes. For  $|\theta| \geq 1/2$ , this is a non-stationary process. For  $\theta \in (0, 1/2)$ , the process is long memory and for  $\theta \in (-1/2, 0)$  this is negative memory. For

$\theta = 0$ , this is a particular case of short-memory process, which corresponds to a stationary ARMA model.

Here in our design,  $X$  is a pure fractional model ARFIMA(0,  $\theta_X$ , 0) with  $\theta_X \in (0, 1/2)$ . This model has the moving average representation:

$$X_t = \mu + \sum_{k=0}^{\infty} \alpha_k B^k \epsilon_k, \text{ with } \alpha_k = \frac{\Gamma(k + \theta_X)}{\Gamma(\theta_X)\Gamma(k + 1)}.$$

When the innovations are Gaussian the marginal distribution of  $X$  is gaussian with mean  $\mu$  and variance

$$\sigma_X^2 = \sigma_\epsilon^2 \frac{\Gamma(1 - 2\theta)}{(\Gamma(1 - \theta))^2}. \quad (12)$$

The joint distribution is more complex, but clearly depends of the memory parameter. Hereafter, we denote by  $\mathcal{M}(p, \theta, \mu, \sigma_\epsilon^2)$  the ARFIMA( $p, \theta, 0$ ) with mean  $\mu$  and variance innovation  $\sigma_\epsilon^2$ . With this notation we choose  $X$  as  $\mathcal{M}(0, \theta_X, 0, 1)$ . Varying the parameters  $\theta$ ,  $\mu$  and  $\sigma_\epsilon^2$ , we investigate null and alternative distributions of the process  $Y$ . The null hypothesis is represented when the process  $Y$  is  $\mathcal{M}(0, \theta_Y, 0, 1)$ , with  $\theta_Y = \theta_X$ , that is both processes  $X$  and  $Y$  are the same. We consider various values  $\theta_X \in \{0.1, 0.2, 0.3, 0.4\}$ . We write  $H_0(\theta_X)$  such a null hypothesis. Classical alternatives such mean deviation or variance deviation, coincide with a difference of marginal distributions for  $X$  and  $Y$ . Such alternatives have been largely studied in Doukhan et al (2019) and are relatively well detected. We consider briefly the variance deviation case in our simulation, where  $X$  is  $\mathcal{M}(0, \theta_X, 0, 1)$  and  $Y$  is  $\mathcal{M}(0, \theta_Y, 0, \sigma_\epsilon^2)$  with  $\sigma_\epsilon^2 \neq 1$ .

Eventually, we focus on a more refined alternative, leading to the same marginal distributions for both processes, but with  $\theta_X \neq \theta_Y$ , that is a difference of their joint distributions. We construct the alternative as follows: we consider the process  $Y$  as  $\mathcal{M}(0, \theta_Y, 0, \sigma_\epsilon^2)$ , such that

$$\sigma_\epsilon^2 = \frac{\Gamma(1 - 2\theta_X)}{(\Gamma(1 - \theta_X))^2} \cdot \frac{\Gamma(1 - 2\theta_Y)}{(\Gamma(1 - \theta_Y))^2}. \quad (13)$$

In such a way  $X$  and  $Y$  have the same marginal distribution  $\mathcal{N}(0, \sigma^2)$  with

$$\sigma^2 = \frac{\Gamma(1 - 2\theta_X)}{(\Gamma(1 - \theta_X))^2},$$

from (12). But their joint distributions differ since  $\theta_X \neq \theta_Y$ . Then our test procedure will attend to detect this departure.

The ARFIMA sequences were computed using the fracdiff package R. We used a burn-in period of 10000. The null distribution of our test statistic  $\tilde{T}_n$  is given by Theorem 4, with  $u_n = n^{1/2+\theta_X}$  and  $\hat{\sigma}_n^2$  the Whittle estimator of the asymptotic variance of  $r_n(1, 1)$  as defined as in [4].

*Empirical levels.* Figure 5 show the empirical level for different values of  $\theta_X$ . We can observe an unstable level for small sample sizes, but the convergence to the nominal level seems correct for values of  $\theta_X$  not too close of 0 or 0.5. This unstable phenomena is more pronounced when the memory parameter was too close to 0, as  $\theta = 0.1$  (proximity to the short memory case), or too close to 0.5, as  $\theta = 0.4$  (proximity to the non stationary case).

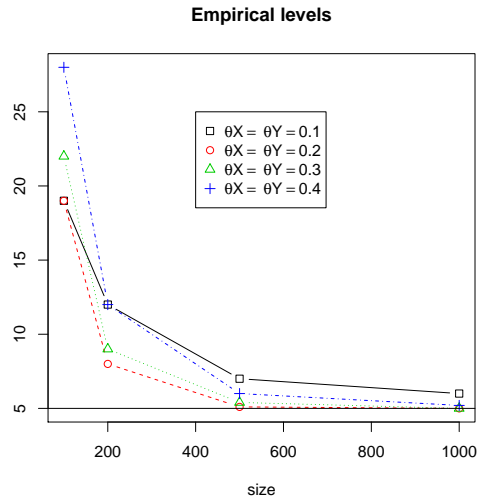


Figure 5: Empirical levels (in % of rejects)  $H_0(\theta_X)$  for  $\theta_X = \theta_Y \in \{0.1, 0.2, 0.3, 0.4\}$ ,  $\sigma_\epsilon^2 = 1$ ,  $\mu = 0$ .

*Empirical powers.* First we consider classical alternatives as memory parameter deviations, or variance deviations. We then consider the case where  $\theta_X \neq \theta_Y$ , but  $\mu = 0$  and  $\sigma_\epsilon^2 = 1$ , that is a change of the memory parameter. In that case, both marginal densities are centered Gaussian, since  $\mu = 0$ , with variance given by (13). We also consider the case where  $\theta_X = \theta_Y$ ,  $\mu = 0$ , but  $\sigma_\epsilon^2 \neq 1$ , that is again a change of the variance, but with the same dependence. In both cases the marginal densities are different and the test should reject the null hypothesis based on the selection of first dimension, that is,  $d = 1$ .

Second we consider the more refined change, with  $\theta_X \neq \theta_Y$  and with  $\sigma_\epsilon^2$  given by (13) with gives the same marginal distributions for both processes but with different joint densities.

Figure 6 shows empirical powers when memory parameters are different. In such a case the marginal densities are different and the test reject the null hypothesis selecting the first dimension and the second polynomial order indicating a difference of the second moments of the margins.

Figures 7-8 present the empirical powers when  $\theta_X = \theta_Y$ ,  $\mu = 0$ , and  $\sigma_\epsilon^2 \in \{0.5, 1.5\}$ . The power is greater when  $\theta_X$  is small. It can be explained by the rate of convergence  $u_n = n^{1/2+\theta_X}$ . Globally the power seems relatively good.

Figure 9 represents the power under the particular alternative from (12), when  $\theta_X \neq \theta_Y$  but with the same marginal distributions for both processes. Clearly the joint distributions differ because of the different range of dependence and we expect that the test procedure will detect such a change among the  $d$  dimensional joint densities. As seen in the figure, the powers are relatively good for large difference between  $\theta_X$  and  $\theta_Y$ . In that case, when the null hypothesis is rejected, the choice of the dimension is more than 90% the third, meaning that the difference is more significant for the three-dimensional joint distribution. We can observe that if  $\theta_X$  and  $\theta_Y$  are too close, the test cannot detect well the alternative, as illustrated when  $\theta_X = 0.3$  and  $\theta_Y = 0.2$ .

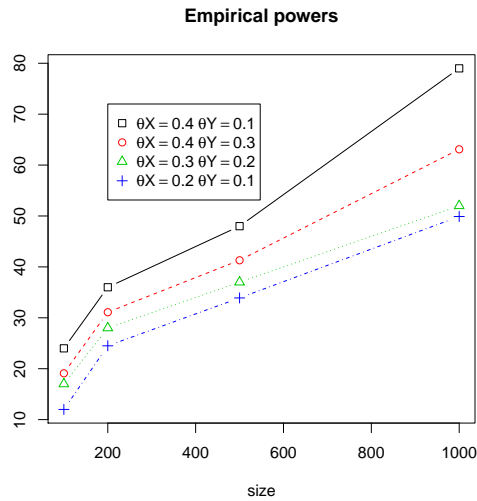


Figure 6: Empirical powers (in % of rejections) for  $\theta_X \neq \theta_Y$ ,  $\mu = 0$  and  $\sigma_\epsilon^2 = 1$ .



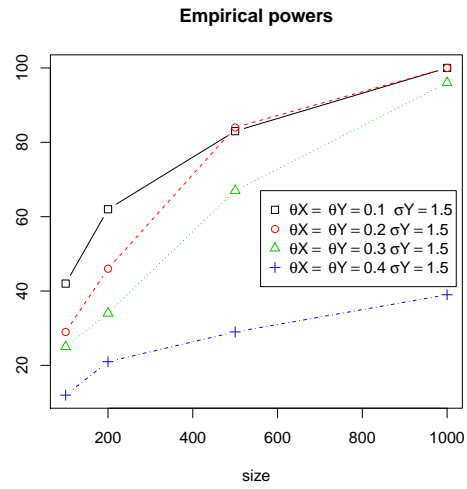


Figure 7: Empirical powers (in % of rejects) for  $\theta_X = \theta_Y$ ,  $\mu = 0$  and  $\sigma_\epsilon^2 = 1.5$ .

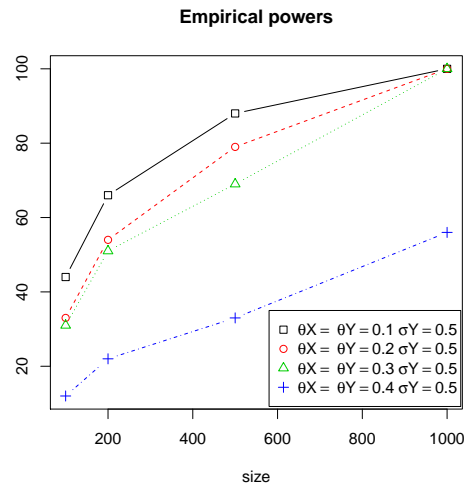


Figure 8: Empirical powers (in % of rejects) for  $\theta_X = \theta_Y$ ,  $\mu = 0$  and  $\sigma_\epsilon^2 = 0.5$ .

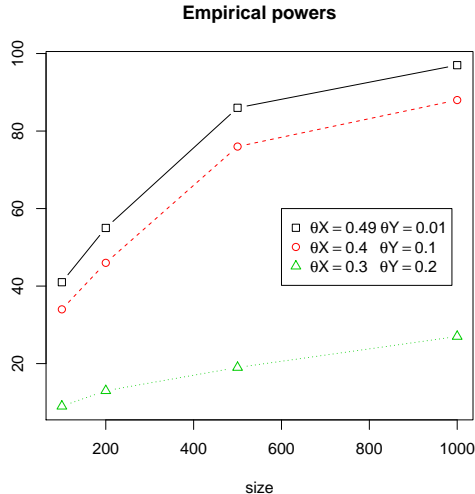


Figure 9: Empirical powers (in % of rejects) for  $\theta_X \neq \theta_Y$  but with same marginal.

## 6. Data study

### 6.1. NYSE versus Down Jones

We consider two financial assets of the US economic sectors studied in Doukhan et al. (2015). The data consist in monthly rates from January 2004 to August 2013. We consider two indices: the Dow Jones Composite Average and the NYSE International 100 Index. The first one is a stock index that tracks 65 prominent companies and the last one tracks the largest 100 non-U.S. common stocks listed on the New York Stock Exchange. To get stationary processes we consider the associated variations processes defined by  $X_{t+1} - X_t$ . We compare the equality of these two increments processes.

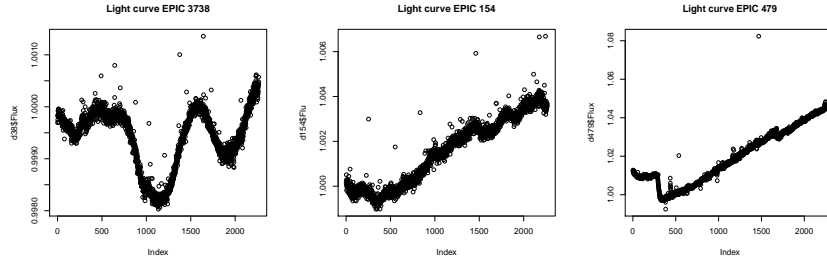
In [4] the test of equality of margins gave a p-value around 0.75, meaning that the margins of the increments of NYSE and Down Jones indexes from 2004 to 2013 can be considered as identically distributed.

We apply our test procedure to detect a possible difference in their joint distributions. We obtain a p-value equal to 0.26, selecting the first dimension and the first order coefficient, and we then do not reject the equality of the processes.

### 6.2. Light curves from Kepler mission

We consider light curves available via the Center for Astrophysics Harvard & Smithsonian (CfA). The data consist of observations of reduced light curves from regions of the Milky Way galaxy. There are 20 campaigns, each with hundreds of data recolted during the K2 mission. We give two illustrations of light curves comparison, with abbreviated names as follows: EPIC3738, EPIC154 and

EPIC479. Figures 6.2 represents the three original light curves. We compare the integrated processes derived from these data, that is we aim to compare the dynamic of such a processes.



Considering the derivative processes as stationary processes, we apply our testing procedure to compare the first two curves and the last two curves. We obtain

- EPIC3738 vs EPIC154: the retained dimension is  $d_n = 2$  and the retained order is  $k_n = 2$ , with a p-value  $< 10^{-6}$ . Then the test detects a significant difference between the bivariate joint distributions of the first two curves. Moreover, since  $k_n = 2$ , the rejection is due to the difference of their moments of second order, that is  $\mathbb{E}(Q_{k,2}(X_1 X_2))$  for  $|k| = 2$ .
- EPIC479 vs EPIC154: The retained dimension is  $d_n = 1$  and the retained order is  $k_n = 1$ , with a p-value = 0.08. The decision here depends on the threshold: we reject at 10% but not at 5%.

In conclusion, we reject clearly the equality of the dynamic of the first two curves, but there is not a strong significant difference between the two last ones.

## 7. Discussion

In this work we both theoretically and numerically addressed the comparison testing of two stationary processes, possibly dependent, and having short or long range memory. The method is automatic in the sense that the dimension of their joint distributions is chosen by a data driven technic. Simultaneously the order of departure is obtained by another data driven step. The dimension can grow with the sample size, but in practice, we can see that testing the first four or five dimensional joint distribution is reasonable. To the best of our knowledge this approach is the first permitting to detect a deviation into these two directions. It completes the previous works on the comparison of the marginal distributions in the sense that, if the equality between two marginal is not rejected, it permit to look for larger dimensional joint distributions. Simulation confirm the easy use of such a procedure and show the good behavior of the test for  $\alpha$ -mixing and ARFIMA processes. The study of real data sets confirms its practical importance: the decision of equality between two increments of

financial assets has been confirmed, and the comparison of series of curves can be easily implemented.

We think that this work could be extended in two main interesting ways. First we could consider the  $K$ -sample problem, as done for instance in [12] in the iid case. That could be adapted to compare  $K$  stationary processes simultaneously with techniques similar to those used in our paper. Coming back to financial assets or light curves, this would allow us to consider more than two indices or more than two regions simultaneously. Another interesting problem would be the use of such a method to detect a rupture in stationarity of processes, with many applications in domains where it is frequent to observe changes in distributions.

### Appendix 1: asymptotic null distribution under SRD and LRD models

Hereafter, we straightforwardly adapt Doukhan et al (2015, 2019)'s to obtain the limit distribution of  $\tilde{T}_n(\delta)$  in the various situations of dependence, namely  $\alpha$ -mixing processes,  $\theta$  dependent processes, function of Gaussian processes, and linear processes. From now we assume that condition **(A1)** is satisfied, that is we fix:

$$K_n = o(\log(n)e_n).$$

Assumption **(A2)** will be replaced case by case by specific conditions depending of the features of the processes.

#### *$\alpha$ -mixing processes*

Let  $Z = (X, Y)$  a bivariate strictly stationary  $\alpha$ -mixing process in the sense of in Rosenblatt (1956), with non-increasing mixing coefficients sequence  $(\alpha(m))_{m>0}$  with possibly dependent coordinates. We fix  $u_n = \sqrt{n}$  and  $\hat{\sigma}_n^2$  a consistent estimator of

$$\sigma_{SRD}^2 = \sum_{t=-\infty}^{+\infty} \mathbb{E}_0 \left( V_{1,1}^{(0)} V_{1,1}^{(t)} \right). \quad (14)$$

Assume that there exist  $C > 0$ ,  $n_0 > 0$  and  $\delta \geq 0$  such that

**(1.i)** for all  $n > n_0$ ,

$$\frac{1}{K_n} \sum_{k=1}^{K_n} \left( \mathbb{E}_0(|V_{k,d}^{(0)}|^{2+\delta}) \right)^{2/(2+\delta)} < C.$$

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

Then we have,

**Corollary 1.** *If (1.i) and (1.ii) hold, then **(A2)** and **(B)** hold, with  $\sigma^2 = \sigma_{SRD}^2$ , so that Theorem 3 obtains.*

*Proof.* The proof follows from Corollary 2 of Doukhan et al. (2015).

*$\theta$ -mixing processes*

Let  $Z = (X, Y)$  be a bivariate strictly stationary  $\theta$ -dependent process in the sense of Dedecker et al. (2007) with possibly dependent coordinate and a non-increasing coefficients sequence  $(\theta(m))_{m>0}$ . We fix  $u_n = \sqrt{n}$  and  $\hat{\sigma}_n^2$  a consistent estimator of  $\sigma_{SRD}^2$  defined by (14). Assume that there exist  $C > 0$ ,  $n_0 > 0$  and  $\delta \geq 0$  such that

(2.i) for all  $n > n_0$ ,

$$\frac{1}{K_n} \sum_{k=1}^{K_n} \left( \mathbb{E}_0(|V_{k,d}^{(0)}|^{2+\delta}) \right)^{1/(1+\delta)} < C.$$

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

(2.iii) The functions  $(Q_{k,1})_{1 \leq k}$  are Lipschitz.

Thus, we have

**Corollary 2.** *If (2.i)–(2.iii) hold, then (A2) and (B) hold, with  $\sigma^2 = \sigma_{SRD}^2$ , so that Theorem 3 obtains.*

*Proof.* The proof follows from Corollary 3 of Doukhan et al. (2015).

*Functions of long-memory Gaussian sequences*

Let  $Z = (X, Y)$  be a bivariate strictly stationary process with independent coordinates  $X$  and  $Y$  satisfying

$$X_t = r_1(N_t) \text{ and } Y_t = r_2(M_t),$$

where  $r_1$  and  $r_2$  are measurable functions,  $(N_t)_{t \in \mathbb{Z}}$  and  $(M_t)_{t \in \mathbb{Z}}$  are independent standard Gaussian long-range dependent processes under  $H_0$ . More precisely, we assume that there exists  $0 < \alpha_N < 1$ ,  $0 < \alpha_M < 1$ , such that  $\mathbb{E}_0(N_0 N_t) \sim |t|^{-\alpha_N}$  and  $\mathbb{E}_0(M_0 M_t) \sim |t|^{-\alpha_M}$ . Here, the null hypothesis is equivalent to  $H_0 : r_1 = r_2 \doteq r$ . Recall that the Hermite rank of a function  $f$  is defined as

$$\min\{k \geq 0 : \mathbb{E}(H_k(N)f(N)) \neq 0\},$$

where  $N$  stands for a standard normal random variable and  $H_k$  denotes the  $k^{\text{th}}$  Hermite polynomial

$$H_k(x) = (-1)^k \exp(x^2/2) (d^k/dx^k) (\exp(-x^2/2)).$$

Let  $m_{k,d}$  be the Hermite rank of  $(Q_{k,d} \circ r)^* = Q_{k,d} \circ r - \mathbb{E}_0(Q_{k,d}(X_0))$  and write  $m = \min\{m_{k,d}, 1 \leq k, 1 \leq d\}$  and we set for the sake of simplicity  $m = m_{1,1}$ .

Notice that under conditions ensuring the square-integrability of  $(Q_{1,1} \circ r)^*$  with respect to the standard Gaussian measure, we have the Mehler formula:

$$(\tilde{Q}_{1,1} \circ r)^* = \sum_{j=m}^{\infty} \frac{q_j}{j!} H_j, \text{ with } q_j = \mathbb{E}_0((\tilde{Q}_{1,1} \circ r)^*(N_0)H_j(N_0)). \quad (15)$$

Let us assume that

**(3.i)** There exist  $C > 0$  and  $n_0 > 0$  such that for all  $n > n_0$ ,

$$\frac{1}{K_n} \sum_{k=1}^{K_n} \mathbb{E}_0(Q_{k,d}(X_0)^2) < C.$$

**(3.ii)**  $m$  is known.

Setting  $\alpha = \min\{\alpha_N, \alpha_M\}$ , the convergence rate of the statistic depends on the rate of decay  $\alpha$ . Fix

$$u_n = \max(\sqrt{n}, n^{1-\alpha m/2})$$

so that

$$r_n(k, d) = \begin{cases} n^{-1/2} \sum_{i=1}^{n-d+1} V_{k,d}^{(i)} & \text{if } \alpha m > 1, \\ n^{\alpha m/2-1} \sum_{i=1}^{n-d+1} V_{k,d}^{(i)} & \text{if } \alpha m < 1. \end{cases}$$

where  $\hat{\sigma}_n^2$  is a consistent estimator of  $\sigma_{SRD}^2$  and  $\hat{\sigma}_n'^2$  is a consistent estimator of (see Doukhan et al. 2015)

$$\sigma_{LRDG}^2 = \frac{2(q_1)^2}{(1-\alpha)(1-\alpha/2)}$$

Then, we have

**Corollary 3.** *If (3.i)–(3.ii) hold, then (A2) and (B) hold, with  $\sigma^2 = \sigma_{SRD}^2$  whenever  $\alpha m > 1$ ,  $\sigma^2 = \sigma_{LRDG}^2$  whenever  $m = 1$  and  $\alpha < 1$ , so that Theorem 3 obtains.*

Corollary 3 does not include the case where  $m > 1$  and  $\alpha m < 1$ . This is because we deliberately restricted our setting to Gaussian limits. In the former case, the limiting distribution is that of an Hermite process at point 1.

*Proof.* The proof follows from Corollary 4 of Doukhan et al. (2015).

*Functions of linear processes*

Let  $Z = (X, Y)$  be a bivariate strictly stationary process with independent coordinates  $X$  and  $Y$  satisfying  $X = r_1(N)$  and  $Y = r_2(M)$  with

$$X_t = \sum_{j < t} \alpha_{t-j} \epsilon_j, \quad \text{and} \quad Y_t = \sum_{j < t} \beta_{t-j} e_j, \quad (16)$$

where the innovations  $\epsilon_i$  and  $e_i$  are i.i.d. and centered standard variables,  $\alpha_j \sim c_\alpha |j|^{\theta_X - 1}$ ,  $\beta_j \sim c_\beta |j|^{\theta_Y - 1}$  with  $\theta_X, \theta_Y < 1/2$ ,  $c_\alpha$  and  $c_\beta$  are real positive constants. For simplicity we shall set  $\theta_Y \leq \theta_X$ .

Following Ho and Sing (1997), we introduce the notion of power rank associated to a measurable function  $L$  such that  $\mathbb{E}(L(N)) = 0$ . We say that  $L$  has power rank  $m \geq 1$  with respect to the linear process  $N$  if

$$L_\infty^{(m)}(0) \neq 0 \quad \text{and} \quad L_\infty^{(r)}(0) = 0, \quad \text{for all } 1 \leq r < m,$$

$$L_\infty^{(r)}(w) = \frac{\partial^r}{\partial w^r} \mathbb{E}(L(w + N_0)). \quad (17)$$

Hereafter, using the same notation as in the preceding subsection, we assume that the power rank of  $(Q_{1,1} \circ r_1)^*$  with respect to  $N$  and that of  $(Q_{1,1} \circ r_2)^*$  with respect to  $M$  both equal 1.

Moreover, we set

**(4.i)**  $(Q_{k,d} \circ r_1)^*$  and  $(Q_{k,d} \circ r_2)^*$  are differentiable with continuous bounded derivative.

**(4.ii)**  $(Q_{k,d} \circ r_1)^*$  and  $(Q_{k,d} \circ r_2)^*$  are  $\beta_{k,d}$ -Lipschitz. and there exists some  $B$  such that for all  $j > 0$ ,  $\frac{1}{j} \sum_{k=1}^j \beta_{k,d}^2 < B$ .

**(4.iii)**  $\mathbb{E}_0(\epsilon_0^4) < \infty$  and  $\mathbb{E}_0(e_0^4) < \infty$ .

**(4.iv)**  $\mathbb{E}_0((Q_{1,1}^*(X_0))^2) < \infty$ .

**(4.v)**  $\mathbb{E}_0(\epsilon_0^8) < \infty$  and  $\mathbb{E}_0(e_0^4) < \infty$ .

**(4.vi)**  $Q_{1,1}^*$  is has a continuous bounded third order derivative.

**(4.vii)**  $\mathbb{E}_0(\epsilon_0^8) < \infty$  and  $\mathbb{E}_0(e_0^8) < \infty$ .

We set

$$u_n = \begin{cases} n^{-1/2}, & \text{if } \theta_X < 0, \\ n^{-1/2 - \theta_X}, & \text{if } \theta_X > 0. \end{cases}$$

Write for the SRD and the LRD cases

$$\sigma_{LSRD}^2 = \sum_{t \in \mathbb{Z}} \gamma_1(t), \quad (18)$$

with  $\gamma_1(t) = \mathbb{E}_0(Q_{1,1}^*(X_0)Q_{1,1}^*(X_t) + Q_{1,1}^*(Y_0)Q_{1,1}^*(Y_t))$ , and

$$\sigma_{LLRD}^2 = c_\alpha^2 \left( (\tilde{Q}_1^*)_\infty^{(1)}(0) \right)^2 C(\delta_X)^2, \quad (19)$$

with  $C(\delta_X) = \sqrt{\frac{\mathcal{B}(1-2\delta_X, \delta_X)}{\delta_X(1+2\delta_X)}}$ , and  $\mathcal{B}(\cdot, \cdot)$  is the Beta function. Here these parameters are estimated with the Whittle estimator as in Doukhan et al. (2019). Combining our Corollary 3 with Propositions 1-3 of Doukhan et al. (2019) we obtain:

**Corollary 4.** *Let  $\theta_Y \leq \theta_X < 0$ . Assume that (4.i) – (4.iv) hold. Then, (A2) and (B) hold, with  $\sigma^2$  given by (18), and Theorem 3 occurs.*

**Corollary 5.** *Let  $\theta_Y < 0 < \theta_X$ . Assume that (4.ii), (4.iv) – (4.vi) hold. Then, (A2) and (B) hold, with  $\sigma^2$  given by (19), and Theorem 3 occurs.*

**Corollary 6.** *Assume that (4.ii), (4.vi) – (4.vii) hold. Then*

**a)** *if  $0 < \theta_Y < \theta_X$ , (A2) and (B) hold with  $\sigma^2$  defined by (19),*

**b)** *if  $0 < \theta_Y = \theta_X$ , (A2) and (B) hold with  $\sigma^2$  defined by  $\sigma^2 = (c_\alpha^2 + c_\beta^2) \left( (Q_{1,1}^*)_\infty^{(1)}(0) \right)^2 C(\theta_X)^2$ ,*

*and in both cases Theorem 3 is obtained.*



## Appendix 2: proofs

*Proof of Theorem 1.* We want to prove that  $\mathbb{P}_0(k_n(d) \geq 2)$  vanishes as  $n \rightarrow +\infty$ . By definition of  $k_n(d)$  we have:

$$\begin{aligned}
\mathbb{P}_0(k_n(d) \geq 2) &= \mathbb{P}_0(\exists j, 2 \leq |j| \leq K_n : T_n(j, d) - [j]_d \log(n) \geq T_n(1, d) - d \log(n)) \\
&\leq \sum_{|j|=2}^{K_n} \mathbb{P}_0(T_n(j, d) - [j]_d \log(n) \geq T_n(1, d) - d \log(n)) \\
&= \sum_{|j|=2}^{K_n} \mathbb{P}_0\left(\sum_{|k| \leq |j|} \tilde{r}_n(k, d)^2 - [j]_d \log(n) \geq \sum_{|k|=1} \tilde{r}_n(k, d)^2 - d \log(n)\right) \\
&\leq \sum_{|j|=2}^{K_n} \mathbb{P}_0\left(\sum_{|k| \leq |j|} \tilde{r}_n(k, d)^2 \geq ([j]_d - d) \log(n)\right) \\
&\leq \sum_{|j|=2}^{K_n} \mathbb{P}_0\left(\sum_{|k| \leq |j|} r_n(k, d)^2 \geq ([j]_d - d) \log(n) e_n\right) \\
&\leq \sum_{|j|=2}^{K_n} \mathbb{P}_0\left([\jmath]_d - d) \sup_{|k| < |j|} r_n(k, d)^2 \geq ([j]_d - d) \log(n) e_n\right) \\
&\leq \sum_{|k|=2}^{K_n} \mathbb{P}_0(r_n(k, d)^2 \geq \log(n) e_n).
\end{aligned}$$

By Markov inequality we deduce that

$$\mathbb{P}_0(k_n(d) \geq 2) \leq \frac{1}{K_n} \sum_{|j|=2}^{K_n} \mathbb{E}_0(r_n(k, d)^2) \frac{K_n}{\log(n) e_n}.$$

Classical calculations using the stationarity of the process yield

$$\begin{aligned}
\mathbb{E}_0(r_n(k, d)^2) &= \frac{1}{u_n^2} \sum_{s=1}^{n-d} \sum_{t=1}^{n-d} \mathbb{E}_0(V_{k,d}^{(s)} V_{k,d}^{(t)}) \\
&= \frac{n-d}{u_n^2} \sum_{|t| < n-d} \left(1 - \frac{|t|}{n-d}\right) \mathbb{E}_0\left(V_{k,d}^{(0)} V_{k,d}^{(t)}\right) \\
&\leq \frac{2(n-d)}{u_n^2} \sigma_{k,d}(n),
\end{aligned} \tag{20}$$

which implies that

$$\mathbb{P}_0(k_n(d) \geq 2) \leq \frac{1}{K_n} \sum_{|k|=2}^{K_n} \sigma_n(k, d) \frac{2(n-d)}{u_n^2} \frac{K_n}{\log(n) e_n}, \tag{21}$$

and the proof follows from **(A1–A2)**.

*Proof of Theorem 2.* Let us prove that for all  $\delta > 0$ ,  $\mathbb{P}_0(d_n(\delta) \geq 2)$  vanishes as  $n \rightarrow +\infty$ . By definition of  $v(n)$  we have:

$$\begin{aligned}
& \mathbb{P}_0(d_n(\delta) \geq 2) \\
&= \mathbb{P}_0 \left( \exists d, 2 \leq d \leq L(n) : \sum_{|k| \leq k_n(d)} \tilde{r}_n(k, d)^2 - [k_n(d)]_d \log(n) \geq \sum_{|k|=1} \tilde{r}_n(k, d)^2 - \log(n) \right) \\
&\leq \sum_{d=2}^{\delta} \mathbb{P}_0 \left( \sum_{|k| \leq k_n(d)} r_n(k, d)^2 \geq ([k_n(d)]_d - 1) \log(n) e_n \right) \\
&= \sum_{d=2}^{\delta} \mathbb{P}_0 \left( \sum_{|k| \leq k_n(d)} r_n(k, d)^2 \geq ([k_n(d)]_d - 1) \log(n) e_n \cap k_n(d) > 1 \right) \\
&\quad + \sum_{d=2}^{\delta} \mathbb{P}_0 \left( \sum_{|k| \leq k_n(d)} r_n(k, d)^2 \geq ([k_n(d)]_d - 1) \log(n) e_n \cap k_n(d) = 1 \right) \\
&\leq \sum_{d=2}^{\delta} \left[ \mathbb{P}_0(k_n(d) > 1) + \mathbb{P}_0 \left( \sum_{|k|=1} r_n(k, d)^2 \geq (d-1) \log(n) e_n \right) \right] \\
&\leq \sum_{d=2}^{\delta} \left[ \mathbb{P}_0(k_n(d) > 1) + \mathbb{P}_0 \left( \sum_{|k|=1} r_n(k, d)^2 \geq \log(n) e_n \right) \right] \\
&= A + B
\end{aligned}$$

From (21), combining **(A2)** and **(A3)** to get

$$\begin{aligned}
A &\leq \sum_{d=2}^{\delta} \frac{1}{K_n} \sum_{|k|=2}^{K_n} \sigma_n(k, d) \frac{2(n-d)}{u_n^2} \frac{K_n}{\log(n) e_n} \\
&\leq M \frac{\delta K_n}{\log(n) e_n}.
\end{aligned}$$

which tends to 0, as  $n \rightarrow +\infty$ . By Markov inequality,

$$B \leq \sum_{d=2}^{K_n} \sum_{|k|=1} \mathbb{E}_0(r_n(k, d)^2) \frac{1}{\log(n) e_n},$$

and from (20) we have

$$B \leq \sum_{d=2}^{\delta} \sum_{|k|=1} \frac{2(n-d)}{u_n^2} \sigma_{k,d}(n) \frac{1}{\log(n) e_n}.$$

We now apply **(A2)** with  $K_n = 1$  to get

$$B \leq M \frac{\delta}{\log(n)e_n},$$

which tends to zero with **(A3)**.

*Proof of Proposition 1.*

*First step.* We first prove that under  $H_1(k_1, d_1)$  we have

$$k_n(d) \xrightarrow{\mathcal{P}} 1, \quad \forall d < d_1 \quad (22)$$

$$k_n(d_1) \xrightarrow{\mathcal{P}} k_1, \quad (23)$$

where  $n$  tend to infinity. The proof of (22) mimics the proof of Theorem 1. To prove (23) we write

$$T_n(k_1, d_1) - [k_1]_{d_1} \log(n) \geq \left( (\tilde{r}_n(k_1, d_1) - \frac{n}{u_n} \delta_1) + \frac{n}{u_n} \delta_1 \right)^2 - [k_1]_{d_1} \log(n).$$

In the SRD case,  $n/u_n = n^{1/2}$ . In the LRD case, there exists  $\beta > 0$  such that  $n/u_n = n^\beta$ . Then in both case we conclude that,  $T_n(k_1, d_1) - [k_1]_{d_1} \log(n) \xrightarrow{\mathcal{P}} +\infty$ , that is for all  $\epsilon > 0$ ,

$$\mathbb{P}(T_n(k_1, d_1) - [k_1]_{d_1} \log(n) > \epsilon) \rightarrow 1,$$

as  $n$  tends to infinity and we get the result.

*Second step.* We prove that under  $H_1(k_1, d_1)$  we have for all  $\delta > 0$ ,  $\mathbb{P}(d_n(\delta) < d_1) \rightarrow 0$ . By construction of  $d_n(\delta)$ , we have for all  $d < d_1$ :

$$\begin{aligned} \mathbb{P}(d_n(\delta) = d) &\leq \mathbb{P}(T_n(k_n(d), d) - [k_n(d)]_d \log(n) \geq T_n(k_n(d_1), d_1) - [k_n(d_1)]_{d_1} \log(n)) \\ &= \mathbb{P}(T_n(k_n(d_1), d_1) - T_n(k_n(d), d) \leq ([k_n(d_1)]_{d_1} - [k_n(d)]_d) \log(n)) \\ &= \mathbb{P}\left(\sum_{j \in \mathbb{N}^{d_1}, |j|=1}^{k_n(d_1)} \tilde{r}_n(j, d_1)^2 - \sum_{j \in \mathbb{N}^d, |j|=1}^{k_n(d)} \tilde{r}_n(j, d)^2 \leq ([k_n(d_1)]_{d_1} - [k_n(d)]_d) \log(n)\right) \end{aligned}$$

Using the embedding property of the orthogonal bases deduced from ??, we get

$$\begin{aligned} \mathbb{P}(d_n(\delta) = d) &\leq \mathbb{P}(\tilde{r}_n(k_1, d_1)^2 \leq ([k_n(d_1)]_{d_1} - [k_n(d)]_d) \log(n)) \\ &= \mathbb{P}\left(\frac{1}{\log(n)} \tilde{r}_n(k_1, d_1)^2 \leq [k_n(d_1)]_{d_1} - [k_n(d)]_d\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{\log(n)}} |\tilde{r}_n(k_1, d_1)| \leq \sqrt{[k_n(d_1)]_{d_1} - [k_n(d)]_d}\right) \end{aligned}$$

We can decompose

$$\begin{aligned} \frac{1}{\sqrt{\log(n)}} \tilde{r}_n(k_1, d_1) &= \frac{1}{\sqrt{\log(n)}} \left( \tilde{r}_n(k_1, d_1) - \frac{n}{u_n} \delta_1 \right) + \frac{n \sqrt{\log(n)}}{u_n} \delta_1 \\ &= A + B. \end{aligned}$$

From **(C)**,  $(\tilde{r}_n(k_1, d_1) - \frac{n}{u_n} \delta_1)$  has a limit distribution and then  $A$  converges to a Dirac in zero. Moreover, in the SRD case,  $n \sqrt{\log(n)}/u_n = \sqrt{n \log(n)}$  and  $B$  diverges since  $\delta_1 \neq 0$ . Similarly, in the LRD case, there exists  $\beta > 0$  such that  $n \sqrt{\log(n)}/u_n = n^\beta \sqrt{\log(n)}$  and then  $B$  diverges. From (22) and (23) we have  $\sqrt{[k_n(d_1)]_{d_1} - [k_n(d)]_d} \xrightarrow{\mathcal{P}} \sqrt{[k_1]_{d_1} - [1]_d} < \infty$  and we conclude that for all  $d < d_1$ ,

$$\mathbb{P}(d_n(\delta) = d) \rightarrow 0.$$

### Appendix 3: Illustration of assumption **(C)**: alternative behaviour for linear transformations

We consider the case of transformations of strictly stationary linear processes here, but this result can be extended to Gaussian processes,  $\alpha$ -mixing processes or  $\theta$ -dependent processes. If

$$Z_t = \sum_{j < t} \alpha_{t-j} \epsilon_j, \quad \alpha_j \sim c|j|^{\delta-1}, \quad (24)$$

where the  $\epsilon_j$ 's are i.i.d. zero centered variables,  $c$  is a real non null constant and  $\delta < 1/2$ , we say that a measurable function  $L$  such that  $\mathbb{E}L(Z) = 0$  has power rank  $k \geq 1$  with respect to the linear process  $Z$  if

$$L_\infty^{(k)}(0) \neq 0 \quad \text{and} \quad L_\infty^{(r)}(0) = 0, \quad \text{for all } 1 \leq r < k,$$

$$L_\infty^{(r)}(w) = \frac{\partial^r}{\partial w^r} \mathbb{E}(L(w + Z_0)). \quad (25)$$

So,  $k = \min\{r \geq 1, L_\infty^{(r)}(0) \neq 0\}$ , with  $L_\infty^{(r)}$  defined by (25). Let us set the following assumptions:

- (i)  $L$  has a continuous bounded derivative of order  $k + 2$ .
- (ii)  $\mathbb{E}(\epsilon_1^{2k \vee 8}) < \infty$
- (iii)  $\mathbb{E}(\epsilon_1^8) < \infty$
- (iv) We have

$$\mathbb{E}(L(Z_1) - L(Z_{1,m}))^2 \xrightarrow{m \rightarrow \infty} 0, \quad Z_{1,m} = \sum_{1 \leq i \leq m} \alpha_i \epsilon_{1-i}.$$

We recall below the result of [6] (see also [?] for improved conditions) for the partial sums of the process  $L(Z)$ :

**Proposition 2** ([6]). *Let  $L$  be a measurable function such that  $\mathbb{E}(L(Z_1)) = 0$  and  $\mathbb{E}(L^2(Z_1)) < \infty$  with  $Z$  defined as in (24). Let  $k$  be the power rank of  $L$ . Assume that (i) holds. Then*

- if  $k(1 - 2\delta) < 1$  and (ii) holds,

$$n^{k(1/2-\delta)-1} \sum_{s=1}^n L(Z_s) \xrightarrow{\mathcal{L}} T_{k,\delta,c,L}^\infty = c^k C(\delta, k) L_\infty^{(k)}(0) Z_{k,1-\delta}(1), \quad (26)$$

where

$$C(\delta, k) = \sqrt{\frac{k!(1 - k(1/2 - \delta))(1 - k(1 - 2\delta))}{(\int_0^\infty (x + x^2)^{\delta-1})^k}}, \quad (27)$$

$Z_{k,1-\delta}(1)$  is the Hermite process at point 1 (see [11]).

- if  $k(1 - 2\delta) > 1$  and (iii)-(iv) hold, then

$$n^{-1/2} \sum_{s=1}^n L(Z_s) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad (28)$$

where

$$\sigma^2 = \lim \frac{1}{n} \text{var} \left( \sum_{s=1}^n L(Z_s) \right) = \sum_{s \in \mathbb{Z}} \mathbb{E}(L(Z_0)L(Z_s)) > 0.$$

If we assume that there exists two measurable functions  $L$  and  $L'$  such that  $X_t = L(Z_t)$  and  $Y_t = L'(Z'_t)$ , where  $Z_t$  and  $Z'_t$  are governed by (24) with evident notations, then (C) follows from Proposition 2 and then Proposition 1 occurs.

## Bibliography

- [1] T. Carleman. *Les fonctions quasi analytiques*. Collection de Monographies sur la Théorie des Fonctions, Gauthier-Villars, Paris, 1926.
- [2] S.S. Dhar, B. Chakraborty, and P. Chauduri. Comparison of multivariate distributions using quantile-quantile plots and related tests. *Bernoulli*, 20:1484–1506, 2014.
- [3] P. Doukhan, I. Grublyte, D. Pommeret, and L. Reboul. Comparing the marginal densities of two strictly stationary linear processes. *Annals of the Institute of Statistical Mathematics*, <https://doi.org/10.1007/s10463-019-00730-6>, 2019.
- [4] P. Doukhan, D. Pommeret, and L. Reboul. Data driven smooth test of comparison for dependent sequences. *Journal of Multivariate Analysis*, 139:147–165, 2015.

- [5] C.W.J Granger and R. Joyeux. An introduction to long-memory series models and fractional differencing. *Journal of Time Series Analysis*, 1:15–29, 1980.
- [6] H. C. Ho and T. Hsing. Limit theorems for functionals of moving averages. *The Annals of Probability*, 25:1636–1669, 1997.
- [7] J. R. M. Hosking. Fractional differencing. *Biometrika*, 68:165–176, 1981.
- [8] 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, 1995.
- [9] T. Ledwina. Data-driven version of Neyman’s smooth test of fit. *Journal of the American Statistical Association*, 89:1000–1005, 1994.
- [10] P.R. Rosenbaum. An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society, serie B*, 67:515–530, 2005.
- [11] D. Surgailis. Long-range dependence and Appell rank. *The Annals of Probability*, 28:478–497, 2000.
- [12] G. Wylupek. Data-driven k sample tests. *Technometrics*, 52:107–123, 2010.