



UNIVERSIDAD CARLOS III DE MADRID

working
papers

Working Paper 05-07
Statistics and Econometrics Series 02
February 2005

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

ON THE COMPARISON OF TIME SERIES USING SUBSAMPLING

Andrés M. Alonso and Elizabeth A. Maharaj*

Abstract

In this paper we propose a procedure based on the subsampling techniques for the comparison of stationary time series that are not necessarily independent. We study a test based on the Euclidean distance between the autocorrelation functions of two series. Consistency of the proposed method is established. We present a Monte Carlo study with the size and the power of the proposed test.

Keywords: stationary time series; comparison; resampling.

*Alonso: Department of Statistics, Universidad Carlos III de Madrid, Spain, E-mail: andres.alonso@uc3m.es; Maharaj: Department of Econometrics and Business Statistics, Monash University, Australia, E-mail: ann.maharaj@buseco.monash.edu.au. This work was partially supported by the Fundación BBVA project 2233 and by the Comunidad de Madrid project 06/0050/2003.

1 Introduction

The comparison of two or more time series is a problem of great interest in many practical situations: (*i*) in geology, for example, it is interesting to detect the differences between the waves produced by earthquakes and by mining explosions; (*ii*) in medicine, the comparison of different sections of a biomedical signal is used as a diagnostic procedure; (*iii*) in economics, it is interesting to compare the interest rates or the inflation rates in different regions or countries.

In this paper, we propose a procedure based on subsampling for testing the equality of the generating processes of two stationary time series that are not necessarily independent. The proposed procedure is different from the methods of Basawa et al. (1984), Maharaj (1996), Guo (1999) and Maharaj (2000) since it does not require the selection and the estimation of models. Also, it is different from the methods of Coates and Diggle (1986), Swanepoel and Van Wyk (1986) and Timmer et al. (1999) since it does not require spectral estimation. With the exception of Maharaj (2000), the above-mentioned methods are only applicable to independent series. In Section 3, we will see that the procedure based on subsampling is valid for both independent and dependent series and it is free of model or spectral estimation.

The paper is organized as follows. Section 2 presents the general approach for hypothesis testing using subsampling (as in Politis et al. (1999)). In Section 3 we develop the procedure of testing the equality of generating processes. Section 4 is devoted to the consistency study of the proposed method. Finally, in Section 5, we includes the results of a Monte Carlo study of the properties of the test.

2 Subsampling method

Politis and Romano (1994) introduce the subsampling method in homogeneous stochastic fields that includes, as a particular case, the strictly stationary series. For simplicity, we describe

the method in the stationary series case:

Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a vector of observations from a stationary series following certain probabilistic model P . The null hypothesis, H_0 , establishes that $P \in \mathbf{P}_0$ while the alternative hypothesis, H_1 , is $P \in \mathbf{P}_1$, where \mathbf{P}_0 and \mathbf{P}_1 are disjoint subclasses of \mathbf{P} such that $\mathbf{P}_0 \cup \mathbf{P}_1 = \mathbf{P}$. Let $\mathbf{X}_j = (X_j, X_{j+1}, \dots, X_{j+l-1})$ be blocks or subsamples of l consecutive observations with $j = 1, 2, \dots, n-l+1$. Let $T_n = \tau_n t_n(X_1, X_2, \dots, X_n)$ be the test statistic where τ_n is a nonrandom normalizing sequence, then we can define the j -th subsampling statistic, $T_l^{(j)} = \tau_l t_l(\mathbf{X}_j)$, as the test statistic T evaluated at the subsample \mathbf{X}_j . The subsampling estimator of the distribution of T_n is:

$$\widehat{G}_{n,l}(x) = \frac{1}{n-l+1} \sum_{j=1}^{n-l+1} I \left\{ T_l^{(j)} \leq x \right\}, \quad (1)$$

where $I\{E\}$ denotes the indicator of the event E . Given the estimated sampling distribution, the critical value the test is obtained as the $1 - \alpha$ quantile of $\widehat{G}_{n,l}(\cdot)$ defined by:

$$g_{n,l}(1 - \alpha) = \inf \left\{ x : \widehat{G}_{n,l}(x) \geq 1 - \alpha \right\}. \quad (2)$$

Finally, we reject the null hypothesis with a nominal level α if and only if $T_n > g_{n,l}(1 - \alpha)$.

The main hypothesis in subsampling is that $G_n(\cdot, P)$ converges weakly to a limit distribution, $G(\cdot, P)$. The Theorem 3.5.1 of Politis et al. (1999) establishes that $g_{n,l}(1 - \alpha)$ converges in probability to the $1 - \alpha$ quantile of $G(\cdot, P)$ for $P \in \mathbf{P}_0$ assuming that $\{X_t\}_{t \in \mathbb{Z}}$ is α -mixing with some restrictions on the normalizing sequences and the subsamples size; specifically $\tau_l/\tau_n \rightarrow 0$, $l \rightarrow \infty$ and $l/n \rightarrow 0$.

3 Hypothesis testing procedure

Let $\{X_t\}_{t \in \mathbb{Z}}$ and $\{Y_t\}_{t \in \mathbb{Z}}$ be two strictly stationary processes that follow the models P_X and P_Y , respectively. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ be vectors of observations from $\{X_t\}_{t \in \mathbb{Z}}$ and $\{Y_t\}_{t \in \mathbb{Z}}$, respectively. We are interesting on the following hypothesis test:

$$\begin{aligned} H_0 : P_X &= P_Y \\ H_1 : P_X &\neq P_Y \end{aligned} \quad (3)$$

i.e., we want to test if the generating process is the same in both series. It is known that the complete probabilistic structure of a stochastic process $\{Z_t\}_{t \in \mathbb{Z}}$ is determined by the set of all finite collections Z_t 's. In linear processes (and particularly, in Gaussian linear processes) much of the information in these joint distributions can be described in terms of the process mean $\mu_Z = E[Z_t]$ and the autocovariance function, $\gamma_{Z,k} = \text{Cov}(Z_t, Z_{t+|k|})$ where $k \geq 0$. In this paper, we consider the test (3) using a statistic defined as function of the estimated autocorrelations calculated on the samples \mathbf{X} and \mathbf{Y} under the (simplifying) assumption that $\mu_X = \mu_Y$.

The proposed test statistic is the following:

$$T_{n,m} = n \sum_{k=1}^m (\hat{\rho}_{X,k} - \hat{\rho}_{Y,k})^2, \quad (4)$$

where $\hat{\rho}_{X,k}$, $\hat{\rho}_{Y,k}$ are the k -th estimated autocorrelations using \mathbf{X} and \mathbf{Y} , respectively, n is the sample size and m is the maximum considered lag. In Appendix we provide the theoretical basis for the subsampling method using the statistics (4). In this case the normalizing constant $\tau_n = n$. Notice that the statistic, $T_{n,m}$, is just the squared Euclidean distance (properly normalized) between the autocorrelation vectors $(\hat{\rho}_{X,1}, \hat{\rho}_{X,2}, \dots, \hat{\rho}_{X,m})$ and $(\hat{\rho}_{Y,1}, \hat{\rho}_{Y,2}, \dots, \hat{\rho}_{Y,m})$. This fact suggests the study of others distances as well.

The general approach presented in Section 2 is applied to the statistic $T_{n,m}$ as follows:

1. Let $\mathbf{X}_j = (X_j, X_{j+1}, \dots, X_{j+l-1})$ and $\mathbf{Y}_j = (Y_j, Y_{j+1}, \dots, Y_{j+l-1})$ with $j = 1, 2, \dots, n - l + 1$ be the subsamples of l consecutive observations from \mathbf{X} and \mathbf{Y} , respectively. We

calculate the j -th subsampling statistic, $T_{l,m}^{(j)}$, by:

$$T_{l,m}^{(j)} = l \sum_{k=1}^m (\widehat{\rho}_{X_j,k} - r_{\widehat{\rho}_{j,k}})^2, \quad (5)$$

where $\widehat{\rho}_{X_j,k}$ and $\widehat{\rho}_{Y_j,k}$ are the k -th estimated autocorrelations using the subsamples \mathbf{X}_j and \mathbf{Y}_j , respectively.

2. We calculate $g_{n,l}(1 - \alpha)$ the $1 - \alpha$ quantile of $\widehat{G}_{n,l}(\cdot)$ using the expression (2).
3. We reject H_0 if and only if $T_{n,m} > g_{n,l}(1 - \alpha)$.

Notice that the proposed algorithm is valid for dependent series since the subsamples $(\mathbf{X}_j, \mathbf{Y}_j)$ can be considered as a vector of size l from the bidimensional process $\{(X_t, Y_t)\}_{t \in \mathbb{Z}}$. In the other hand, when the processes $\{X_t\}_{t \in \mathbb{Z}}$ and $\{Y_t\}_{t \in \mathbb{Z}}$ are independent, we can consider a larger number of subsamples modifying the above algorithm as follows:

1. Let $\mathbf{X}_i = (X_i, X_{i+1}, \dots, X_{i+l-1})$ with $i = 1, 2, \dots, n - l + 1$ and $\mathbf{Y}_j = (Y_j, Y_{j+1}, \dots, Y_{j+l-1})$ with $j = 1, 2, \dots, n - l + 1$ be the subsamples of l consecutive observations from \mathbf{X} and \mathbf{Y} , respectively. We calculate the (i, j) -th subsampling statistic, $T_{l,m}^{(i,j)}$, by:

$$T_{l,m}^{(i,j)} = l \sum_{k=1}^m (\widehat{\rho}_{X_i,k} - \widehat{\rho}_{Y_j,k})^2, \quad (6)$$

where $\widehat{\rho}_{X_i,k}$ and $\widehat{\rho}_{Y_j,k}$ are the k -th estimated autocorrelations using the subsamples \mathbf{X}_i and \mathbf{Y}_j , respectively.

2. We calculate $g_{n,l}(1 - \alpha)$ the $1 - \alpha$ quantile of $\widehat{G}_{n,l}(\cdot)$ using the following expression:

$$\widehat{G}_{n,l}(x) = \frac{1}{(n - l + 1)^2} \sum_{i=1}^{n-l+1} \sum_{j=1}^{n-l+1} I \left\{ T_{l,m}^{(i,j)} \leq x \right\}. \quad (7)$$

3. We reject H_0 if and only if $T_{n,m} > g_{n,l}(1 - \alpha)$.

4 Consistency results

The asymptotic validity of subsampling method for hypothesis testing is provided by the Theorem 5.1 of Politis et. al. (2001):

Theorem 5.1 (Politis et. al. (2001)):

- (i) Assume, for $P \in \mathbf{P}_0$, $G_n(P)$ converges weakly to a continuous limit law $G(P)$, whose corresponding cumulative distribution function is $G(\cdot, P)$ and whose $1 - \alpha$ quantile is $g(1 - \alpha, P)$. Assume that $l/n \rightarrow 0$ and $l \rightarrow \infty$ as $n \rightarrow \infty$. Also assume that $\alpha_Z(k) \rightarrow 0$ as $k \rightarrow \infty$, where $\alpha_Z(\cdot)$ is the mixing sequence corresponding $\{Z_t\}$. If $G(\cdot, P)$ is continuous at $g(1 - \alpha, P)$ and $P \in \mathbf{P}_0$ then $g_{n,l}(1 - \alpha) \rightarrow g(1 - \alpha, P)$ in probability and $\Pr_P\{T_n > g_{n,l}(1 - \alpha)\} \rightarrow \alpha$ as $n \rightarrow \infty$.
- (ii) Assume the test statistics is constructed so that $t_n(X_1, X_2, \dots, X_n) \rightarrow t(P)$ in probability, where $t(P)$ is a constant which satisfies $t(P) = 0$ if $P \in \mathbf{P}_0$ and $t(P) > 0$ if $P \in \mathbf{P}_1$. Assume that $l/n \rightarrow 0$, $l \rightarrow \infty$ and $\tau_l/\tau_n \rightarrow 0$ as $n \rightarrow \infty$. Also assume that $\alpha_Z(k) \rightarrow 0$ as $k \rightarrow \infty$, where $\alpha_Z(\cdot)$ is the mixing sequence corresponding $\{Z_t\}$. Then if $P \in \mathbf{P}_1$, the rejection probability satisfies $\Pr_P\{T_n > g_{n,l}(1 - \alpha)\} \rightarrow 1$ as $n \rightarrow \infty$.

The following proposition and its corollaries establish the required assumptions of the previous theorem for the statistics defined by (4).

Proposition 1: Suppose that (X_1, X_2, \dots, X_n) and (Y_1, Y_2, \dots, Y_n) are samples of two stationary zero mean processes with strong mixing coefficient $\alpha_X(\cdot)$ and $\alpha_Y(\cdot)$, respectively. Also, assume that one of the two following conditions hold:

- (A) $\Pr\{|X_1| \leq c\} = 1$ and $\Pr\{|Y_1| \leq c\} = 1$ for some $c \in (0, \infty)$, and $\sum_{k=1}^{\infty} \alpha_X(k) < \infty$ and $\sum_{k=1}^{\infty} \alpha_Y(k) < \infty$.
- (B) $E|X_1|^{4+2\delta} < \infty$ and $E|Y_1|^{4+2\delta} < \infty$ and $\sum_{k=1}^{\infty} \alpha_X(k)^{\delta/(2+\delta)} < \infty$ and $\sum_{k=1}^{\infty} \alpha_Y(k)^{\delta/(2+\delta)} < \infty$ for some $\delta \in (0, \infty)$.

Then, for any fixed nonnegative integer m ,

$$n^{1/2}(\tilde{\gamma}_{X,0} - \gamma_{X,0}, \tilde{\gamma}_{X,1} - \gamma_{X,1}, \dots, \tilde{\gamma}_{X,m} - \gamma_{X,m}, \tilde{\gamma}_{Y,0} - \gamma_{Y,0}, \tilde{\gamma}_{Y,1} - \gamma_{Y,1}, \dots, \tilde{\gamma}_{Y,m} - \gamma_{Y,m}) \quad (8)$$

is asymptotically normal with mean zero and covariance matrix $\mathbf{C}_{X,Y}$, where $\mathbf{C}_{X,Y}$ is a $2(m+1) \times 2(m+1)$ matrix with entries given by (13) and (14).

Proof: For sake of simplicity we present the proof for the following vector:

$$n^{1/2}(\tilde{\gamma}_{X,i} - \gamma_{X,i}, \tilde{\gamma}_{X,j} - \gamma_{X,j}, \tilde{\gamma}_{Y,k} - \gamma_{Y,k}, \tilde{\gamma}_{Y,l} - \gamma_{Y,l}), \quad (9)$$

where indexes i, j, k and l are in $\{0, 1, \dots, m\}$.

In order to prove that (9) is asymptotically normal with mean zero and covariance matrix

$$\mathbf{C} = \begin{bmatrix} c_{X,i,i} & c_{X,i,j} & c_{XY,i,k} & c_{XY,i,l} \\ c_{X,j,i} & c_{X,j,j} & c_{XY,j,k} & c_{XY,j,l} \\ c_{XY,k,i} & c_{XY,k,j} & c_{Y,k,k} & c_{Y,k,l} \\ c_{XY,l,i} & c_{XY,l,j} & c_{Y,l,k} & c_{Y,l,l} \end{bmatrix}, \quad (10)$$

we will use a Cramer-Wold argument, i.e., we will consider arbitrary linear combination and establish the asymptotical univariate normality.

Let be $\lambda_i, \lambda_j, \lambda_k$ and λ_l arbitrary real constants, then the linear combination of (9) satisfies,

$$\begin{aligned} & n^{1/2} (\lambda_i(\tilde{\gamma}_{X,i} - \gamma_{X,i}) + \lambda_j(\tilde{\gamma}_{X,j} - \gamma_{X,j}) + \lambda_k(\tilde{\gamma}_{Y,k} - \gamma_{Y,k}) + \lambda_l(\tilde{\gamma}_{Y,l} - \gamma_{Y,l})) \\ &= n^{1/2} \left(\sum_{t=1}^{n-m} \lambda_i (X_t X_{t+i} - \gamma_{X,i}) + \sum_{t=1}^{n-m} \lambda_j (X_t X_{t+j} - \gamma_{X,j}) \right. \\ & \quad \left. + \sum_{t=1}^{n-m} \lambda_k (Y_t Y_{t+k} - \gamma_{Y,k}) + \sum_{t=1}^{n-m} \lambda_l (Y_t Y_{t+l} - \gamma_{Y,l}) \right) \\ &= n^{1/2} \sum_{t=1}^{n-m} (\lambda_i (X_t X_{t+i} - \gamma_{X,i}) + \lambda_j (X_t X_{t+j} - \gamma_{X,j}) \\ & \quad + \lambda_k (Y_t Y_{t+k} - \gamma_{Y,k}) + \lambda_l (Y_t Y_{t+l} - \gamma_{Y,l})). \end{aligned} \quad (11)$$

We define $Z_t = \lambda_i (X_t X_{t+i} - \gamma_{X,i}) + \lambda_j (X_t X_{t+j} - \gamma_{X,j}) + \lambda_k (Y_t Y_{t+k} - \gamma_{Y,k}) + \lambda_l (Y_t Y_{t+l} - \gamma_{Y,l})$.

Then, $\{Z_t\}$ is a zero mean stationary process with mixing coefficient satisfying:

$$\alpha_Z(h) \leq \alpha_X(h-i) + \alpha_X(h-j) + \alpha_Y(h-k) + \alpha_Y(h-l). \quad (12)$$

The definition of Z_t and (12) implies that if $\{X_t\}$ and $\{Y_t\}$ satisfy the condition (A) or (B) then $\{Z_t\}$ also satisfies similar conditions. In the case of condition (B) we have $E|Z_1|^{2+\delta} < \infty$. Hence we can apply a central limit theorem for strong mixing process (see, v.g., Theorem A.8

of Lahiri (2003)) to conclude that $n^{1/2} \sum_{t=1}^{n-m} Z_t$ is asymptotically normal with mean zero and variance $\sigma_\infty^2 = \sum_{t=-\infty}^{\infty} \text{cov}(Z_1, Z_{1+t}) = \boldsymbol{\lambda}' \mathbf{C} \boldsymbol{\lambda}$, where $\boldsymbol{\lambda} = (\lambda_i, \lambda_j, \lambda_k \lambda_l)'$. Then, the Cramer-Wold theorem implies that (9) is asymptotically normal with mean zero and covariance matrix \mathbf{C} . It only rest to give the expression for the entries of matrix \mathbf{C} .

For indexes in X or Y , i.e., $(i, i), (i, j), (j, j), (k, k), (k, l)$ or (l, l) , the expression coincide with the original Bartlett's formula:

$$\begin{aligned}
c_{X,i,j} &\equiv \lim_{n \rightarrow \infty} n \text{cov}(\tilde{\gamma}_{X,i}, \tilde{\gamma}_{X,j}) \\
&= \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^{n-m} \sum_{s=1}^{n-m} \gamma_{X,s-t} \gamma_{X,s-t+j-i} + \gamma_{X,s-t+j} \gamma_{X,s-t-i} + \\
&\quad + \kappa_X(s-t, i, j-i) \\
&= \sum_{t=-\infty}^{\infty} \gamma_{X,t} \gamma_{X,t+j-i} + \gamma_{X,t+j} \gamma_{X,t-i} + \kappa_X(t, i, j-i),
\end{aligned} \tag{13}$$

where $\kappa_X(t, i, j-i)$ denotes the fourth joint cumulant of the distribution of $(X_h, X_{h+i}, X_{h+t}, X_{h+t+j})$. Notice that the double sum is simplified by noting that in each term s and t only occur in the form $s-t$, see Section 5.3.3 of Priestley (1981). Similar expressions can be obtained for the others pairs of indexes.

For indexes in X and Y , i.e., $(i, k), (j, k), (i, l)$ or (j, l) we can use the expression (5.3.16) in Section 5.3.3 of Priestley (1981) to obtain:

$$\begin{aligned}
c_{XY,i,k} &= \lim_{n \rightarrow \infty} n \text{cov}(\tilde{\gamma}_{X,i}, \tilde{\gamma}_{Y,k}) \\
&= \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^{n-m} \sum_{s=1}^{n-m} \gamma_{XY,s-t} \gamma_{XY,s-t+j-i} + \gamma_{XY,s-t+j} \gamma_{XY,s-t-i} + \\
&\quad + \kappa_{XY}(s-t, i, j-i) \\
&= \sum_{t=-\infty}^{\infty} \gamma_{XY,t} \gamma_{XY,t+i-j} + \gamma_{XY,t-j} \gamma_{XY,t+i} + \kappa_{X,Y}(t, i, j-i),
\end{aligned} \tag{14}$$

where $\gamma_{XY,\cdot}$ denotes the cross-covariance function of and $\kappa_{X,Y}(t, i, j-i)$ denotes the fourth joint cumulant of the distribution of $(X_h, X_{h+i}, Y_{h+t}, Y_{h+t+j})$. Notice that the double sum is simplified by noting that in each term s and t only occur in the form $s-t$, see Section 5.3.3 of Priestley (1981). Similar expressions can be obtained for the others pairs of indexes. Notice that $c_{XY,i,k}$ are zero for independent processes. ■

Remark 1: Notice that the above proposition was established for the modified covariance estimator $\tilde{\gamma}_{X,j} = n^{-1} \sum_{t=1}^{n-m} X_t X_{t+j}$, i.e. assuming a known process mean, instead of the usual covariance estimator $\hat{\gamma}_{X,j} = n^{-1} \sum_{t=1}^{n-j} (X_t - \bar{X}_n)(X_{t+j} - \bar{X}_n)$. It is straightforward to establish that $\tilde{\gamma}_{X,j}$ and $\hat{\gamma}_{X,j}$ have the same limit distribution by using $n^{1/2}(\tilde{\gamma}_{X,j} - \hat{\gamma}_{X,j}) = n^{1/2} \bar{X}_n^2 - n^{-1/2} \bar{X}_n \sum_{t=1}^{n-j} (X_t + X_{t+j}) + o_P(n^{-1})$.

Remark 2: The Proposition 1 is a generalization of Theorem 3.1 in Romano and Thombs (1996) for bivariate stationary processes. Notice that condition (B) is used in Romano and Thombs (1996) but the results using condition (A) are new. The condition (A) is weaker than (B) in the restriction imposed on the mixing sequences.

Corollary 1: *Under the assumptions of Proposition 1, for any fixed nonnegative integer m , we have,*

$$n^{1/2}(\tilde{\rho}_{X,1} - \rho_{X,1}, \tilde{\rho}_{X,2} - \rho_{X,2}, \dots, \tilde{\rho}_{X,m} - \rho_{X,m}, \tilde{\rho}_{Y,1} - \rho_{Y,1}, \tilde{\rho}_{Y,2} - \rho_{Y,2}, \dots, \tilde{\rho}_{Y,m} - \rho_{Y,m}) \quad (15)$$

is asymptotically normal with mean zero and covariance matrix $\mathbf{R}_{X,Y}$, where $\mathbf{R}_{X,Y}$ is a $2m \times 2m$ matrix with entries given by (19) and (20).

Proof: Since $\tilde{\rho}_{X,i} = \tilde{\gamma}_{X,i}/\tilde{\gamma}_{X,0}$ and $\tilde{\rho}_{Y,k} = \tilde{\gamma}_{Y,k}/\tilde{\gamma}_{Y,0}$ we can use expression (5.3.34) in Section 5.3.4 of Priestley (1981) to obtain:

$$n^{1/2}(\tilde{\rho}_{X,i} - \rho_{X,i}) = \gamma_{X,0}^{-1} n^{1/2}(\tilde{\gamma}_{X,i} - \gamma_{X,i}) - \rho_{X,i} \gamma_{X,0}^{-1} n^{1/2}(\tilde{\gamma}_{X,0} - \gamma_{X,0}) + o_P(1) \quad (16)$$

and

$$n^{1/2}(\tilde{\rho}_{Y,k} - \rho_{Y,k}) = \gamma_{Y,0}^{-1} n^{1/2}(\tilde{\gamma}_{Y,k} - \gamma_{Y,k}) - \rho_{Y,k} \gamma_{Y,0}^{-1} n^{1/2}(\tilde{\gamma}_{Y,0} - \gamma_{Y,0}) + o_P(1). \quad (17)$$

Then, from Proposition 1 we have that the vector:

$$n^{1/2}(\tilde{\rho}_{X,i} - \rho_{X,i}, \tilde{\rho}_{X,j} - \rho_{X,j}, \tilde{\rho}_{Y,k} - \rho_{Y,k}, \tilde{\rho}_{Y,l} - \rho_{Y,l}), \quad (18)$$

is asymptotically normal with mean zero and covariance matrix \mathbf{R} with entries defined by

$$\begin{aligned}
r_{X,i,j} &\equiv \lim_{n \rightarrow \infty} n \operatorname{cov}(\tilde{\rho}_{X,i}, \tilde{\rho}_{X,j}) \\
&= \lim_{n \rightarrow \infty} \gamma_{X,0}^{-2} n \operatorname{cov}(\tilde{\gamma}_{X,i} - \rho_{X,i} \tilde{\gamma}_{X,0}, \tilde{\gamma}_{X,j} - \rho_{X,j} \tilde{\gamma}_{X,0}) \\
&= \gamma_{X,0}^{-2} (c_{X,i,j} - \rho_{X,i} c_{X,0,j} - \rho_{X,j} c_{X,0,i} + \rho_{X,i} \rho_{X,j} c_{X,0,0}),
\end{aligned} \tag{19}$$

where $c_{X,\cdot,\cdot}$ is obtained from (13), and

$$\begin{aligned}
r_{XY,i,k} &\equiv \lim_{n \rightarrow \infty} n \operatorname{cov}(\tilde{\rho}_{X,i}, \tilde{\rho}_{Y,k}) \\
&= \lim_{n \rightarrow \infty} \gamma_{X,0}^{-1} \gamma_{Y,0}^{-1} n \operatorname{cov}(\tilde{\gamma}_{X,i} - \rho_{X,i} \tilde{\gamma}_{X,0}, \tilde{\gamma}_{Y,k} - \rho_{Y,k} \tilde{\gamma}_{Y,0}) \\
&= \gamma_{X,0}^{-1} \gamma_{Y,0}^{-1} (c_{XY,i,k} - \rho_{X,i} c_{XY,0,j} - \rho_{Y,k} c_{XY,i,0} + \rho_{X,i} \rho_{Y,k} c_{XY,0,0}),
\end{aligned} \tag{20}$$

where $c_{XY,\cdot,\cdot}$ is obtained from (14).

As in Proposition 1, we only derive the expression for the pairs of indexes (i, j) and (i, k) but similar expressions can be obtained for the others pairs of indexes. \blacksquare

Remark 3: Using a similar argument to Remark 1, we can establish that the modified correlation estimator $\tilde{\rho}_{X,j}$ and the usual correlation estimator $\hat{\rho}_{X,j} = \hat{\gamma}_{X,j} / \hat{\gamma}_{X,0}$ have the same limit distribution.

Remark 4: The Corollary 1 is a generalization of Theorem 3.2 in Romano and Thombs (1996) for bivariate stationary processes.

Corollary 2: Under the assumptions of Proposition 1, for any fixed nonnegative integer m , we have that $T_{n,m}$ is asymptotically distributed, under $H_0 : \rho_{X,1} = \rho_{Y,1}, \rho_{X,2} = \rho_{Y,2}, \dots, \rho_{X,m} = \rho_{Y,m}$, as a $\sum_{k=1}^m \lambda_k \chi_1^2$, where each χ_1^2 variate is distributed independently of every other and the λ 's are the real nonzero eigenvalues of the matrix $\mathbf{R}_{X,Y} \mathbf{H}$. $\mathbf{R}_{X,Y}$ is the asymptotical covariance matrix of (15) and $\mathbf{H} = \begin{bmatrix} \mathbf{I}_m & -\mathbf{I}_m \\ -\mathbf{I}_m & \mathbf{I}_m \end{bmatrix}$, where \mathbf{I}_m is the $m \times m$ identity matrix.

Proof: $T_{n,m}$ can be written as follows

$$\begin{aligned}
T_{n,m} &= n \sum_{k=1}^m \left((\widehat{\rho}_{X,k} - \rho_{X,k}) - (\widehat{\rho}_{Y,k} - \rho_{Y,k}) - (\rho_{Y,k} - \rho_{X,k}) \right)^2 \\
&= n \sum_{k=1}^m \left((\widehat{\rho}_{X,k} - \rho_{X,k}) - (\widehat{\rho}_{Y,k} - \rho_{Y,k}) \right)^2 + n \sum_{k=1}^m (\rho_{Y,k} - \rho_{X,k})^2 + \\
&\quad + n \sum_{k=1}^m (\rho_{Y,k} - \rho_{X,k})(\widehat{\rho}_{X,k} - \rho_{X,k}) + n \sum_{k=1}^m (\rho_{Y,k} - \rho_{X,k})(\widehat{\rho}_{Y,k} - \rho_{Y,k}) \\
&= S_1 + S_2 + S_3 + S_4.
\end{aligned} \tag{21}$$

The first summand, S_1 , can be expressed as

$$S_1 = n^{1/2}(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho})' \mathbf{H} n^{1/2}(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}), \tag{22}$$

where $(\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho})$ denotes the following $2m \times 1$ vector

$$(\widehat{\rho}_{X,1} - \rho_{X,1}, \widehat{\rho}_{X,2} - \rho_{X,2}, \dots, \widehat{\rho}_{X,m} - \rho_{X,m}, \widehat{\rho}_{Y,1} - \rho_{Y,1}, \widehat{\rho}_{Y,2} - \rho_{Y,2}, \dots, \widehat{\rho}_{Y,m} - \rho_{Y,m})'. \tag{23}$$

Then, from the Theorem 2.1 of Box (1954) on quadratic forms and the Remark 3 it follows that S_1 is asymptotically distributed as a $\sum_{k=1}^m \lambda_k \chi_1^2$, where each χ_1^2 variate is distributed independently of every other and the λ 's are the real nonzero eigenvalues of the matrix $\mathbf{R}_{X,Y} \mathbf{H}$.

The second summand, $S_2 = n \sum_{k=1}^m (\rho_{Y,k} - \rho_{X,k})^2$, is zero under H_0 .

For the third and the fourth summands, Corollary 1 implies that $S_3 = O_P(n^{-1/2})$ and $S_4 = O_P(n^{-1/2})$ since they are linear combination of vectors $(\widehat{\rho}_{X,1} - \rho_{X,1}, \widehat{\rho}_{X,2} - \rho_{X,2}, \dots, \widehat{\rho}_{X,m} - \rho_{X,m})$ and $(\widehat{\rho}_{Y,1} - \rho_{Y,1}, \widehat{\rho}_{Y,2} - \rho_{Y,2}, \dots, \widehat{\rho}_{Y,m} - \rho_{Y,m})$, respectively.

Finally, the Slutsky's lemma establish that $T_{n,m}$ and S_1 have the same limit distribution, under H_0 . ■

Remark 5: The Corollary 2 establishes the required assumption for Theorem 5.1(i) of Politis et al. (2001), i.e., the distribution of $T_{n,m}$, $G_n(P)$, converges weakly to a continuous limit law $G(P) = \sum_{k=1}^m \lambda_k \chi_1^2$.

Corollary 3: *Under the assumptions of Proposition 1, for any fixed nonnegative integer m , we have that $t_n = t_n(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n)$ converges, in probability, to a constant $t(P)$ which satisfies $t(P) = 0$ under $H_0 : \rho_{X,1} = \rho_{Y,1}, \rho_{X,2} = \rho_{Y,2}, \dots, \rho_{X,m} = \rho_{Y,m}$, and $t(P) > 0$ under $H_1 : \rho_{X,k} \neq \rho_{Y,k}$, for some $k \in \{1, 2, \dots, m\}$.*

Proof: From (21) we have

$$t_n = n^{-1}(S_1 + S_2 + S_3 + S_4), \quad (24)$$

and from the proof of Corollary 2 we have that $n^{-1}S_1 = O_P(n^{-1})$, $S_3 = O_P(n^{-3/2})$ and $S_4 = O_P(n^{-3/2})$. Then, $t_n = \sum_{k=1}^m (\rho_{Y,k} - \rho_{X,k})^2 + O_P(n^{-1})$. It is clear that constant $t(P) = \sum_{k=1}^m (\rho_{Y,k} - \rho_{X,k})^2$ satisfies $t(P) = 0$ under H_0 , and $t(P) > 0$ under H_1 . ■

Remark 6: The Corollary 3 establishes the required assumption for Theorem 5.1(ii) of Politis et al. (2001).

5 Simulation results

In this section, we study the behavior of the proposed testing procedure in terms of size and power for finite samples. We will use the following models:

Model 1: $X_t = \phi X_{t-1} + \varepsilon_t$, with $\phi = 0, 0.2, 0.4, 0.6, 0.8$

and where ε_t i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$ and σ_ε^2 is fixed such that $\text{Var}(X_t) = 1$. We consider two sample sizes, $n = 256$ and 512 , the nominal size $\alpha = 0.05$ and the maximum lag $m = 10$. In future research, we will study the selection of m . We expect that the power of the test based on $T_{n,m}$ will decrease with m , as in the case in goodness-of-fit test (see, v.g. Peña and Rodríguez (2002)).

In this paper, we present the simulations results for independent series using the algorithm based on the estimation (7). Notice that the number of subsamples, $(n - l + 1)^2$, can be high,

e.g., with $n = 256$ and $l = 32$ we have 50625 subsamples, therefore we will use $B = 1000$ random subsamples to approximate the expression (7).

In Tables 1 and 2, we present the results for the size of the test using $l = 4, 8, 16, 32, 64, 128$ for $n = 256$ and $l = 4, 8, 16, 32, 64, 128, 256$ for $n = 512$. To estimate the sizes we use 1000 independent replicas; therefore the estimated sizes between 0.0365 and 0.0635 are not significantly different from the nominal 0.05. In the cases $l = 4$ and 8, we use $m = 3$ and 7, respectively. In the fourth column of both tables we report the size of the test using the calibration procedure proposed in Politis et al. (1999). Additionally, we present the results for the procedures proposed by Diggle and Fisher (1991) and Maharaj (2000) in the second and the third column, respectively.

=====> Tables 1 and 2 around here. <=====

As we expect, the results are dependent on the subsample lengths. We obtain sizes near to the nominal value, 0.05, for the larger values of l . The calibration procedure, in almost all the cases, obtains a size values not statistically different from the nominal value. The Maharaj's procedure tends to overestimate the size of the test (in all the cases for $n = 256$ and in three cases for $n = 512$). Notice that the Maharaj's method uses the additional information of the autoregressive structure of the generating process. Also the Diggle and Fisher's procedure tends to overestimate the size of the test when the sample size is 512. Notice that the Diggle and Fisher's method is only applicable to independent series.

In Tables 3 and 4, we present the results of the power of the following hypothesis test: H_0 : AR(1) with $\phi = 0.0$ (white noise) against H_1 : AR(1) with $\phi \neq 0.0$ taking $\phi = 0.2, 0.4, 0.6$ and 0.8. As we expect, we observe that the power increases as ϕ increases since the models in the alternative hypothesis are more different from the model in the null hypothesis. For the sample size $n = 256$ with $\phi \geq 0.6$ and for $n = 512$ with $\phi \geq 0.4$, the powers with the

calibration procedure are close to 1. Notice that Diggle and Fisher's and Maharaj's methods have higher powers than the proposed procedure but their size are generally overestimated.

=====> Tables 3 and 4 around here. <=====

6 Conclusion

In this paper we have proposed a procedure based on the subsampling techniques for the comparison of stationary time series that are not necessarily independent. We have established the asymptotical consistency of the proposed method. In the simulation experiments, we have confirmed the competitive behavior of the subsampling test procedure with regard to the procedures proposed by Diggle and Fisher (1991) and Maharaj (2000).

References

- [1] Basawa, I., Billard, L. and Srinivasan, R. (1984) Large-sample test of homogeneity for time series, *Biometrika* **71**, 203–206.
- [2] Box, G.E.P. (1954) Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification, *Annals of Mathematical Statistics*, **25**, 290–302.
- [3] Coates, D.S. and Diggle, P.J. (1986) Tests for comparing two estimated spectral densities, *Journal of Time Series*, **7**, 7–20.
- [4] Diggle, P.J. and Fisher, N.I. (1991) Nonparametric comparison of cumulative periodograms, *Applied Statistics*, **40**, 423–434.
- [5] Guo, J.H. (1999) A nonparametric test for the parallelism of two first-order autoregressive processes, *Australian and New Zealand Journal of Statistics*, **41**, 59–65.
- [6] Lahiri, S.N. (2003) *Resampling methods for dependent data*, Springer-Verlag.

- [7] Maharaj, E.A. (1996) A significance test for classifying ARMA models, *Journal of Statistical Computation and Simulation*, **54**, 305–331.
- [8] Maharaj, E.A. (2000) Clusters of time series, *Journal of Classification*, **17**, 297–314.
- [9] Peña, D. and Rodríguez, J. (2002) A powerful portmanteau test of lack of fit for time series, *The Journal of the American Statistical Association*, **97**, 601–610.
- [10] Politis, D.N. and Romano, J.P. (1994) Large sample confidence regions based on subsamples under minimal assumptions, *The Annals of Statistics*, **22**, 2031–2050.
- [11] Politis, D.N., Romano, J.P. and Wolf, M. (1999) *Subsampling*. Springer–Verlag.
- [12] Politis, D.N., Romano, J.P. and Wolf, M. (2001) On the asymptotic theory of subsampling, *Statistica Sinica*, **11**, 1105–1124.
- [13] Priestley, M.B. (1981) *Spectral analysis and time series*, Academic Press.
- [14] Romano, J.P. and Thombs, L.A. (1996) Inference for autocorrelation under weak assumptions, *The Journal of the American Statistical Association*, **91**, 590–600.
- [15] Swanepoel, J.W.H. and Van Wyk, J.W.J. (1986). *The comparison of two spectral density functions using the bootstrap*. *Journal of Statistical Computation and Simulation* **24**, 271–282.
- [16] Timmer, J., Lauk, M., Vach, W. and Lucking, C.H. (1999). *A test for a difference between spectral peak frequencies*. *Computational Statistics and Data Analysis* **30**, 45–50.

AR(1) : ϕ	D_{DF}	D_M	l_{cal}	$l = 4$	$l = 8$	$l = 16$	$l = 32$	$l = 64$	$l = 128$
0.0	0.060	0.080	0.051	0.642	0.654	0.335	0.121	0.062	0.067
0.2	0.058	0.069	0.062	0.642	0.647	0.336	0.155	0.092	0.087
0.4	0.055	0.082	0.047	0.608	0.675	0.388	0.153	0.069	0.070
0.6	0.064	0.073	0.057	0.583	0.664	0.458	0.224	0.086	0.066
0.8	0.061	0.069	0.073	0.558	0.702	0.506	0.285	0.121	0.091

Table 1: Estimated size of the test ($n = 256$).

AR(1) : ϕ	D_{DF}	D_M	l_{cal}	$l = 8$	$l = 16$	$l = 32$	$l = 64$	$l = 128$	$l = 256$
0.0	0.065	0.073	0.046	0.655	0.322	0.127	0.064	0.047	0.069
0.2	0.080	0.053	0.068	0.685	0.404	0.173	0.090	0.071	0.071
0.4	0.103	0.073	0.048	0.657	0.384	0.182	0.086	0.048	0.060
0.6	0.083	0.066	0.050	0.711	0.466	0.234	0.108	0.057	0.069
0.8	0.110	0.055	0.051	0.670	0.501	0.277	0.116	0.060	0.072

Table 2: Estimated size of the test ($n = 512$).

AR(1) : ϕ	D_{DF}	D_M	l_{cal}	$l = 4$	$l = 8$	$l = 16$	$l = 32$	$l = 64$	$l = 128$
0.2	0.526	0.619	0.212	0.891	0.868	0.660	0.393	0.246	0.205
0.4	0.977	0.994	0.761	1.000	0.999	0.987	0.930	0.806	0.627
0.6	0.998	1.000	0.988	1.000	1.000	1.000	1.000	0.991	0.944
0.8	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995

Table 3: Estimated power of the test ($n = 256$).

AR(1) : ϕ	D_{DF}	D_M	l_{cal}	$l = 8$	$l = 16$	$l = 32$	$l = 64$	$l = 128$	$l = 256$
0.2	0.771	0.908	0.420	0.978	0.890	0.688	0.543	0.432	0.343
0.4	1.000	1.000	0.975	1.000	1.000	0.999	0.995	0.977	0.899
0.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997
0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: Estimated power of the test ($n = 512$).