Data driven smooth test of comparison for dependent sequences

P. Doukhan\textsuperscript{a}, D. Pommeret\textsuperscript{b}, L. Reboul\textsuperscript{b}

\textsuperscript{a}University Cergy-Pontoise, Mathematics, office E 5.28, UCP site Saint-Martin, 2 Bd. Adolphe Chauvin, 95000 Cergy-Pontoise.
\textsuperscript{b}Aix-Marseille University, Campus de Luminy, 13288 Marseille cedex 9.

Abstract

In this paper we propose a smooth test of comparison for the marginal distributions of two possibly dependent strictly stationary sequences. We first state a general test procedure. Several cases of dependence are then investigated, allowing to cover various real situations. The test is applied to both simulated data and real datasets obtained from financial markets.

Keywords: Smooth test, Schwarz’s rule, Strictly stationary process, \( m \)-dependence, Strong mixing, \( \theta \)-dependence, Long memory, Gaussian processes

1. Introduction

The comparison of two or more time series is a problem of great interest in many practical situations. The independence property both within and between series is generally assumed in comparison methods as well as the assumption of identically drawn observations. These assumptions are generally too restrictive in practice. For instance, most of economic or financial time series exhibit dependence and a drift in time. It is however often realistic to consider phenomenons on which time has only transitory effects, replacing independent and identically drawn assumptions by that of strict stationarity of the sequences. Moreover, even when stationarity can look like a heavy restriction at first glance, there are often simple transformations of the series that can be plausibly assumed to be stationary such as linear detrending or first differentiating.

In this paper, we wish to compare the marginals of two possibly dependent time phenomenons modeled by strictly stationary time processes \( X \) and \( Y \) such that \( Z = (X, Y) \) is a bivariate dependent process. When the phenomenons are known to be strictly stationary with a known in advance within dependence structure, denoting by \( f_X \) and \( f_Y \) the unknown marginal densities...
at any given time of \( X \) and \( Y \), our comparison problem relies in testing the nonparametric hypothesis
\[
H_0 : f_X = f_Y, \tag{1}
\]
versus the omnibus alternative that the two marginal distributions are different, based on a vector of observations of each process.

This two-sample nonparametric testing problem has been widely studied when there is no dependence within and between series (see [15], [10] and references therein). Most of the usual tests rely on ranks or empirical CDF (see for instance [15]). For possibly cross-dependent data, a performing test strategy inspired from Neyman’s smooth test [20] was recently proposed by [10]. In this paper, the authors considered the series expansions of \( f_X \) and \( f_Y \) along an appropriated family of orthogonal functions. Testing (1) thus reduces to the parametric testing problem that the coefficients of the same order in both expansions are equal. Then, they adapted a data-driven method (see [16]) to choose the optimal dimension of the orthogonal family. It consists of a modified Schwartz’s Bayesian information criterion. Finally, they obtained a test statistic and derived its asymptotic distribution.

In the context of within dependence, the development of practicable tests becomes usually more difficult. In the one sample case, [13], then [18] recently proposed goodness-of-fit tests for the marginal distribution of observations arising from an \( \alpha \)-mixing discrete time stochastic process. [18]’s test is a data-driven version of Neyman’s smooth test adapted to dependent data.

In this paper, we propose a test strategy for (1) generalizing [18, 10]’s results to several cases of dependence of the observations. We first develop a general theory in which the test statistic is defined as a normalized sum of the squared differences between the estimated coefficients in the expansions of \( f_X \) and \( f_Y \) in an appropriated basis. For each particular dependence structure within observations, the normalization is chosen in such a way that a central limit theorem obtains. The number of components to keep in the sum is selected by the way of an information criterion inspired from [10] and [18]. We investigate several dependence structures, covering short and long memory processes. In so doing, we provide a wide panorama of dependence situations which can be adapted to various concrete phenomenons, such as economical data series. For the practical implementation of the test, some modifications in the definition of our test statistic and of our selection rule are considered and discussed.

Eventually, our method is illustrated on simulations and real data arising from short memory financial index time series: the Dow Jones Composite Average, the NASDAQ Composite and the NYSE International 100 Index.

The rest of the paper is organized as follows. In Section 2 we develop the methodology of the test and a data-driven statistic is introduced. Main results are stated in Section 3. In Section 4 we apply our general methodology to several situations of dependence. In Section 5 the practical implementation of the test procedure is discussed. Sections 6 and 7 propose a short simulation study and an application to real financial data. Finally, a conclusion is given in Section 8 and Section 9 is devoted to the proofs.
2. Methodology

2.1. Construction of the test statistic

Let \( Z = (Z_t)_{t \in \mathbb{Z}} \) be a discrete time process defined on some probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \) and taking values in \( \mathbb{R}^2 \). We assume throughout that \( Z \) is strictly stationary. Let \( \nu \) be a given probability measure with density \( h \) with respect to some reference measure \( \lambda \) (Lebesgue’s or counting measure for instance). We set \( Z = (X,Y) \), where \( X \) and \( Y \) are possibly cross-dependent. We denote by \( f_X \) and \( f_Y \) the respective common unknown marginal densities of the \( X_t \)'s and \( Y_t \)'s with respect to \( \lambda \) and assume that they belong to \( L^2(\nu) \).

In this setup, we wish to test (1) based on the observation of a stretch \((Z_1, \ldots, Z_s, \ldots, Z_n)\) of \( Z \). For that task, we consider the expansions of \( f_X \) and \( f_Y \) along a dense family \((Q_j)_{j \in \mathbb{N}}\) of orthonormal functions in \( L^2(\nu) \):

\[
f_X = \sum_{j \geq 0} a_j Q_j \quad \text{and} \quad f_Y = \sum_{j \geq 0} b_j Q_j ,
\]

with

\[
a_j = \mathbb{E}(\tilde{Q}_j(X_1)) = \int_{\mathbb{R}} Q_j(t) f_X(t) d\nu(t) , \quad b_j = \mathbb{E}(\tilde{Q}_j(Y_1)) = \int_{\mathbb{R}} Q_j(t) f_Y(t) d\nu(t),
\]

and \( \tilde{Q}_j = hQ_j \) for all \( j \in \mathbb{N} \). It is clear that \( H_0 \) can be rewritten as \( H_0 : a_i = b_i \), for all \( i = 1, 2, \cdots \). Then we consider the parametric testing problems:

\[
H_0^{(d(n))} : a_j = b_j, \quad \text{for all} \quad j = 1, \cdots, d(n), \quad (3)
\]

against

\[
H_1^{(d(n))} : \exists j \in \{1, \cdots, d(n)\}, \quad a_j \neq b_j,
\]

where \( \lim_{n \to \infty} d_n = \infty \).

In view to define a test strategy for (3), let us set

\[
\Phi_j(x,y) = \tilde{Q}_j(x) - \tilde{Q}_j(y), \quad V^{(j)}_t = \Phi_j(X_t, Y_t), \quad V_i(k) = (V^{(j)}_t)_{1 \leq j \leq d(n)} , \quad V(k) = (V_i(k))_{i \in \mathbb{Z}},
\]

for all \( k = 1, \cdots, d(n) \). Next, define the \( k \)-dimensional random vector of the empirical mean of observations

\[
U_n(k) = (U^{(j)}_n)_{1 \leq j \leq k} = \frac{1}{u_n} \sum_{s=1}^{n} V_s(k),
\]

where \((u_n)_{n>0}\) is an appropriately chosen positive norming sequence such that \( \lim_{n \to \infty} u_n = \infty \). We will consider the Neyman’s type test statistic defined as follows:

\[
N_n(k) = \|U_n(k)\|^2 = \sum_{j=1}^{k} \left( U^{(j)}_n \right)^2 , \quad k = 1, \cdots, d(n), \quad (4)
\]
According to the dependence structure of $Z$, we generally expect to obtain the limiting law of $U_n(k)$ by application of a central limit theorem, with an appropriate choice of $u_n$ and additional moment assumptions (see [10] and [21]). Therefore the limiting distribution of $N_n(k)$, or of a rescaled version of it, is known. When $d(n)$ is fixed, this distribution has been studied in [13].

2.2. A data-driven selection rule

In view to select the number of components $k$, we use in the sequel a data-driven selection rule: at first step, we select among every $k = 1, \ldots, d(n)$ the value $K_n$ that minimizes the information criterion

$$K_n = \min \left\{ \arg \max_{1 \leq k \leq d(n)} (N_n(k) - k \log(n)) \right\}.$$  

(5)

Once $K_n$ is determined, we use the test statistic

$$\tilde{N}_n = N_n(K_n).$$  

(6)

Criterion (5) has been used in [16] for independent observations and consists in this case of a modified version of [24]'s Bayesian information rule, based on an expansion of the maximum likelihood function. The extension of this rule to the paired and strong-mixing contexts have been heuristically justified by [10] and [18] respectively (see Remarks 3 of both papers).

The test statistic (6) is an extension to the two-sample case and to general orthogonal families of [18]'s test statistic $R_k$. It suffers some disadvantages that we will discuss in Section 5.

3. Main results

3.1. Assumptions

From now on, we make the assumption that $d(n) = o(\log(n))$. Let us denote by $r_n(k)$ the partial sum of the auto-covariance series:

$$r_n(k) = \sum_{t=0}^{n-1} \left| \mathbb{E}_0 (V_0^{(k)} V_t^{(k)}) \right|,$$  

(7)

where $\mathbb{E}_0$ is the expectation under $H_0$. We will need the following assumptions:

(A): $\frac{1}{d(n)} \sum_{k=1}^{d(n)} r_n(k) = O \left( \frac{u_n^2}{n} \right)$.

(B): $U_n(1) \overset{L}{\to} U$ under $H_0$, where $U$ is a random variable whose distribution possibly depends on a nuisance parameter.

Assumption (B) amounts to say that a central limit theorem holds for the partial sum of the process $V^{(1)}$. Assumption (A) informs about the asymptotic behavior of the partial sums of the auto-covariance series. Notice that when
processes $V^{(k)}$, $k = 1, \ldots, d(n)$ have short-memory, $u_n = \sqrt{n}$ and \((A)\) obtains by definition. In a long-memory context, \((A)\) makes a control of the rate of divergence of $r_n(k)$ to infinity.

**Remark 1.** Assumption \((A)\) can be weakened using a control of the second moment of the test statistic (see (20)). More precisely, the results of this Section still hold when there exists $C > 0$ such that

$$
\frac{1}{d(n)} \sum_{k=1}^{d(n)} \mathbb{E}_0 \left( |U_n^{(k)}|^2 \right) < C,
$$

or

$$
\sup r_n(k) < C.
$$

### 3.2. Convergence under the null hypothesis

Hereafter, we give the asymptotic distribution of $\tilde{N}_n$ under the null.

**Theorem 1.** Assume that \((A)\) holds. Thus, under $H_0$,

$$
K_n \xrightarrow{p} 1. \quad (8)
$$

As a consequence of Theorem 1, one has the following

**Corollary 1.** Let assumptions \((A)\) and \((B)\) hold. Then

$$
\tilde{N}_n \xrightarrow{L} U^2. \quad (9)
$$

Therefore, we can use the quantiles of $U^2$ to calibrate our test, as soon as its distribution is parameter free. Otherwise, we shall build estimated quantiles by simply plugging-in a suitable estimator of the unknown nuisance parameter in the limiting distribution.

### 3.3. Convergence under contiguous alternatives

For suitable alternatives, the test based on the limiting quantile is consistent. Namely, let us set $\delta_k = a_k - b_k$ and consider the following two alternatives:

\[
\begin{align*}
H_1^* & : \forall k \geq 1, \delta_k = o(u_n/n), \\
H_1^{**} & : \exists K > 1 \text{ such that } \forall k \neq K, \delta_k = o(u_n/n) \text{ and } \delta_K = O(\sqrt{h(n)u_n/n^\beta}), \text{ with } \beta < 1.
\end{align*}
\]

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

*(A*): There exists some $C^* > 0$ such that
\[ \frac{1}{d(n)} \sum_{k=1}^{d(n)} E_1(|U_n^{(k)} - \delta_k|^p) < C^*, \]

\((B^*)\): \( U_n^{(K)} - \frac{n}{u_n} \delta_K = O_{\mathbb{P}_1}(\sqrt{\log(n)}) \).

**Theorem 2.** Assume that \((A^*)\) holds. Then,

- Under \(H_1^*\), \( K_n \xrightarrow{\mathbb{P}} 1 \). Moreover, if \((B)\) holds, \( \tilde{N}_n \xrightarrow{\mathbb{L}} U^2 \).
- Under \(H_1^{**}\), \( K_n \xrightarrow{\mathbb{P}} K \). Moreover, if \((B^*)\) holds, \( \tilde{N}_n \xrightarrow{\mathbb{P}} +\infty \).

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

**Remark 2.**

- Theorem 2 still holds if we change \(H_1^{**}\) into \(H_1^{***}: \exists K > 1 \) such that
  \[
  \forall j < K, \delta_j = o(\sqrt{u_n/n}), \\
  \forall j > K, \delta_j = o\left(\log(n)^{1/2}u_n(nd(n)^{1/2})^{-1}\right), \\
  \text{and} \delta_K = O(\sqrt{\log(n)}u_n/n^\beta), \text{and } \beta < 1.
  \]
  Moreover, if we replace the condition \(\beta < 1\) by \(\beta \geq 1\) we obtain \(K_n \xrightarrow{\mathbb{P}} 1\).
  - Notice that \((B^*)\) holds as soon as \(V^{(K)}\) is ergodic.

**4. Testing dependent processes**

In this section, we apply the previous general theoretical results to various dependence structures of the stationary process \(Z\). We first investigate short-memory cases, in which dependence within observations dies out sufficiently fast to allow the absolute convergence of the auto-covariances series \( (\lim_{n \to \infty} r_n(k)) < \infty \). In such cases, for convenient classes of functions \(\Phi_k\), processes \( V^{(k)} = \Phi_k(Z) \) inherit the short memory properties of \(Z\) and standard central limit theorems ensuring \((B)\) exist under suitable conditions. Hereafter, we consider \(m\)-dependent processes and particular classes of mixing and weakly dependent processes to illustrate such context. Heredity properties no more hold in the long memory case. More precisely, when \(Z\) is long memory, the \(V^{(k)}\)'s not necessarily inherit this property. In particular, no general central limit theorem is available for the \(V^{(k)}\)'s so that our test strategy cannot generally be handled. In the sequel, we present the tractable particular case of a bivariate process \(Z\) with gaussian independent coordinates \(X\) and \(Y\). Finally, we briefly indicate other possible applications. In particular, we mention the case of Infinite...
moving average processes, which will be developed in a Complementary Note. Hereafter, we set

$$\sigma^2 = \sum_{t=-\infty}^{+\infty} E_0(V_0(1)V_t(1)),$$

(10)

$$\sigma'^2 = \frac{2(q_1(1))^2}{(1-\alpha)(1-\alpha/2)},$$

(11)

where $q_1(1)$ is defined by (14). Table 1 below summarizes the application studied in this Section.

<table>
<thead>
<tr>
<th>Dependence</th>
<th>Conditions</th>
<th>Under $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>$u_n = \sqrt{n}$, and $\exists C &gt; 0,$</td>
<td>$\tilde{N}_n \to U^2$</td>
</tr>
<tr>
<td>$m$-dependence</td>
<td>$\frac{1}{d(n)} \sum_{k=1}^{d(n)} E_0(V_0^{(k)})^2 &lt; C$</td>
<td>$U^2 \sim N(0, \sigma^2)$</td>
</tr>
<tr>
<td>$\alpha$-Mixing</td>
<td>$u_n = \sqrt{n}$, and $\exists C &gt; 0, \delta \geq 0,$</td>
<td>$\tilde{N}_n \to U^2$</td>
</tr>
<tr>
<td>coefficients</td>
<td>$\frac{1}{d(n)} \sum_{k=1}^{d(n)} (E_0</td>
<td>V_0^{(k)}</td>
</tr>
<tr>
<td>$\theta$-dependence</td>
<td>$u_n = \sqrt{n}$, and $\exists C &gt; 0, \delta \geq 0,$</td>
<td>$\tilde{N}_n \to U^2$</td>
</tr>
<tr>
<td>coefficients</td>
<td>$\frac{1}{d(n)} \sum_{k=1}^{d(n)} (E_0</td>
<td>V_0^{(k)}</td>
</tr>
<tr>
<td>$\theta(m)$</td>
<td>$\sum_{m&gt;0} m^{2/\delta} \alpha(m) &lt; \infty$</td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td>$u_n = \max(\sqrt{n}, n^{1-\alpha m/2}),$</td>
<td>$\tilde{N}_n \to U^2$</td>
</tr>
<tr>
<td>Hermite rank</td>
<td>and $\exists C &gt; 0, \delta \geq 0,$</td>
<td>$U^2 \sim N(0, \sigma^2)$ if $\alpha m &gt; 1$</td>
</tr>
<tr>
<td>$m$</td>
<td>$\frac{1}{d(n)} \sum_{k=1}^{d(n)} (E_0</td>
<td>V_0^{(k)}</td>
</tr>
<tr>
<td>$\theta(m)$</td>
<td>$\sum_{m&gt;0} m^{1/\delta} \theta(m) &lt; \infty$</td>
<td></td>
</tr>
<tr>
<td>$\theta(m)$</td>
<td>$E_0(X_0X_t)$ and $E(Y_0Y_t) \sim t^{-\alpha}$</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Summary of the studied cases

4.1. **Independent and $m$-dependent sequences**

We say that a unidimensional strictly stationary process $(X_t)_{t \in \mathbb{Z}}$ is $m$-dependent for some $m \geq 0$ if any finite-dimensional margins $(X_{j_1}, \ldots, X_{j_n})$ and $(X_{k_1}, \ldots, X_{k_l})$ such that $k_1 - j_n > m$ are independent. In particular, a stationary sequence is 0-dependent if and only if it is i.i.d. $m$-dependent processes include several classical models such as moving average processes. We assume in the sequel that $X$ and $Y$ are two possibly cross-dependent $m$-dependent sequences and we
fix \( u_n = \sqrt{n} \) so that
\[
U_n(k) = n^{-1/2} \sum_{s=1}^{n} V_s(k).
\]

**Corollary 2.** Assume that there exists some \( C > 0 \) such that

(i) \[
\frac{1}{d(n)} \sum_{k=1}^{d(n)} \mathbb{E}_0(V_0(k))^2 < C.
\]

Thus, under \( H_0 \)
\[
\tilde{N}_n \xrightarrow{L} U^2, \text{ with } U \sim \mathcal{N}(0, \sigma^2).
\]

### 4.2. Mixing sequences

We assume in the sequel that \( Z \) satisfies a mixing condition. More specifically, we assume here strong or \( \alpha \)-mixing in [23]'s sense. Formally, setting \( \mathcal{P} = \sigma(Z_t, t \leq 0), \mathcal{F} = \sigma(Z_t, t \geq m) \) and defining the decreasing sequence of strong mixing coefficients of \( Z \) by

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

we say that \( Z \) is \( \alpha \)-mixing if it satisfies the condition:

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

A lot of classical models satisfy this condition. In particular, the important class of linear stochastic processes are strong mixing, provided that they are based on innovation random variables with a Lebesgue-integrable characteristic function (see [26]). We fix \( u_n = \sqrt{n} \) so that
\[
U_n(k) = n^{-1/2} \sum_{s=1}^{n} V_s(k).
\]

**Corollary 3.** Assume that \( Z \) is a strictly stationary \( \alpha \)-mixing process with non-increasing mixing coefficients sequence \((\alpha(m))_{m>0}\) and such that

(i) There exists \( C > 0 \) and \( \delta > 0 \) such that
\[
\frac{1}{d(n)} \sum_{k=1}^{d(n)} \left( \mathbb{E}_0(|V_0(k)|^{2+\delta}) \right)^{2/(2+\delta)} < C.
\]

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

Thus, under \( H_0 \)
\[
\tilde{N}_n \xrightarrow{L} U^2, \text{ with } U \sim \mathcal{N}(0, \sigma^2).
\]
4.3. Weakly dependent sequences

Although mixing properties are satisfied by fairly general models, they are not easy to check and do not always hold. Namely, they do not cover the case of linear processes with discrete innovations (see e.g. [1]). A less restrictive condition is to assume that \( Z \) is weakly dependent as defined in [9]. Among various kind of weak dependence structures, we focus here on \( \theta \)-dependence (see [5] for illustrations). Formally, let us define the Lipschitz modulus of a function \( g \) from \( \mathbb{R}^d \) into \( \mathbb{R} \) by

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

where \( \|x\|_1 = \sum_{i=1}^d |x_i| \). For \( (u,v) \in \mathbb{N}^2 \), let \( \mathcal{F}_u \) and \( \mathcal{G}_v \) respectively denote the set of measurable functions from \( (\mathbb{R}^2)^u \) into \( \mathbb{R} \) that are bounded by 1 and the set of functions from \( (\mathbb{R}^2)^v \) into \( \mathbb{R} \) that are bounded by 1 and have finite Lipschitz modulus. Let \( \{\theta(m), m \geq 0\} \) be a decreasing sequence tending to 0 as \( m \) goes to infinity. We say that \( Z \) is \( \theta \)-dependent if for all \( (u,v) \) and all set of indices \( i_1 \leq \ldots \leq i_u \leq j_1 - m \leq j_v \leq \ldots \leq j_v \), we have for all \( f \in \mathcal{F}_u, g \in \mathcal{G}_v \),

\[
|\text{cov}(f(P), g(F))| \leq v\text{Lip}(g)\theta(m),
\]

with \( P = (Z_{i_1}, \ldots, Z_{i_u}) \) and \( F = (Z_{j_1}, \ldots, Z_{j_v}) \). As previously, we fix \( u_n = \sqrt{n} \) so that

\[
U_n(k) = n^{-1/2} \sum_{s=1}^n V_s(k).
\]

**Corollary 4.** Assume that \( Z \) is a strictly stationary \( \theta \)-dependent process with non-increasing coefficients sequence \( (\theta(m))_{m>0} \) and such that

(i) There exists \( C > 0 \) and \( \delta \geq 0 \) such that

\[
\frac{1}{d(n)} \sum_{k=1}^{d(n)} \left( \mathbb{E}_0( |V_0(k) |^{2+\delta}) \right)^{1/(1+\delta)} < C.
\]

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

(iii) The functions \( (\Phi_k)_{1 \leq k} \) are Lipschitz.

Thus, under \( H_0 \)

\[
\tilde{N}_n \xrightarrow{c} U^2, \text{ with } U \sim \mathcal{N}(0, \sigma^2).
\]

4.4. The Gaussian case

We assume in the sequel that \( (Z_t)_{t \in \mathbb{Z}} \) is a strictly stationary process with independent coordinates \( (X_t)_{t \in \mathbb{Z}} \) and \( (Y_t)_{t \in \mathbb{Z}} \) (see Subsection 4.5 for extensions) such that under \( H_0 \),
There exists \( \alpha_X > 0, \alpha_Y > 0 \), such that \( \mathbb{E}_0(X_0 X_t) \sim t^{-\alpha_X} \) and \( \mathbb{E}_0(Y_0 Y_t) \sim t^{-\alpha_Y} \).

For all \( t \in \mathbb{Z} \), \( X_t \sim \mathcal{N}(0, 1) \).

Moreover, we assume

There exists some \( C > 0 \) such that

\[
\frac{1}{d(n)} \sum_{k=1}^{d(n)} \mathbb{E}_0(\tilde{Q}_k(X_0)^2) < C.
\]

The last condition implies that \( \mathbb{E}_0(\tilde{Q}_k(X_0)^2) < \infty \) so that \( \tilde{Q}_k \) and then \( \tilde{Q}_k^* = Q_k - \mathbb{E}_0(\tilde{Q}_k(X_0)) \) can be expanded in a basis of Hermite polynomials \((H_j)_{j \geq 0}^\prime\):

\[
\tilde{Q}_k^* = \sum_{j=m_k}^{\infty} \frac{q_j^{(k)}}{j!} H_j, \tag{13}
\]

with \( q_j^{(k)} = \mathbb{E}_0(\tilde{Q}_k^*(X_0) H_j(X_0)) \). \tag{14}

Notice that since \( \tilde{Q}_k^* \) is centered, the indice \( m_k \) of the first non-zero coefficient in (13), usually called Hermite rank, is strictly positive. Let us set \( m = \min\{m_k, 1 \leq k\} \), \( \alpha = \min\{\alpha_X, \alpha_Y\} \), then fix

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

so that

\[
U_n(k) = \begin{cases} 
    n^{-1/2} \sum_{s=1}^{n} V_s(k) & \text{if } \alpha m > 1, \\
    n^{\alpha m/2-1} \sum_{s=1}^{n} V_s(k) & \text{if } \alpha m < 1.
\end{cases}
\]

For the sake of simplicity, we assume in the sequel that \( m = m_1 \). Thus, we have

**Corollary 5.** Assume (i)-(iii).

- If \( \alpha m > 1 \), thus, under \( H_0 \)

  \[ \tilde{N}_n \overset{\mathcal{L}}{\rightarrow} U^2, \text{ with } U \sim \mathcal{N}(0, \sigma^2). \]

- If \( m = 1 \) and \( \alpha < 1 \), then, under \( H_0 \),

  \[ \tilde{N}_n \overset{\mathcal{L}}{\rightarrow} U^2 \text{ with } U \sim \mathcal{N}(0, \sigma'^2). \]

Notice that \( m_1 = m = 1 \) obtains in practical applications, in which the usual choice for \( \tilde{Q}_k \) is the Hermite function \( \tilde{Q}_k = H_k(x)g(x) \), where \( g \) is the standard Gaussian density. The generalization to \( m > 1 \) is straightforward (see the proof of Corollary 4).
4.5. Extensions

The aforementioned applications are not exhaustive.

Firstly, the results obtained for \( \alpha \)-mixing and \( \theta \)-dependent sequences may be straightforwardly extended to several other short-memory examples, satisfying other kind of mixing or weakly-dependence conditions, by adapting the assumptions of the Corollaries.

Secondly, in the Gaussian case, the independence of \( X \) and \( Y \) may be relaxed with additional technique. More precisely, when \( X \) and \( Y \) are dependent, condition (A) still holds (the proof is unchanged). In the short memory case \( \alpha m > 1 \), Assumption (B) is still obtained using [3]'s Theorem. Namely, the vector \( (U_n^{(X)}, U_n^{(Y)}) \) defined in the proof of Corollary 5 converges to a centered gaussian vector with known covariance. However, in the long memory context \( \alpha m < 1 \), [7]'s Theorem 1 can no more be applied. Its extension to the vectorial case need additional technique upon multiple Itô-Wiener integrals (see [8] and [11]).

Thirdly (see the Complementary Note), similar results hold for independent infinite moving average processes

\[
X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j} \quad \text{and} \quad Y_t = \sum_{j=0}^{\infty} d_j \epsilon_{t-j},
\]

where the \( \epsilon_i \)'s and the \( \epsilon_i \)'s are i.i.d. centered standard random variables, \( (c_j) \) and \( (d_j) \) are positive sequences such that \( c_j \sim j^{-\alpha} \) and \( d_j \sim j^{-\alpha} \) for some \( \alpha > 1/2 \).

Finally, in the Gaussian (resp. Infinite moving average) case, it is easily seen from the proof that similar results hold for processes \( Z \) such that \( X_t = l(N_t) \) and \( Y_t = l(M_t) \), where \( l \) is a measurable function, \( (N_t)_{t \in \mathbb{Z}} \) and \( (M_t)_{t \in \mathbb{Z}} \) are standard gaussian (resp. Infinite moving average) independent processes. For that task, we just need to replace \( \tilde{Q}_k \) by \( \tilde{Q}_k \circ l \) in the assumptions.

5. Practical implementation of the test

The computation of the test first requires the choice of a family \( (Q_j)_{j \geq 0} \) and a reference measure \( \nu \). Next, the nuisance parameter in limiting distribution of the test statistic has to be estimated (see Table 1). We discuss below how to deal with these choices. Moreover, we propose slight modifications in the definition of the test statistic and the information rule allowing to improve in some cases the performances of the test.

5.1. Choosing the orthogonal basis

For given dependence structure of \( Z \), the choice of \( (Q_j)_{j \geq 0} \) and \( \nu \) has to comply with the assumptions of the related corollary. In practice, this choice depends on the support of the distribution under the null of the studied model. Hereafter, we indicate relevant choices for Corollaries 2-4.
• When $Z$ takes values in $\mathbb{R}^2$, we can choose the standard normal distribution for $x$ with its associated Hermite polynomials $\{Q_j, j = 0, 1, 2, \cdots \}$ with first terms $Q_0 = 1$, $Q_1(x) = x$ and $Q_2(x) = x^2 - 1$. Then $\tilde{Q}_j$ is the $j$th Hermite function. We have
\[
E_0(\|V_0^{(k)}\|^{2+\delta}) = E_0(\|\tilde{Q}_k(X) - \tilde{Q}_k(Y)\|^{2+\delta}) \leq E_0(\|\tilde{Q}_k(X) - \tilde{Q}_k(Y)\|^{2+\delta}/N),
\]
for any integer $N > 2 + \delta$. Now let us observe that there exists a constant $c > 0$ such that for $s, t \geq 0$
\[
E_0(|\tilde{Q}_k(X)^s\tilde{Q}_k(Y)^t|) \leq c k^{-s-t}/12.
\] (16)
This is implied by Theorem 8.91.3 in Szegö (1939). It follows that
\[
E_0(\|V_0^{(k)}\|^{2+\delta}) \leq \left( \sum_{l=0}^{N} (\sum_{l=0}^{N} (\sum_{l=0}^{N} c k^{-N/12})^{2+\delta}/N \right)
\]
\[
= \bar{c} 2^{2+\delta} k^{-2\delta}/12
\]
for some $\bar{c} > 0$ so that Assumption (i) of Corollaries 2-4 is satisfied. Using (16) with $t = 0$, (iii) of Corollary 5 is also satisfied. Finally it remains to check the Lipschitz condition (iii) of Corollary 4, which is immediate since $\max |\tilde{Q}_k(x)| \leq ck^{-1/12}$ from (16).

• If $Z$ takes values in $(0, +\infty)^2$, we can choose the exponential distribution with mean 1 for $\nu$ with its associated Laguerre polynomials $\{Q_j, j = 0, 1, 2, \cdots \}$ with first terms $Q_0 = 1$, $Q_1(x) = 1 - x$ and $Q_2(x) = 0.5(3 - x)(1 - x) - 1$. Then $\tilde{Q}_j$ is the $j$th Laguerre function satisfying $\max |\tilde{Q}_k(x)| \leq C k^{-1/4}$, for some constant $C$, and we can adapt the previous calculus to verify the conditions of the corollaries.

• If $Z$ takes bounded values, say in $(0, 1)^2$, we can consider the uniform distribution for $\nu$ and its associated Legendre polynomials for $\{Q_j, j = 0, 1, 2, \cdots \}$. We have $|Q_k(x)| \leq \sqrt{2k + 1}$. Thus Condition (iii) of Corollary 5 is fulfilled. Moreover, $E_0(\|V_0^{(k)}\|^{2+\delta}) \leq (2\sqrt{2k + 1})^{2+\delta} < \infty$ and Condition (i) of Corollaries 2-4 are satisfied. Finally, the Lipschitz condition (iii) of Corollary 4 is clearly satisfied.

Remark 3. When $Z$ is continuous, another approach is to consider a bi-orthonormal basis of wavelets, say $\{\phi_i, \psi_{i,j}; i, j \in \mathbb{Z} \}$ (see Daubechies, 1992). Note here that the measure $\nu$ is the Lebesgue one but we could change $\phi_i, \psi_{i,j}$ into $\phi_i/h, \psi_{i,j}/h$ to keep our assumptions on $\nu$. The density expansions would take the form
\[
f_X = \sum_{i \in \mathbb{Z}} \langle f_X, \phi_i \rangle \phi_i + \sum_{i \in \mathbb{N}, j \in \mathbb{Z}} \langle f_X, \psi_{i,j} \rangle \psi_{i,j}, \quad f_Y = \sum_{i \in \mathbb{Z}} \langle f_Y, \phi_i \rangle \phi_i + \sum_{i \in \mathbb{N}, j \in \mathbb{Z}} \langle f_Y, \psi_{i,j} \rangle \psi_{i,j}.
\]
5.2. Estimating nuisance parameters

In each case studied in Section 4, the limiting normal distribution of the test statistic requires the estimation of an unknown variance parameter. In practice, a test strategy for (3) is to define the renormalized test statistic

\[ \tilde{R}_n = \frac{\tilde{N}_n}{\tilde{\sigma}^2}, \]

where \( \tilde{\sigma}^2 \) is a consistent estimator of the variance \( (\sigma^2, \sigma'^2) \). This statistic has a chi-square limiting distribution with one degree of freedom, so the limiting quantiles can be used to calibrate the test.

To estimate \( \tilde{\sigma}^2 \), classical kernel based estimators of the spectral density may be used (see e.g. [19], [2] and references therein). They take the form

\[ \tilde{\sigma}^2 = \sum_{|s|<n} K \left( \frac{s}{\ell_n} \right) \tilde{\sigma}_s^2, \tag{17} \]

with

\[ \tilde{\sigma}_s^2 = \begin{cases} \frac{1}{n} \sum_{h=1}^{n-s} V_h^{(1)} V_{h+s}^{(1)} & \text{if } 0 \leq s \leq n, \\ \frac{\tilde{\sigma}^2}{s} & \text{if } -n \leq s < 0. \end{cases} \]

In Formula (17), \( \ell_n \) is a bandwidth satisfying \( \ell_n \to +\infty \) and \( \ell_n/n \to 0 \); \( K \) is a symmetric kernel function satisfying \( K(0) = 1 \), \( |K(x)| \leq 1 \) for all \( x \in \mathbb{R} \), \( K \) is continuous at \( x = 0 \) and at almost all other points of \( \mathbb{R} \). Consistency of \( \tilde{\sigma}^2 \) has been obtained by several authors under more or less strong conditions on \( V^{(1)} \) and \( \ell_n \) (see e.g. [2]'s Theorem 1). In Sections 6 and 7, we use the quadratic spectral kernel

\[ K(x) = \frac{25}{12 \pi^2 x^2} \left( \frac{\sin 6\pi x}{6\pi x} - \cos 6\pi x \right), \]

which has been shown by [2] to possess large sample optimality properties, with an automatic bandwidth selection method proposed in this former paper.

A detailler As far as the estimation of \( \sigma'^2 \) is concerned, the quantity \( q_1^{(1)} \) appearing in (11) can be easily estimated using the stationarity of the process and \( H = 1 - \alpha/2 \) is the Hurst parameter of a Brownian motion, whose estimation is classical (see e.g. [25]).

5.3. Improving the test statistic

Our main test statistic \( \tilde{N}_n \) defined by (6) is an extension to the two-sample case of \( R_{Smod} \) proposed by [18] (Corollary 2, page 7). In practice, we have seen that the test consists in computing

\[ \tilde{R}_n = R_n(K_n), \]

where

\[ R_n(k) = (\tilde{\sigma}^2)^{-1} N_n(k), \quad K_n = \min \left\{ \arg \max_{1 \leq k \leq d(n)} (N_n(k) - k \log(n)) \right\}. \]
Another statistic defined by these authors \((N_{Sm2}, \text{Corollary 3})\) can be generalized as
\[
\tilde{M}_n = M_n(S_n),
\]
where
\[
M_n(k) = (\hat{\sigma}^2)^{-1} N_n(k), \quad S_n = \min \{ \arg \max_{1 \leq k \leq d(n)} (M_n(k) - k \log(n)) \},
\]
and it may be proved that under \(H_0\), \(S_n \overset{p}{\rightarrow} 1\) so that the limiting distribution of both \(\tilde{R}_n\) and \(\tilde{M}_n\) is the Chi-squared law with one degree of freedom.

However, it is clearly seen that for moderate sample size \(n\) the value of \(\sigma^2\) will influence the solutions of the decision rules, leading to unstable empirical levels for both tests. Namely, small values of \(\sigma^2\) lead to large values of \(S_n\) and then a higher rejection rate than the prescribed level for \(\tilde{M}_n\), while large values of the parameter lead to the same phenomenon for \(\tilde{R}_n\). Namely, small values of \(\sigma^2\) lead to large values of \(K_n\) and then a higher rejection rate.

In order to overcome this drawback we propose to use a third statistic invariant with respect to the variability of the components involved in (4). Namely, we consider the normalized statistic
\[
\tilde{G}_n = G_n(C_n),
\]
where
\[
G_n(k) = \sum_{j=1}^k \frac{U^{(j)}_n}{\hat{\sigma}_j^2}, \quad C_n = \min \{ \arg \max_{1 \leq k \leq a(n)} (G_n(k) - k \log(n)) \}, \quad (18)
\]
and \(\hat{\sigma}_j^2\) is a bounded consistent estimator of \(\sigma_j^2 = \sum_{i=1}^\infty \text{Cov}(V_0^{(j)} V_t^{(j)})\). In order to deal with instability we add a trimming, replacing \(\hat{\sigma}_j^2\) by
\[
\hat{\sigma}_{j, e_n}^2 = \sup(\hat{\sigma}_j^2, e_n)
\]
where \(e_n \rightarrow 0\), as \(n\) tends to infinity.

Let \(G_{n, e_n}(k)\) and \(C_{n, e_n}\) the quantities obtained by replacing \(\hat{\sigma}_j^2\) by \(\hat{\sigma}_{j, e_n}^2\) in (18). We have

**Proposition 1.** Assume that (A) holds. Thus, under \(H_0\), if \(d(n) = o(e_n \log(n))\) we have
\[
C_{n, e_n} \overset{p}{\rightarrow} 1.
\]

6. Simulation study

In order to evaluate the finite-sample performances of our tests, we run Monte-Carlo simulations on several sample sizes and models. The simulated models are vector autoregressive processes with Gaussian marginal distributions. So, \(Z\) is strong mixing. We consider different degrees of between and within dependence. For each model and sample size, we compute and compare the empirical levels and powers of several tests. The nominal level is fixed at \(\alpha = 5\%\). We present below the models, competitor tests and the obtained results.
6.1. Models

The simulated examples are based on the observation of a sequence of size \( n \in \{100, 200, 300\} \) of a stationary bivariate vector autoregressive process of order one \( \{(Z_t) = (X_t, Y_t), t \in \mathbb{Z}\} \) defined by:

\[
Z_t = C + \Theta Z_{t-1} + \epsilon_t, t \in \mathbb{Z}.
\]  

(19)

Here, \( C = (0, c)' \), \( \Theta \) 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 \( \Sigma = \Theta \Sigma \Theta' \) 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, \( (Z_t)_{t \in \mathbb{Z}} \) is a bivariate strictly stationary and strong mixing Gaussian process with mean vector \( \mathbb{E}(Z_t) = \Theta \mathbb{E}(Z_{t-h}) = \Theta^h \Sigma \). Namely, \( X_1 \) follows a Gaussian law with mean zero and variance 1 while \( Y_1 \) follows a Gaussian law with mean \( \mu = c/(1 - \theta_Y) \) and variance \( \sigma_Y^2 \). Within sample dependence is controlled by \( \theta_X \) and \( \theta_Y \) and increases with their absolute values. In order to maintain a fixed level of within dependence between the null and the alternative, we chose \( \theta^* = (\theta_X, \theta_Y^*) \in \{(0, 0), (0.3, 0.3), (0.6, 0.6)\} \). The case \( \theta^* = (0, 0) \) 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 \( \Sigma \) and increases with its absolute value. We chose \( v \in \{-0.5, 0, 0.5\} \). The case \( v = 0 \) corresponds to independent sequences. Finally, the values of \( c \) and \( \sigma_Y \) control the marginal distribution of \( Y_1 \). The null hypothesis corresponds to the case \( (c, \sigma_Y) = (0, 1) \). Hereafter, we only display powers against a scale alteration of the null hypothesis, corresponding to \( (c, \sigma_Y) = (0, 0.5) \).

6.2. Tests

We first considered the three tests discussed in Subsection 5.3. We denote \( G, R \) and \( M \) the tests based on \( \tilde{G}_n, \tilde{R}_n \) and \( \tilde{M}_n \) respectively. Moreover, we computed the test procedure described in [10], that we call \( T \). This test is designed to compare paired independent samples and has been shown to achieve good performances with respect to classical tests such as Wilcoxon signed rank test on a wide range of marginal distributions including those studied here. It allows to observe what is going on if we omit to take into account the within-sample dependence. The computation of the test statistics first requires the choice of \( d(n) \). A previous study showed that the empirical levels and powers obtained do not depend on \( d(n) \) for sufficiently large values of this parameter. In practice, \( d(n) \) were set at 10. Secondly, according to the support \( \mathbb{R} \) of the process considered in our simulation study, we used here the standard Gaussian distribution and its associated Hermite polynomials. The unknown variance parameter where estimated as indicated in Section 5.
6.3. Empirical levels

Setting $\sigma_Y = 1$ and $c = 0$, the data were drawn from processes having standard Gaussian marginal distributions $f_X$ and $f_Y$ with respect to Lebesgues measure and we investigated the empirical level of the tests for $n \in \{50, 100, 200, 300\}$. Thus, for each value of $n$, $\theta^*$ and $\nu$, we computed the test statistic based on the sample and compared it to the 5%-critical value of its approximated distribution under $H_0$. The empirical level of the test were defined as the percentage of rejection of the null hypothesis over 10000 replications of the test statistic.

For small sample size ($n \leq 50$) we observed empirical levels much larger than the theoretical one. The reason is clearly due to the fact that all penalizing statistics $K_n, C_n, S_n$ are equal or greater than 1. This problem is discussed in [17] who proposed a second order approximation of the statistic’s distribution. However, in the dependent case this correction cannot be done since such an approximation depends on the within dependence structure of $Z$. In addition, as underlined by [18], the dependent case requires larger data sets as for the independent case, and hence small sample corrections are often not of major interest.

So, we finally restricted our attention to sample sizes $n = 100, 200, 300$. Empirical levels are reported in Table 2. It can be observed that in most cases, $G$ overestimates the nominal level. However, from the whole experience this overestimation is limited and stable with respect to the degree of both within and between dependence, which is not the case for all other competitors: first it can be observed that the empirical level of $T$ explodes for large values of $\theta^*$, pointing out that $T$ is clearly not adapted to within dependent situations. Concerning $M$ and $R$, we recover the importance played by $\hat{\sigma}$ in the selection of the numbers of components $K_n$. In our simulation, $\hat{\sigma}$ is largely less than 1 (around $10^{-2}$). Small values of $\hat{\sigma}$ lead to a less concentration of $S_n$ around dimension 1 so that $M$ tends to overestimate the level. As for $R$, the convergence to the asymptotic null distribution can be very slow, and a very large sample size may be necessary to detect alternative hypotheses as we will see in our power study (for an explanation of this phenomenon, see [18] p. 12). We should have assist to the inverse phenomenon if $\hat{\sigma}$ had been greater than one, leading to slow convergence for $M$.

In conclusion, despite its empirical level slightly too high, statistic $G$ offers a greater stability than the other statistics.

6.4. Empirical powers

The empirical powers are given in Table 3. We did not display the values corresponding to largest empirical levels. It is shown that $G$ detects reasonably a scale change in the margin and is equivalent, or sometimes better than $M$ in cases of within-dependence. A general remark here is that $G$ is slightly better when $\nu \neq 0$; that is, when there exists a dependence between sequences. Concerning $R$ it can be seen that the convergence is very slow and alternatives are not detected. Finally, for $M$, it seems that alternatives are better detected for small values of $\theta^*$ than for large ones.
Table 2: Empirical levels (in %) for $T, G, M$ and $R$. Within-sample (resp. cross-) dependency is measured by $\theta^*$ (resp. $v$).

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>n</th>
<th>100</th>
<th>200</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta^*$</td>
<td>v-0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>(-0.6,-0.6)</td>
<td>T</td>
<td>7.4</td>
<td>7.6</td>
<td>7.8</td>
</tr>
<tr>
<td>(-0.6,-0.6)</td>
<td>G</td>
<td>7.3</td>
<td>7.0</td>
<td>8.1</td>
</tr>
<tr>
<td>(-0.6,-0.6)</td>
<td>M</td>
<td>14.3</td>
<td>21.9</td>
<td>25.2</td>
</tr>
<tr>
<td>(-0.6,-0.6)</td>
<td>R</td>
<td>3.7</td>
<td>3.5</td>
<td>4.1</td>
</tr>
<tr>
<td>(-0.3,-0.3)</td>
<td>T</td>
<td>3.8</td>
<td>3.8</td>
<td>4.1</td>
</tr>
<tr>
<td>(-0.3,-0.3)</td>
<td>G</td>
<td>7.0</td>
<td>7.1</td>
<td>7.2</td>
</tr>
<tr>
<td>(-0.3,-0.3)</td>
<td>M</td>
<td>5.0</td>
<td>7.6</td>
<td>9.9</td>
</tr>
<tr>
<td>(-0.3,-0.3)</td>
<td>R</td>
<td>4.1</td>
<td>3.9</td>
<td>4.0</td>
</tr>
<tr>
<td>(0,0)</td>
<td>T</td>
<td>4.6</td>
<td>4.5</td>
<td>4.6</td>
</tr>
<tr>
<td>(0,0)</td>
<td>G</td>
<td>7.4</td>
<td>7.3</td>
<td>7.8</td>
</tr>
<tr>
<td>(0,0)</td>
<td>M</td>
<td>4.6</td>
<td>5.7</td>
<td>7.4</td>
</tr>
<tr>
<td>(0,0)</td>
<td>R</td>
<td>4.5</td>
<td>4.6</td>
<td>4.8</td>
</tr>
<tr>
<td>(0.3,0.3)</td>
<td>T</td>
<td>10.1</td>
<td>10.2</td>
<td>9.0</td>
</tr>
<tr>
<td>(0.3,0.3)</td>
<td>G</td>
<td>8.0</td>
<td>9.1</td>
<td>9.2</td>
</tr>
<tr>
<td>(0.3,0.3)</td>
<td>M</td>
<td>5.6</td>
<td>6.5</td>
<td>7.7</td>
</tr>
<tr>
<td>(0.3,0.3)</td>
<td>R</td>
<td>5.6</td>
<td>6.0</td>
<td>6.4</td>
</tr>
<tr>
<td>(0.6,0.6)</td>
<td>T</td>
<td>26.3</td>
<td>26.1</td>
<td>24.6</td>
</tr>
<tr>
<td>(0.6,0.6)</td>
<td>G</td>
<td>9.1</td>
<td>9.4</td>
<td>9.3</td>
</tr>
<tr>
<td>(0.6,0.6)</td>
<td>M</td>
<td>6.7</td>
<td>6.9</td>
<td>8.8</td>
</tr>
<tr>
<td>(0.6,0.6)</td>
<td>R</td>
<td>6.7</td>
<td>6.5</td>
<td>7.2</td>
</tr>
</tbody>
</table>
Table 3: Empirical powers (in %) for $T, G, R$ and $M$ in the case $c = 0$ and $\sigma^2_Y = 0.5$

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>n</th>
<th>100</th>
<th>200</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta^*$</td>
<td>$v$ -0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>(-0.6,-0.6)</td>
<td>T</td>
<td>84.6</td>
<td>65.0</td>
<td>85.6</td>
</tr>
<tr>
<td>(-0.6,-0.6)</td>
<td>G</td>
<td>72.8</td>
<td>48.0</td>
<td>72.6</td>
</tr>
<tr>
<td>(-0.6,-0.6)</td>
<td>M</td>
<td>3.5</td>
<td>3.5</td>
<td>3.7</td>
</tr>
<tr>
<td>(-0.3,-0.3)</td>
<td>T</td>
<td>90.8</td>
<td>68.6</td>
<td>90.6</td>
</tr>
<tr>
<td>(-0.3,-0.3)</td>
<td>G</td>
<td>89.4</td>
<td>67.3</td>
<td>88.7</td>
</tr>
<tr>
<td>(-0.3,-0.3)</td>
<td>M</td>
<td>52.6</td>
<td>69.1</td>
<td>94.3</td>
</tr>
<tr>
<td>(-0.3,-0.3)</td>
<td>R</td>
<td>3.8</td>
<td>3.9</td>
<td>3.7</td>
</tr>
<tr>
<td>(0.0)</td>
<td>T</td>
<td>92.5</td>
<td>72.3</td>
<td>91.7</td>
</tr>
<tr>
<td>(0.0)</td>
<td>G</td>
<td>92.1</td>
<td>73.2</td>
<td>91.5</td>
</tr>
<tr>
<td>(0.0)</td>
<td>M</td>
<td>27.9</td>
<td>56.8</td>
<td>91.7</td>
</tr>
<tr>
<td>(0.0)</td>
<td>R</td>
<td>5.1</td>
<td>4.2</td>
<td>4.3</td>
</tr>
<tr>
<td>(0.3,0.3)</td>
<td>T</td>
<td>88.2</td>
<td>67.2</td>
<td>88.3</td>
</tr>
<tr>
<td>(0.3,0.3)</td>
<td>G</td>
<td>12.2</td>
<td>38.5</td>
<td>84.8</td>
</tr>
<tr>
<td>(0.3,0.3)</td>
<td>M</td>
<td>5.9</td>
<td>6.1</td>
<td>6.4</td>
</tr>
<tr>
<td>(0.3,0.3)</td>
<td>R</td>
<td>5.9</td>
<td>6.1</td>
<td>6.4</td>
</tr>
<tr>
<td>(0.6,0.6)</td>
<td>T</td>
<td>66.2</td>
<td>44.9</td>
<td>70.2</td>
</tr>
<tr>
<td>(0.6,0.6)</td>
<td>G</td>
<td>7.3</td>
<td>20.3</td>
<td>62.8</td>
</tr>
<tr>
<td>(0.6,0.6)</td>
<td>M</td>
<td>6.5</td>
<td>6.2</td>
<td>7.2</td>
</tr>
</tbody>
</table>
7. Financial data

In this real study we analyze several financial assets of the US economic sectors. Data are available from Bloomberg and consists in monthly rates from January 2004 to August 2013. We consider three indices: the Dow Jones Composite Average, the NASDAQ Composite and the NYSE International 100 Index. The first one is a stock index that tracks 65 prominent companies, the NASDAQ is an indicator of the performance of stocks of technology companies and growth 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 will compare the marginal distributions of these three increments processes.

7.1. NASDAQ versus Dow Jones

The three test statistic $T$, $G$, $M$ concluded to reject the equality of the two increments processes, with p-values less than $10^{-12}$, $10^{-4}$ and $10^{-5}$, and with values of $S_n$ equal to 6, 2 and 2, respectively. This means that at least a second-order difference was detected by $G$ and $M$. $R$ is the one that does not reject the equality of the margins with a p-value around 10% and $S_n = 1$. This is due to the very small convergence of its distribution under alternatives. Following the test statistic $G$ we can reject the equality of the variations between NASDAQ and Dow Jones.

7.2. NASDAQ versus NYSE

The three test statistic $T$, $G$, $M$ concluded to the rejection of the equality of the two increments processes, with all p-values less than $10^{-7}$. Moreover $S_n$ was equal to 4, 4 and 2 for $T$, $G$ and $R$, respectively. The p-value associated to $R$ was equal to 0.048 with $S_n = 1$. Following our test statistic $G$ we reject the equality between the two increments processes.

7.3. NYSE versus Dow Jones

All statistics have almost the same p-value around 0.75, with $S_n = 1$. In conclusion, concerning NYSE and Dow Jones indexes from 2004 to 2013, their increments can be considered as identically distributed.

8. Discussion

In this paper we extend the two-sample problem to different within dependence structures, covering short and long memory cases. A general result is proposed. The implementation of the test is simple, it only requires the choice of a polynomial basis. The nuisance parameter, that is the variance of the asymptotic statistic distribution, can be estimated by standard techniques. The test can be applied in practical situations, as illustrated in our data study. Concerning the choice of the selection rule, some refinements could be envisaged by the use of an Akaike’s like criterion, as proposed in [14]. Among the proposed
versions of our main test statistic, it seems that the best choice is $\tilde{G}_n$, that takes into account the variability of all components, making the test more robust with respect to the value of the variance parameter.

In theory the test is consistent against contiguous alternatives. Simulation study showed good performances of the test statistic $G$ for the weak dependence. It also pointed out the instability of the other statistics. These simulations show that the test can be implemented without difficulty. There would still be other numerical studies to achieve, but this aspect is beyond the scope of this paper. Finally, an extension to the K-sample case could be envisaged, following the approach of [27]. Thus, this extension would allow us to compare our three financial data sets simultaneously. Another issue that could be considered would be that of stable processes, which is not studied in this paper. Then a two-sample test would have many applications, as for instance for extreme values.

9. Proofs

9.1. Proofs of Section 2

9.1.1. Proof of Theorem 1

In the following, we denote by $P_0$ the probability under $H_0$. First notice that it is enough to prove (8). Then applying (B), (9) obtains. Let us show that $P_0(K_n \geq 2)$ tends to zero as $n$ tends to infinity. By definition of $K_n$ and Markov’s inequality, we have for all $p > 0$

$$P_0(K_n \geq 2) = P_0 \left( \max_{2 \leq k \leq d(n)} \{N_n(k) - k \log(n)\} \geq N_n(1) - \log(n) \right)$$

$$= P_0 \left( \exists k, 2 \leq k \leq d(n), N_n(k) - k \log(n) \geq N_n(1) - \log(n) \right)$$

$$= P_0 \left( \exists k, 2 \leq k \leq d(n), \sum_{j=2}^{k} (U_n^{(j)})^2 \geq (k - 1) \log(n) \right)$$

$$\leq P_0 \left( \exists k, 2 \leq k \leq d(n), (U_n^{(k)})^2 \geq \log(n) \right)$$

$$\leq \sum_{k=2}^{d(n)} P_0 \left( (U_n^{(k)})^2 \geq \log(n) \right)$$

$$\leq \frac{1}{\log(n)} \sum_{k=2}^{d(n)} E_0 \left( (U_n^{(k)})^2 \right).$$
Notice that since \((Z_t)_{t \in \mathbb{Z}}\) is strictly stationary, the same happens for \((V^{(k)}_t)_{t \in \mathbb{Z}}\). Therefore, classical calculations using the stationarity of the process yield

\[
E_0 \left( (U_n^{(k)})^2 \right) = \frac{1}{u_n} \sum_{s=1}^{n} \sum_{t=1}^{n} E_0 (V^{(k)}_s V^{(k)}_t) = \frac{n}{u_n^2} \sum_{|t| < n} (1 - \frac{|t|}{n}) E_0 \left( V^{(k)}_0 V^{(k)}_t \right) \leq \frac{2n}{u_n^2} r_n(k),
\]

so that

\[
P_0(K_n \geq 2) \leq C o(1),
\]

by (A) and (8) of Theorem 1 obtains.

9.1.2. Proof of Theorem 2

Under \(H^*_1\), following the proof of Theorem 1, then using standard power inequalities, we have

\[
P_1(K_n \geq 2) \leq \frac{1}{\log(n)} \sum_{k=2}^{d(n)} E_1 |U_1^{(k)}|^2 \leq \frac{2d(n)}{\log(n)} \left( C^* + \frac{n \beta^*}{u_n} |\delta_k|^2 \right).
\]

Therefore, \(P_1(K_n \geq 2) = o(1)\) by Assumption (A*).

Under \(H^*_1\), one has with similar arguments \(P_1(K_n > K) = o(1)\).

On the other hand,

\[
P_1(K_n < K) = P_1 (\exists k < K, N_n(k) - k \log(n) \geq N_n(K) - K \log(n)) \leq \sum_{k=1}^{K-1} P_1 \left( \sum_{j=k}^{K} (U_n^{(j)})^2 \leq (K - k) \log(n) \right)
\]

\[
\leq K P_1 \left( \frac{U_n^{(K)}}{\sqrt{\log(n)}} \leq \sqrt{K} \right).
\]

Moreover, one has

\[
\frac{U_n^{(K)}}{\sqrt{\log(n)}} = \frac{1}{\sqrt{\log(n)}} (U_n^{(K)} - \frac{n \delta_K}{u_n} + \frac{n \delta_K}{u_n \sqrt{\log(n)}}),
\]

so that \(\frac{U_n^{(K)}}{\sqrt{\log(n)}}\) tends to infinity as \(n\) tends to infinity.

It follows that \(K_n \to K\) and that the test statistic \(\tilde{N}_n\) converges to infinity as soon as (B*) holds.
9.2. Proofs of Section 4

9.2.1. Proof of Corollary 2

Let us first notice that for given $k > 0$, the process $(V_t^{(k)})_{t \in \mathbb{Z}}$ is $m$-dependent too. Therefore, $\mathbb{E}_0(V_0^{(k)}V_s^{(k)}) = 0$ for all $s > m$.

- With $r_n(k)$ defined by (7), we have by Cauchy-Schwarz’s inequality
  \[ r_n(k) \leq \sum_{j=-m}^{m} |\mathbb{E}_0(V_0^{(k)}V_j^{(k)})| \leq K\mathbb{E}_0(V_0^{(k)})^2, \]
  with $K = 1$ if $m = 0$ and $K = 2m + 1$ otherwise, so that (A) holds with $u_n = \sqrt{n}$.

- The fact that (B) holds straightforwardly arises from [12]'s central limit theorem for $m$-dependent sequences with $U_a$ centered gaussian variable with variance $\sigma^2$. The convergence of the renormalized statistic is obtained by Slutsky’s theorem.

9.2.2. Proof of Corollaries 3 and 4

In the sequel, $\|\|_q$ denotes the $L_q$ norm and we define $Q_k$ from $[0, 1]$ into $\mathbb{R}^+$ as the quantile function of $|V_0^{(k)}|$, that is the inverse function of $x \rightarrow P_0(|V_0^{(k)}| > x)$. Moreover, we define $G_k$ from $\mathbb{R}^+$ into $[0, 1]$ as the inverse function of $x \rightarrow \int_0^x Q_k(t)dt$. Finally, for any sequence $(\delta_i)$ of nonnegative numbers, set
  \[ \delta^{-1}(u) = \sum_{i \geq 0} 1_{u < \delta_i}. \]

- We first check (A) in both cases.

  - In the strong mixing case, let us notice that $(V_t^{(k)})$ is strong mixing too since it can be expressed as a measurable function of $Z$. Moreover, denoting by $\alpha_k(r)$ its mixing coefficients sequence, one has $\alpha_k(r) \leq \alpha(r)$. Starting from [22]'s covariance inequality given in Theorem 1.1, one obtains as in (1.23) p 13 of this book,
    \[ r_n(k) \leq 4 \int_0^1 (\alpha^{-1}(u) \wedge n)Q_k^2(u)du, \]
    so that $r_n(k)$ converges with
    \[ \lim_{n \to \infty} r_n(k) \leq 4 \int_0^1 \alpha^{-1}(u)Q_k^2(u)du. \]
    The right hand side of the above equation can be upper bounded following calculations p.15 of [22]. Namely, we obtain the analogue of bound (1.25a) p.16 of this paper:
    \[ \lim_{n \to \infty} r_n(k) \leq K||V_0^{(k)}||_{2+\delta}^2 \left( \sum_{t \geq 0} (t + 1)^{2/\delta} \alpha(t) \right)^{\delta/(2+\delta)}, \]
    \[ \text{(21)} \]
with $K = 4 \exp(2/(2 + \delta))$, so that (A) obtains by (i)-(ii), with $u_n = \sqrt{n}$.

- In the $\theta$-dependent case, first note that (iii) implies the $\theta$-dependence of $(V_t^{(k)})$, with coefficients sequence $\theta_k(r) \leq \theta(r)$. Starting from the covariance inequality of [4]'s Proposition 1, we obtain a bound as in inequality (5.2) in [4]'s Proposition 2. Namely,

$$r_n(k) \leq 2 \int_0^\|V_0^{(k)}\|_1 (\theta/2)^{-1} (u \wedge n) Q_k \circ G_k(u) du,$$

so that

$$\lim_{n \to \infty} r_n(k) \leq 2 \int_0^\|V_0^{(k)}\|_1 (\theta/2)^{-1} (u) Q_k \circ G_k(u) du.$$

Next, following the proof of [4]'s Lemma 2, we finally get

$$\lim_{n \to \infty} r_n(k) \leq K \|V_0^{(k)}\|^{2+\delta}/(1+\delta) \left( \sum_{t \geq 0} (t + 1)^{1/4} \theta(t) \right)^{\delta/(1+\delta)}, \quad (22)$$

with $K = 2^{(4+\delta)/(2+\delta)}$, so that (A) obtains by (i)-(ii), with $p = 2$ and $u_n = \sqrt{n}$.

- Under conditions (i) and (ii) and since $\sigma^2 \neq 0$, Assumption (B) straightforwardly follows from central limit theorems stated in [6]'s Corollary 1 for the strong mixing case and [4]'s Theorem 2 (after application of Lemma 2 and Corollary 1 of this paper) in the $\theta$-dependent case. Namely, the limit of $U_n(1)$ under $H_0$ is a centered gaussian variable with variance $\sigma^2$.

9.2.3. Proof of Corollary 5

Set

$$S_n^{(k)} = \sum_{s=1}^n V_s^{(k)} = u_n U_n^{(k)} = \sum_{s=1}^n (\bar{Q}_k^*(X_s) - \bar{Q}_k^*(Y_s)).$$

- Let us first prove (A). One has

$$\mathbb{E}_0 \left( S_n^{(k)} \right)^2 \leq 2 \left( \mathbb{E}_0 \left( \sum_{s=1}^n \bar{Q}_k^*(X_s) \right)^2 + \mathbb{E}_0 \left( \sum_{s=1}^n \bar{Q}_k^*(Y_s) \right)^2 \right). \quad (23)$$

Let us bound one of the two terms at the right hand side of (23). Using the stationarity of $\bar{Q}_k(X_s)$ yields

$$\mathbb{E}_0 \left( \sum_{s=1}^n \bar{Q}_k^*(X_s) \right)^2 = \sum_{|s| < n} (n - |s|) \mathbb{E}_0(\bar{Q}_k^*(X_0)\bar{Q}_k^*(X_s)). \quad (24)$$
Let $H_l$ be the $l$th Hermite’s polynomial with recurrence formula $xH_j(x) = H_{j+1}(x) + jH_{j-1}(x)$. By (ii), $\widetilde{Q}_k^*$ can be expanded as

$$\widetilde{Q}_k^*(x) = \sum_{l=0}^{\infty} \frac{q_l^{(k)}}{l!} H_l(x) = \sum_{l=m_k}^{\infty} \frac{q_l^{(k)}}{l!} H_l(x),$$

with $q_l^{(k)} = \mathbb{E}(\tilde{Q}_k^*(X_0)H_l(X_0))$. Notice that since $\tilde{Q}_k^*(X_0)$ is centered, we necessarily have $m_k \geq 1$. Therefore,

$$\mathbb{E}_0(\tilde{Q}_k^*(X_0)\tilde{Q}_k^*(X_s)) = \mathbb{E}_0 \left( \sum_{l=m_k}^{\infty} \sum_{r=m_k}^{\infty} \frac{q_l^{(k)}}{l!} \frac{q_r^{(k)}}{r!} H_l(X_0)H_r(X_s) \right),$$

$$= \sum_{l=m_k}^{\infty} \frac{(q_l^{(k)})^2}{l!} (\mathbb{E}_0(X_0X_s))^l. \tag{25}$$

The last equality arises from Mehler’s formula. Using the same scheme for the other term at the right hand side of (23), we get

$$\mathbb{E}_0 \left( S_n^{(k)} \right)^2 \leq 2n \sum_{l=m_k}^{\infty} \frac{(q_l^{(k)})^2}{l!} \sum_{|s| \leq n} \left( 1 - \frac{|s|}{n} \right) \left( |\mathbb{E}_0(X_0X_s)|^l + |\mathbb{E}_0(Y_0Y_s)|^l \right). \tag{26}$$

Since $X$ and $Y$ are standard processes, we have $|\mathbb{E}_0(X_0X_s)| \leq 1$ and $|\mathbb{E}_0(Y_0Y_s)| \leq 1$ so that $|\mathbb{E}_0(X_0X_s)|^l \leq |\mathbb{E}_0(X_0X_s)|^{m_k} \leq |\mathbb{E}_0(X_0X_s)|^{m}$ for all $l \geq m_k$. Next, we have

$$\sum_{l=m_k}^{\infty} \frac{(q_l^{(k)})^2}{l!} = \mathbb{E}_0(\tilde{Q}_k^*(X_0))^2 \leq E_0(\tilde{Q}_k^*(X_0))^2.$$

Therefore, by Assumption (i),

$$\mathbb{E}_0 \left( S_n^{(k)} \right)^2 \leq 4n\mathbb{E}_0(\tilde{Q}_k^*(X_0))^2 \sum_{s=0}^{n-1} (|\mathbb{E}_0(X_0X_s)|^m + |\mathbb{E}_0(Y_0Y_s)|^m),$$

$$\leq 8An\mathbb{E}_0(\tilde{Q}_k^*(X_0))^2 \sum_{s=0}^{n-1} s^{-am}, \tag{27}$$

with $A = \max(A_X, A_Y)$. Therefore, when $am > 1$, there exists $K > 0$ such that $\mathbb{E}_0(\tilde{S}_n^{(k)})^2 \leq Kn\mathbb{E}_0(\tilde{Q}_k^*(X_0))^2$ so that (A) obtains with $u_n = \sqrt{n}$ by (ii) and (iii). When $am < 1$, there exists some $K'$ such that $\mathbb{E}_0(\tilde{S}_n^{(k)})^2 \leq K'n^{2-am}$ so that (A) obtains with $u_n = n^{1-am/2}$ by (ii) and (iii).

- The proof of (B) when $m = m_1$ straightforwardly follows from known results:
– When $\alpha m > 1$, we have by (i)

$$\sum_{t \in \mathbb{Z}} |E_0(X_0X_t)|^m < \infty,$$

and since $E_0(\tilde{Q}^*_1(X_0))^2$ exists by (iii), [3]'s theorem for stationary vectors yields

$$U_n^{(X)} = \frac{1}{\sqrt{n}} \sum_{s=1}^{n} \tilde{Q}^*_1(X_s) \xrightarrow{\mathcal{L}} U_X,$$

similarly,

$$U_n^{(Y)} = \frac{1}{\sqrt{n}} \sum_{s=1}^{n} \tilde{Q}^*_1(Y_s) \xrightarrow{\mathcal{L}} U_Y,$$

where $U_X$ and $U_Y$ are independent centered Gaussian random variables with respective variances

$$\sigma^2_X = \sum_{s \in \mathbb{Z}} E_0(\tilde{Q}^*_1(X_0)\tilde{Q}^*_1(X_s)),$$

and $\sigma^2_Y$ defined in the same way. Since processes $X$ and $Y$ are independent, so are $U_n^{(X)}$ and $U_n^{(Y)}$ and $(U_n^{(X)}, U_n^{(Y)})$ converges in distribution to $(U_X, U_Y)$. Thus,

$$U_n^{(1)} = U_n^{(X)} - U_n^{(Y)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma), \quad (28)$$

with $\sigma^2 = \sigma^2_X + \sigma^2_Y$.

– When $\alpha m < 1$, (i) and (iii) implies that [7]'s Theorem 1 holds. Namely, one has:

$$U_n^{(X)} = n^{m\alpha/2-1} \sum_{j=0}^{n-1} \tilde{Q}^*_1(X_s) \xrightarrow{\mathcal{L}} K_{m,\alpha}^{(1)} \frac{q_{m}}{m!} Z_X,$$

where $Z_X$ has the same distribution as an Hermite process $\{U_{m,1-\alpha/2}(t), t \in \mathbb{R}\}$ of order $m$ and parameter $1 - \alpha m/2$ at point 1; that is,

$$U_{m,1-\alpha/2}(t) = D^{-m/2} K_{m,\alpha}^{-1} \int_{\mathbb{R}^m} \frac{e^{i t (x_1 + \ldots + x_m)} - 1}{i (x_1 + \ldots + x_m)} |x_1|^{\alpha - 1} \ldots |x_m|^{\alpha - 1} dW(x_1) \ldots dW(x_m),$$

where $D = 2\Gamma(\alpha) \cos(\frac{\alpha\pi}{2})$,

$$K_{m,\alpha} = \frac{m!}{(1-\alpha m)(1-\alpha m/2)}, \quad (29)$$

and $W$ is the spectral measure of the white noise process. When $m = 1$, the process $U_{1,1-\alpha/2}(t)$ is the fractional brownian motion.
with Hurst parameter $1 - \alpha/2$, so that $Z_X$ is a standard Gaussian variable. The same result holds for $u_n(Y)$. So, applying the same scheme as for (28), one has that

$$U_n(1) \xrightarrow{L} U, \text{ with } U = q_1^{(1)} K_{1,\alpha} (Z_X - Z_Y),$$

where $Z_X$ and $Z_Y$ are two independent standard Gaussian variables. This gives the result.

### 9.2.4. Proof of Proposition 1

Following the proof of Theorem 1 we get

$$\mathbb{P}_0(C_{n,e_n} \geq 2) \leq \frac{1}{\log(n) \max(\hat{\sigma}^2_{e_n})} \sum_{k=2}^{d(n)} \mathbb{E}_0 \left( |U_n^{(k)}|^2 \right)$$

$$\leq \frac{1}{\log(n)e_n} \sum_{k=2}^{d(n)} \mathbb{E}_0 \left( |U_n^{(k)}|^2 \right).$$


