International Journal of Wavelets, Multiresolution and Information Processing
© World Scientific Publishing Company

COMPARING TIME-VARYING AUTOREGRESSIVE STRUCTURES OF LOCALLY STATIONARY PROCESSES

GLADYS E. SALCEDO
Department of Mathematics, University of Quindío
Armenia, Quindío, Colombia
gsalcedo@ime.usp.br

JOÃO R. SATO
Institute of Mathematics and Statistics, University of São Paulo,
Rua do Matão 1010, 05508-090, São Paulo, Brazil.
jsato@ime.usp.br

PEDRO A. MORETTIN
Institute of Mathematics and Statistics, University of São Paulo,
Rua do Matão 1010, 05508-090, São Paulo, Brazil.
pam@ime.usp.br

CLÉLIA M. TOLOI
Institute of Mathematics and Statistics, University of São Paulo,
Rua do Matão 1010, 05508-090, São Paulo, Brazil.
clelia@ime.usp.br

Received (17 February 2006)
Accepted (4 July 2007)
Communicated by (xxxxxxxxxx)

In this paper a novel statistical test is introduced to compare two locally stationary time series. The proposed approach is a Wald test considering time-varying autoregressive modelling and function projections in adequate spaces. The covariance structure of the innovations may be also time-varying. In order to obtain function estimators for the time-varying autoregressive parameters, we consider function expansions in splines and wavelet bases. Simulation studies provide evidence that the proposed test has a good performance. We also assess its usefullness when applied to a financial time series.

Keywords: Hypotheses testing; locally stationary processes; time-varying AR models; splines; wavelets.

AMS Subject Classification: 22E46, 53C35, 57S20

1. Introduction

Comparison of time series has been a problem of interest in many studies. For different purposes, it is very important to know if there are similarities or differences between two or more time series. Several techniques have been proposed in order to identify these similarities for both stationary and non-stationary series. For example, if two series come from the same process, we can obtain better estimators by pooling the data sets [1,11]. As a classification problem, group of series can be discriminated in different clusters according to the degree of similarity, see for example Ref. 19, 14, 15, 16 and 2. If the focus of a study is monitoring some geophysical variable, as salinity in different deep sea levels, and these time series are similar, it will be enough to take samples in only one level, reducing time and costs of the study [22]. As a financial application, Maharaj [17,18] compared interest rates patterns among various countries.

Most of the existing techniques for comparing time series are applicable to stationary time series or to non-stationary ones that can be transformed to stationary by some transformation as differencing. On the other hand, these procedures are basically based on comparing either the spectral densities of pair of time series [3,9], or the coefficients of an autoregressive(AR) model [11,14,16].

Since many phenomena in applied sciences show a non-stationary behavior, for example, the second order structure changes over time, realizations of these processes are not easily transformable to stationarity. In order to evaluate the similarity among time series that are non-stationary in variance, Maharaj [17,18] compared the evolutionary spectra and the sums of squared wavelet coefficients at different times or scales, respectively, using randomization tests. Huang et al. [13] using the SLEX (smooth localized complex exponentials) model developed an approach for discrimination and classification of non-stationary time series and for Dahlhaus locally stationary series, Sakiyama and Taniguchi [21] and Shumway [24] discussed the problem of discrimination.

On the other hand, since locally stationary series can be represented by autoregressive models with time-varying coefficients [6], in this paper we propose a new approach to evaluate the similarity among two locally stationary series comparing their autoregressive parameters. Further, this new approach includes the case where the variance and covariances of the innovations also change over time.

The remainder of the paper is organized as follows. In section 2, a brief description about both locally stationary processes and approximation theory of functions is presented. In section 3, the novel testing procedure is described. Some simulation results are presented in section 4 and an application to financial time series is illustrated in section 5. Finally, in section 6, some conclusions are given.

2. Background

Autoregressive models form a very important class of stationary models, due to the fact that they can model a wide variety of phenomena, they are easy to estimate and interprete and the asymptotic properties of autoregressive estimators are well understood. In the context of locally stationary processes, we consider the autoregressive model with varying-time coefficients. It is given by the following difference equation:

$$\sum_{i=0}^{p} a_i\left(\frac{t}{T}\right)\left(X_{t-i,T}-\mu\left(\frac{t-i}{T}\right)\right)=\sigma\left(\frac{t}{T}\right)\varepsilon_t,\ t\in\mathcal{Z}, \tag{2.1}$$

where $a_0 \equiv 1$ and $\{\varepsilon_t, t = 1, \ldots, T\}$ are independent random variables with mean zero and variance one. We assume that the functions $\sigma(u)$ and $a_i(u)$, where $u = \frac{t}{T}$, are continuous on $\mathcal{R}$ with $\sigma(u) = \sigma(0), a_i(u) = a_i(0)$ for $u < 0,\ \sigma(u) = \sigma(1), a_i(u) = a_i(1)$ for $u > 1$, and differentiable for $u \in (0, 1)$ with bounded derivatives. Here, $\mathcal{Z}$ denotes the set of all integers and $\mathcal{R}$ the set of all real numbers.

2.1. *Locally Stationary Processes*

Stationarity has always been a main assumption in the theoretical treatment of time series. For example, the well-known ARMA models and the classical Cramér spectral representation are different ways to represent a stationary time series. Nevertheless, many phenomena in the applied science show a non-stationary behavior (e.g. economics, oceanography, medicine), the second order structure of these processes is no longer time-shift invariant but changes over time. Priestley [20] considered processes having a time-varying spectral representation

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} A_t(\omega) d\xi(\omega), \quad t \in \mathcal{Z},$$

with an orthogonal increment process $\xi(\omega)$ and a time-varying transfer function $A_t(\omega)$. But, within the approach of Priestley, asymptotic considerations are not possible.

In the representation (2.1), if $T \to \infty$, it means that we have in the sample $X_{1,T}, X_{2,T}, \ldots, X_{T,T}$ more and more "observations" for the local structure of $a_i(\cdot)$ at each time point.

Dahlhaus [7] defined the following class of non-stationary processes having a time-varying spectral representation.

Definition: A sequence of stochastic processes $\{X_{t,T},\ t = 1, \ldots, T,\}$ is called locally stationary with transfer function A^0 and trend μ if there exists a representation

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} e^{i\omega t} A^0_{t,T}(\omega) d\xi(\omega), \tag{2.2}$$

where:

- $\xi(\omega)$ is a stochastic process on $[-\pi, \pi]$ with $\overline{\xi(\omega)} = \xi(-\omega),\ E(\xi(\omega)) = 0$, with orthonormal increments, i.e.

  $$Cov[d\xi(\omega), d\xi(\omega')] = \delta(\omega - \omega')$$

  such that

  $$Cum\{d\xi(\omega_1), \ldots, d\xi(\omega_k)\} = \eta\left(\sum_{j=1}^{k}\omega_j\right) g_k(\omega_1, \ldots, \omega_{k-1}) d\omega_1 \ldots d\omega_k,$$

where $Cum\{\ldots\}$ denotes the cumulant of $k-$th order, $g_1(\omega) = 0,\ g_2(\omega) = 1,\ |g_k(\omega_1, \ldots, \omega_{k-1})| \leq const_k$ for all k and $\eta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j)$ is the period 2π extension of the Dirac delta function.

- There exists a constant K and a 2π-periodic function $A : [0, 1] \times \mathcal{R} \to \mathcal{C}$ with $A(u, -\omega) = \overline{A(u, \omega)}$, and
$$\sup_{t,\omega} |A^0_{t,T}(\omega) - A\left(\frac{t}{T}, \omega\right)| \leq KT^{-1}$$
for all t, where $\mathcal{C}$ denotes the set of complex numbers.

The functions $A(u, \omega)$ and $\mu(\omega)$ are assumed to be continuous in u, in order to guarantee that the process has a locally stationary behavior. The evolutionary spectrum of $X_{t,T}$ is defined as $f_x(u, \omega) = |A(u, \omega)|^2$.

Dahlhaus [6] proved that in (2.1), $X_{t,T}$ has the representation (2.2) with

$$A(u, \omega) = \frac{\sigma(u)}{\sqrt{2\pi}} \left(1 + \sum_{j=1}^{p} a_j(u) e^{-ij\omega}\right)^{-1},$$

where

$$\varepsilon_t = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{i\omega t} d\xi(\omega), \quad t \in \mathcal{Z}.$$

For simplicity, in the sequel we assume that $\mu(u) = 0$.

2.2. *Function Expansions*

In most of the approximation theory of functions, the purpose is to expand any function belonging to a specified space by linear combinations of some basis functions which generally form an orthonormal basis for that space. For example, sines and cosines form a basis for $L^2[0, 2\pi]$, B-splines and wavelets form a basis for $L^2(\mathcal{R})$. In order to approximate a smooth function $f \in L^2[0, 1]$, as will be our interest, we introduce two appropriate methods, polynomial splines and wavelets.

2.2.1. *Splines*

A real function $s(x)$ is called a spline function (or simply “spline”) of degree $r \geq 0$ on an interval χ with knot points $x_0 < x_1 < \ldots < x_{M+1}$, where x_0 and x_{M+1} are the two end points of χ, if

- $s(x)$ is a polynomial of degree not greater than r on each of the intervals $[x_m, x_{m+1}],\ m = 0, 1, \ldots, M$, with the polynomial pieces joining smoothly at the knot points;

- $s(x)$ globally has $r-1$ continuous derivatives for $r \geq 1$.

A piecewise constant function, linear spline, quadratic spline and cubic spline correspond to $r = 0, 1, 2, 3$ respectively.

The collection $S_r(x_1, \ldots, x_M)$ whose elements are spline functions of degree r and knot sequence $\{x_1, \ldots, x_M\}$, forms a linear function space and it is called *spline space.* It can be demonstrated that B-splines form a basis of spline spaces with the advantage that they are splines which have the smallest possible support, i.e. B-splines are zero on a large set (Schumaker, 1981).

Thus, if $f(x)$ is a smooth function, it can be well approximated by a spline function $f^*(x)$ in the sense that $sup_{x\in\mathcal{X}}|f(x) - f^*(x)| \to 0$ as the number of knots of the spline tends to infinity. Hence, there is a set of basis functions $\psi_k(\cdot)$ (e.g. B-splines) and constants $c_k, \quad k = 1, \ldots, K$, such that

$$f(x) \approx f^*(x) = \sum_{k=1}^{K} c_k \psi_k(x), \tag{2.3}$$

where K depends on the number of knots and the order of the B-splines.[25]

2.2.2. *Wavelets*

Suppose we have a scaling function $\phi(x)$ and a wavelet $\psi(x)$ such that, defining the following translated and scaled transformations,

$$\widetilde{\phi}_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \quad \widetilde{\psi}_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \quad j, k \in \mathcal{Z}$$

we can obtain the collection $\{\widetilde{\phi}_{j,k}\} \cup \{\widetilde{\psi}_{j,k}\}_{j\geq l; k\in\mathcal{Z}}$ which forms an orthonormal basis of $L^2(\mathcal{R})$, for some coarse scale l. Daubechies [10] describes the way to construct compactly supported ϕ and ψ that generate an orthonormal system and have space-frequency localization.

Since we are interested in functions that are defined on the compact interval $[0, 1]$, it is necessary to consider an orthonormal system that spans $L^2[0, 1]$. The main idea is to periodize on $[0, 1]$ the above wavelets defined on $L^2(\mathcal{R})$ as given by Cohen et al.[5] where

$$\phi_{j,k}(x) = \sum_n \widetilde{\phi}_{j,k}(x - n), \quad \psi_{j,k}(x) = \sum_n \widetilde{\psi}_{j,k}(x - n)$$

are the periodized wavelets and generate a multiresolution level ladder $V_0 \subset V_1 \subset \ldots$, in which the spaces V_j are generated by the $\widetilde{\phi}_{j,k}$. From now on we use $\phi_{j,k}$ and $\psi_{j,k}$ instead of $\widetilde{\phi}_{j,k}$ and $\widetilde{\psi}_{j,k}$, respectively.

Accordingly, we can expand any function $f \in L^2[0, 1]$ in an orthogonal series

$$f(x) = \alpha_{0,0}\phi(x) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} c_{j,k}\psi_{j,k}(x) \tag{2.4}$$

where

$$\alpha_{0,0} = \int_0^1 f(x)\phi(x)dx, \quad c_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$$

are called the wavelet coefficients.

In practice, we approximate the expansion in (2.4) by the finite sumation

$$f(x) = \alpha_{0,0}\phi(x) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} c_{j,k}\psi_{j,k}(x), \tag{2.5}$$

where J is an arbitrary smoothing parameter. The choice of J is based on the expected smoothing degree of the function f, and how much of details can be ignored. In general, J is the highest resolution level such that $2^J \leq \sqrt{T} \leq 2^{J+1}$. In fact, as in all non-parametric smoothing, the variance of the approximation increases monotonically as a function of J. Nevertheless, low values of J lead to a highly biased approximation. See Ref. 12 for details.

3. Hypothesis Testing Procedure

Let $\{X_{t,T}, t = 1, \dots, T\}$ and $\{Y_{t,T}, t = 1, \dots, T\}$ be two locally stationary time series with zero mean, which come from a time-varying AR process with the same order p, i.e. they can be represented as

$$\begin{aligned} X_{t,T} &= \sum_{i=1}^{p} a_i\left(\frac{t}{T}\right) X_{t-i,T} + \varepsilon_t^x, \\ Y_{t,T} &= \sum_{i=1}^{p} b_i\left(\frac{t}{T}\right) Y_{t-i,T} + \varepsilon_t^y, \end{aligned} \tag{3.1}$$

with the corresponding assumptions for $\{\varepsilon_t^x\}, \{\varepsilon_t^y\}, a_i(u), b_i(u)$ with $u = \frac{t}{T}$, $i = 0, 1, \dots, p$, described in section 2.1. The processes $\{X_{t,T}, t = 1, \dots, T\}$ and $\{Y_{t,T}, t = 1, \dots, T\}$ can be correlated or not.

Our aim is to decide if the two series were generated by the same time-varying AR process, i.e. the hypotheses to be tested are

$$\begin{aligned} &H_0 : a_i(u) = b_i(u) \text{ for all } i = 1, \dots, p, \quad u \in (0,1) \\ &H_1 : a_i(u) \neq b_i(u) \text{ for either some } i = 1, \dots, p, \text{ or some } u \in (0,1). \end{aligned} \tag{3.2}$$

The procedure proposed by Maharaj [16] to compare two stationary AR processes cannot be extended directly due to the fact that in this case we have a number of parameters that is much larger than number of observations. To overcome this problem, we can expand the functions $a_i(u)$ and $b_i(u)$, $i = 1, \dots, p$, from (3.1) using (2.3) or (2.5). To illustrate the procedure, we will use wavelets. We obtain the following equations,

$$X_{t,T} = \sum_{i=1}^{p} \left(\alpha_{0,0}\phi\left(\frac{t}{T}\right) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} c_{j,k}^{(i)}\psi_{j,k}\left(\frac{t}{T}\right) \right) X_{t-i,T} + \varepsilon_t^x + s_t^x,$$

$$Y_{t,T} = \sum_{i=1}^{p} \left(\delta_{0,0}\phi\left(\frac{t}{T}\right) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} d_{j,k}^{(i)}\psi_{j,k}\left(\frac{t}{T}\right) \right) Y_{t-i,T} + \varepsilon_t^y + s_t^y \qquad (3.3)$$

where $\{\alpha_{0,0}^{(i)}, c_{j,k}^{(i)}\}$ and $\{\delta_{0,0}^{(i)}, d_{j,k}^{(i)}\}, 0 \leq j \leq J, 0 \leq k \leq 2^j - 1$, are the coefficients for the corresponding functions $a_i(u)$ and $b_i(u)$, $i = 1, \ldots, p$; ε_t^x, ε_t^y as before and s_t^x, s_t^y are errors due to the functional approximation (truncated expansions). However, these errors decay rapidly to zero as the time series length increases [8].

The last $T - p$ observations of the model fitted to $\{X_{t,T}, t = 1, \ldots, T\}$ series can be written in the matrix form

$$\mathbf{X} = \Psi_x \mathbf{c} + \varepsilon_x + \mathbf{s}_x, \qquad (3.4)$$

where

$$\begin{aligned}
\mathbf{X} &= (X_{p+1,T},\ X_{p+2,T},\ \ldots, X_{T-1,T},\ X_{T,T})', \\
\mathbf{c} &= (\alpha_{0,0}^{(1)}, c_{0,0}^{(1)}, \ldots, c_{J,2^J-1}^{(1)}, \alpha_{0,0}^{(2)}, c_{0,0}^{(2)}, \ldots, c_{J,2^J-1}^{(2)}, \ldots, \alpha_{0,0}^{(p)}, c_{0,0}^{(p)}, \ldots, c_{J,2^J-1}^{(p)})', \\
\varepsilon_x &= \left(\varepsilon_{p+1}^x, \varepsilon_{p+2}^x, \ldots, \varepsilon_T^x\right)',\ \mathbf{s}_x = \left(s_{p+1}^x, s_{p+2}^x, \ldots, s_T^x\right)',
\end{aligned}$$

and the Ψ_x matrix corresponds to $\Psi_x = (\Psi_x^{(1)}, \Psi_x^{(2)}, \ldots, \Psi_x^{(p)})$ with

$$\Psi_x^{(i)} = \begin{bmatrix}
\phi\left(\frac{p+1}{T}\right) X_{p+1-i,T} & \psi_{0,0}\left(\frac{p+1}{T}\right) X_{p+1-i,T} & \ldots & \psi_{J,2^J-1}\left(\frac{p+1}{T}\right) X_{p+1-i,T} \\
\phi\left(\frac{p+2}{T}\right) X_{p+2-i,T} & \psi_{0,0}\left(\frac{p+2}{T}\right) X_{p+2-i,T} & \ldots & \psi_{J,2^J-1}\left(\frac{p+2}{T}\right) X_{p+2-i,T} \\
\vdots & & \ddots & \vdots \\
\phi\left(\frac{T-1}{T}\right) X_{T-1-i,T} & \psi_{0,0}\left(\frac{T-1}{T}\right) X_{T-1-i,T} & \ldots & \psi_{J,2^J-1}\left(\frac{T-1}{T}\right) X_{T-1-i,T} \\
\phi\left(\frac{T}{T}\right) X_{T-i,T} & \psi_{0,0}\left(\frac{T}{T}\right) X_{T-i,T} & \ldots & \psi_{J,2^J-1}\left(\frac{T}{T}\right) X_{T-i,T}
\end{bmatrix},$$

and $i = 1, 2, \ldots, p$.

We can also represent the $T - p$ observations $Y_{p+1,T}, Y_{p+2,T}, \ldots, Y_{T-1,T}, Y_{T,T}$, from (3.3) by

$$\mathbf{Y} = \Psi_y \mathbf{d} + \varepsilon_y + \mathbf{s}_y, \qquad (3.5)$$

where the quantities $\mathbf{Y}$, Ψ_y, $\mathbf{d}, \varepsilon_y$ and $\mathbf{s}_y$ are similarly defined. Furthermore,

$$\begin{aligned}
E(\varepsilon_x) &= E(\varepsilon_y) = \mathbf{0}, \\
E(\varepsilon_x \varepsilon_x') &= Diag\left(\sigma_x^2\left(\frac{p+1}{T}\right), \sigma_x^2\left(\frac{p+2}{T}\right), \ldots, \sigma_x^2\left(\frac{T}{T}\right)\right) = \Sigma_{xx}, \\
E(\varepsilon_y \varepsilon_y') &= Diag\left(\sigma_y^2\left(\frac{p+1}{T}\right), \sigma_y^2\left(\frac{p+2}{T}\right), \ldots, \sigma_y^2\left(\frac{T}{T}\right)\right) = \Sigma_{yy}.
\end{aligned}$$

We will assume further that the innovations of the two models are contemporaneously correlated in time, i.e.

$$E(\varepsilon_x \varepsilon_y') = Diag\left(\sigma_{xy}\left(\frac{p+1}{T}\right), \sigma_{xy}\left(\frac{p+2}{T}\right), \ldots, \sigma_{xy}\left(\frac{T}{T}\right)\right) = \Sigma_{xy}.$$

The dimensions of $\mathbf{X}, \varepsilon_x, \mathbf{s}_x, \mathbf{Y}, \varepsilon_y$ and $\mathbf{s}_y$ are $(T-p)\times 1$, of $\mathbf{c}$ and $\mathbf{d}$ are $pL \times 1$, of Ψ_x and Ψ_y are $(T-p)\times pL$ and of each matrix $\Sigma_{..}$ is $(T-p)\times(T-p)$, where $L = 2^{J+1}$ corresponds to the number of wavelet coefficients used in each approximation.

Therefore, joining the models (3.4) and (3.5) we obtain

$$\mathbf{Z} = \boldsymbol{\Psi}\beta + \varepsilon + \mathbf{s},$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}, \ \boldsymbol{\Psi} = \begin{bmatrix} \Psi_x & \mathbf{0} \\ \mathbf{0} & \Psi_y \end{bmatrix}, \ \beta = \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}, \ \varepsilon = \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} \ \text{ and } \mathbf{s} = \begin{bmatrix} \mathbf{s}_x \\ \mathbf{s}_y \end{bmatrix},$$

with $E(\varepsilon) = \mathbf{0}$ and

$$E(\varepsilon\varepsilon') = \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{bmatrix}.$$

Thus, the generalized least squares estimator of β is given by

$$\hat{\beta} = \left[\Psi'\Sigma^{-1}\Psi\right]^{-1}\Psi'\Sigma^{-1}\mathbf{Z}.$$

The dimensions of $\mathbf{Z}, \mathbf{s}$ and ε are $2(T-p)\times 1$, of β is $2pL \times 1$ and of Ψ is $2(T-p)\times 2pL$.

Assuming that the process ε follows a multivariate Normal distribution with mean $\mathbf{0}$ and matrix of variances and covariances Σ, this model is a particular case of Sato et al. [23], and analogously it can be shown that $\sqrt{T}(\hat{\beta}-\beta)$ has asymptotic distribution $N(\mathbf{0}, \Sigma^*)$ where

$$\Sigma^* = lim_{T\to\infty} Var(\sqrt{T}\hat{\beta}) = plim\Big(\frac{\Psi'\Sigma^{-1}\Psi}{T}\Big)^{-1}.$$

Hence, the hypotheses given in (3.2) are equivalent to

$$H_0 : \mathbf{c} = \mathbf{d}$$
$$H_1 : \mathbf{c} \neq \mathbf{d}$$

or equivalently to

$$H_0 : \mathbf{C}\beta = \mathbf{0}$$
$$H_1 : \mathbf{C}\beta \neq \mathbf{0}$$

with $\mathbf{C} = [I_{pL} \quad -I_{pL}]$. It follows that $\sqrt{T}(\mathbf{C}\hat{\beta} - \mathbf{C}\beta)$ is asymptotically $N_{pL}(\mathbf{0}, \mathbf{C}\Sigma^*\mathbf{C}')$.

Defining now υ by

$$\upsilon = [\mathbf{C}(\Psi'\Sigma^{-1}\Psi)^{-1}\mathbf{C}']^{-\frac{1}{2}}(\mathbf{C}\hat{\beta} - \mathbf{C}\beta),$$

it is straightforward that under H_0, υ has asymptotically a $N(\mathbf{0}, I_{pL})$. Consequently under H_0, the statistic

$$W = \upsilon'\upsilon = (\mathbf{C}\hat{\beta})'[\mathbf{C}(\Psi'\Sigma^{-1}\Psi)^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\beta})$$

has asymptotically a $\chi^2_{(pL)}$ distribution. Hence, for a given significance level α, we reject H_0 if $P(W > w) < \alpha$, where w is the observed value of the W statistic.

Nevertheless, all the procedures described previously assume that the time-varying variance and covariance structures of ε_x and ε_y are known. In practice, these structures have also to be estimated.

Since we assume that the innovations ε^x_t have mean zero and variance $\sigma^2_x(t/T)$, a reasonable estimator for $\sigma^2_x(t/T)$ is the squared residual $r^2_x(t/T)$ obtained in the least square estimation procedure. Analogously, an estimator for $\sigma^2_y(t/T)$ and $\sigma_{xy}(t/T)$ are given by $r^2_y(t/T)$ and $r_x(t/T)r_y(t/T)$ respectively. Thus, to estimate the matrix Σ we initially consider the truncated wavelets expansion for $\sigma^2_x(t/T), \sigma^2_y(t/T)$ and $\sigma_{xy}(t/T)$ as follows

$$\sigma^2_x(\frac{t}{T}) = \alpha^x_{0,0}\phi(\frac{t}{T}) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} c^x_{j,k}\psi_{j,k}(\frac{t}{T}),$$

$$\sigma^2_y(\frac{t}{T}) = \alpha^y_{0,0}\phi(\frac{t}{T}) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} c^y_{j,k}\psi_{j,k}(\frac{t}{T}),$$

$$\sigma_{xy}(\frac{t}{T}) = \alpha^{xy}_{0,0}\phi(\frac{t}{T}) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} c^{xy}_{j,k}\psi_{j,k}(\frac{t}{T}).$$

The criterion for choosing the maximum scale J is the same described previously. Then, we apply a linear regression using the truncated wavelet expansion as explanatory variables and the series $r^2_x(t/T)$, $r^2_y(t/T)$ and $r_x(t/T)r_y(t/T)$ as responses, in order to obtain consistent estimates of $\sigma^2_x(t/T)$, $\sigma^2_y(t/T)$ and $\sigma_{xy}(t/T)$.

Finally, similar to Ref. 23, we propose a generalization of the Cochrane and Orcutt [4] procedure given by the following iterative estimation algorithm:

i) Apply the generalized least square estimation considering Σ equal to the identity matrix;

ii) obtain the residuals $r_x(t/T)$ and $r_y(t/T)$, $t = 1, \ldots, T$;

iii) smooth the squared and cross residuals considering the wavelet linear regression obtaining an estimate of Σ;

iv) apply the generalized least square estimation considering the estimated matrix of Σ;

v) return to 2 until numerical convergence of parameters or a maximum number of iterations.

4. Simulations

In this section, the results of some sets of simulations are presented. The simulations were used to evaluate the adequacy of asymptotic null distribution and the performance of the proposed test. All simulations were performed using the package *splines* and *wavethresh* of the statistical software R. We consider the Haar(D1), D2 and D8 (*Double Extreme Phase with periodic boundary condition*) wavelets.

The first group of simulations evaluates the estimation procedure and is based on a time-varying autoregressive model of order one, with coefficients

$$b_1(\frac{t}{T}) = a_1(\frac{t}{T}) = \sin\left(\frac{2\pi t}{T}\right)/3. \tag{4.1}$$

The processes generated in the simulation are given by

$$X_{t,T} = a_1\left(\frac{t}{T}\right) X_{t-1,T} + \varepsilon_t^x,$$
$$Y_{t,T} = b_1\left(\frac{t}{T}\right) Y_{t-1,T} + \varepsilon_t^y$$

where ε_t^x and ε_t^y are Gaussian white noises, and

$$\sigma_x^2\left(\frac{t}{T}\right) = \left(1 + \frac{\cos(\frac{2\pi t}{T})}{4}\right)^2,$$
$$\sigma_y^2\left(\frac{t}{T}\right) = 0.25\left(1 + \frac{\cos(\frac{2\pi t}{T})}{4}\right)^2 + 0.25\left(1 + \frac{\cos(\frac{2\pi t}{T})}{4}\right)^6,$$
$$\sigma_{xy}\left(\frac{t}{T}\right) = 0.5\left(1 + \frac{\cos(\frac{2\pi t}{T})}{4}\right)^4.$$

The results of the generalized least square estimation procedure for $T = 256, L = 8$ show a good performance of the estimation procedure. The results are presented in Fig. 1, 2, 3 and 4 and are based on 1000 simulations. The dotted lines represent the confidence interval of one standard deviation. The functions with support in $[0, 1]$ are rescaled to the original interval $[1, T]$. The expectation for all the estimated curves are close to the theoretical ones. In terms of low frequency, the results are reasonable even using the Haar and D2 wavelets. However, the estimatives using B-splines show a variability higher than the others at the boundaries.

In order to evaluate the asymptotic null distribution of the proposed statistic, one thousand of time-varying autoregressive series of order one were simulated, considering coefficients $b_1\left(\frac{t}{T}\right) = a_1\left(\frac{t}{T}\right)$, where $a_1\left(\frac{t}{T}\right)$, ε_t^x and ε_t^y were the same as in the previous simulations. The proposed Wald test was applied for each generated series ($L = 4$, $T = 128$ and the $D8$ wavelet) and the histogram, kernel density estimates and theoretical distribution are presented in Fig. 5.

Focusing on the evaluation of suggested Wald test power, we generate time-varying autoregressive series of order one, with coefficients $b_1\left(\frac{t}{T}\right) = (1+\lambda)a_1\left(\frac{t}{T}\right)$,

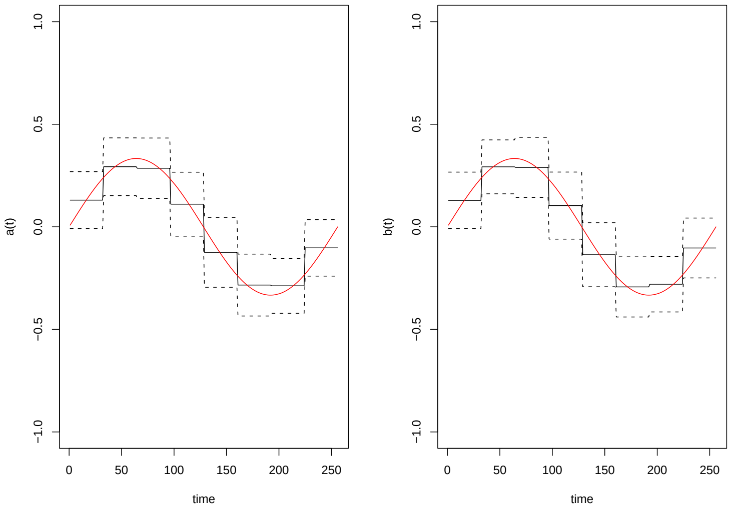

Fig. 1. Theoretical and estimated autoregressive functions using the Haar wavelet basis. The solid lines describe the average of estimated curves and theoretical, respectively. The dashed lines describe the confidence interval of one standard deviation.

where $\lambda \in \{0, 0.2, 0.4, ..., 1.8\}$ and $a_1\left(\frac{t}{T}\right)$, ε_t^x and ε_t^y as before. Hence, the difference between the two autoregressive structures increases as the value of λ increases. For $\lambda = 0$ we can evaluate the size of the test. The null hypotheses is that the two series have the same time-varying autoregressive structure. Table 1 presents the results for the different bases with $T \in \{64, 128, 256\}$; $L \in \{4, 8, 16\}$ and 1000 simulations.

In general, the proposed Wald test for comparing the two autoregressive structures has a good performance. The effect of the class of basis functions on the Wald test is illustrated in Figure 6 ($L = 8$ and $T = 128$). Despite the fact that B-splines seem to be more powerful, the sample length seems to be not enough for a good approximation of the Wald statistic to the asymptotic distribution. Assuming the size of the test $\alpha = 0.05$, the test based on the B-splines has a proportion of rejections of 0.091. On the other hand, the tests based on wavelets lead to an acceptable asymptotic approximation, and they have almost the same performance. Figure 7 shows the effect of the number (L) of basis functions considered in the Wald test, for the D8 wavelet and $T = 128$. The figure illustrates that the power of the test decreases as L increases. The estimation bias is certainly reduced as L increases, but the number of degrees of freedom is strongly reduced, resulting on less power. As expected, Figure 8 shows that the Wald test power increases as the sampling rate increases. The simulation results described in Table 1 point towards a similar power and performance of the test for all the basis functions considered, for a large

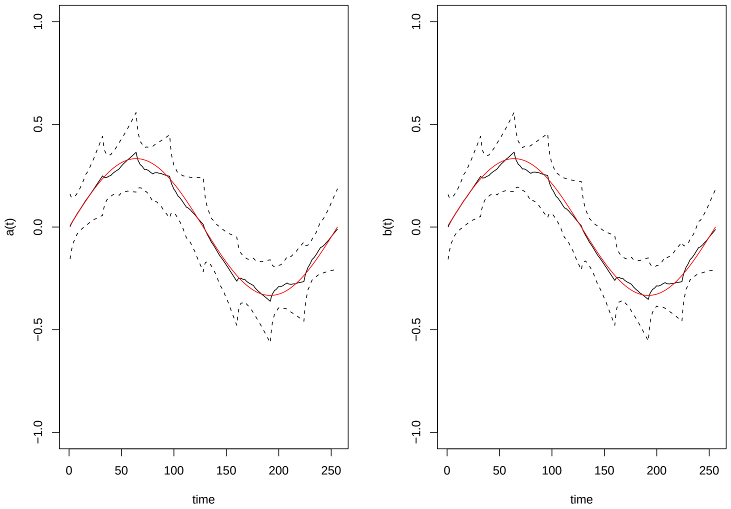

Fig. 2. Theoretical and estimated autoregressive functions using the D2 wavelet basis. The solid lines describe the average of estimated curves and theoretical, respectively. The dashed lines describe the confidence interval of one standard deviation.

sampling rate. However, the test based on wavelet bases seems to have a better performance for small samples.

Furthermore, we are also interested in evaluate the performance of the test in the case of time-variant AR(2) processes. Thus, we considered the following processes:

$$
\begin{aligned}
X_{t,T} &= a_1\left(\frac{t}{T}\right) X_{t-1,T} + a_2\left(\frac{t}{T}\right) X_{t-2,T} + \varepsilon_t^x, \\
Y_{t,T} &= \left(b_1\left(\frac{t}{T}\right) + \lambda\right) Y_{t-1,T} + \left(b_2\left(\frac{t}{T}\right) + \lambda\right) Y_{t-2,T} + \varepsilon_t^y
\end{aligned}
$$

where ε_t^x and ε_t^y are Gaussian white noises with $\sigma_x^2(\cdot), \sigma_y^2(\cdot)$ and $\sigma_{xy}(\cdot)$ as before.

The power of the test was evaluated using 1000 simulations considering $L \in \{4, 8\}$, $T \in \{128, 256\}$ and $\lambda \in \{0, 0.1, 0.2, 0.3, 0.4\}$. It is important to highlight that when $\lambda = 0$, the two processes have the same AR(2) structure. The results of this set of simulations are shown in Table 2.

Table 2 suggests that the approach is adequate to test the structure of AR(2) processes. However, as the number of parameters to be estimated is the double compared to the AR(1) case, the length of observed time series must be larger for a good approximation of the Wald statistic asymptotic distribution.

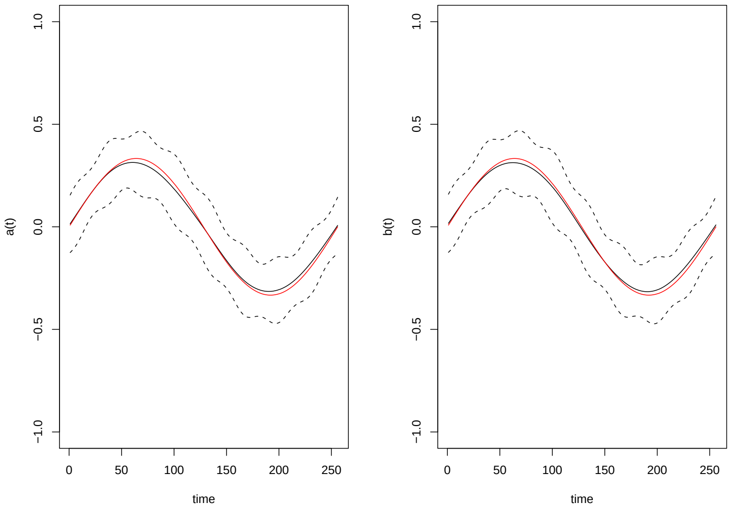

Fig. 3. Theoretical and estimated autoregressive functions using the D8 wavelet basis. The solid lines describe the average of estimated curves and theoretical, respectively. The dashed lines describe the confidence interval of one standard deviation.

5. Application to Real Data

In this section two applications of the proposed test are shown. In the first one, we considered financial time series from Italy and USA stock markets. The second example refers to the analysis of brain signals from two subjects, acquired in a functional magnetic resonance imaging experiment.

5.1. *Application 1*

The volatility structure of financial assets is very useful to quantify risks. Their predictions are frequently used to determine the value at risk of portfolios, which in many countries, are required by government financial institutions. Define the log-volatility time series as

$$z_t = log(r_t^2 + \gamma), \tag{5.1}$$

where r_t is the log-return of prices or indexes, and γ is an arbitrary constant greater than zero. This constant is necessary due to the fact that the log-returns may assume zero values.

This illustrative application is based on the analysis of log-volatilities of the national stock market indexes from Italy and USA. We considered $\gamma = 0.001$, and daily log-returns from September first 1999 to August twelve 2003, resulting on a

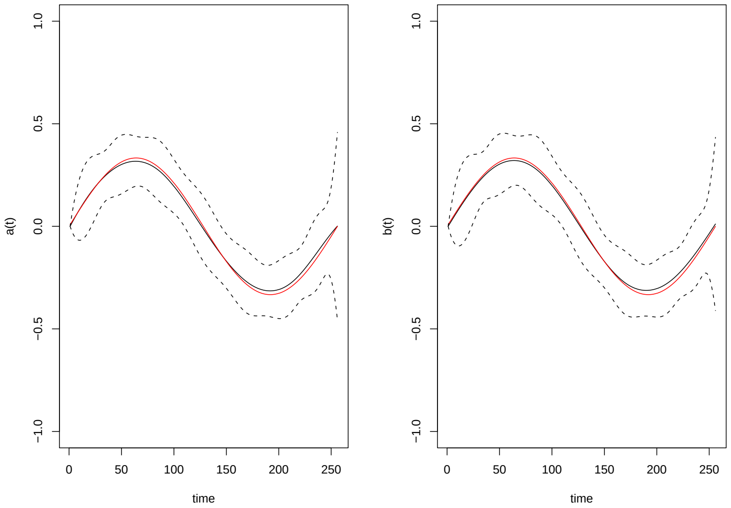

Fig. 4. Theoretical and estimated autoregressive functions using the B-splines basis. The solid lines describe the average of estimated curves and theoretical, respectively. The dashed lines describe the confidence interval of one standard deviation.

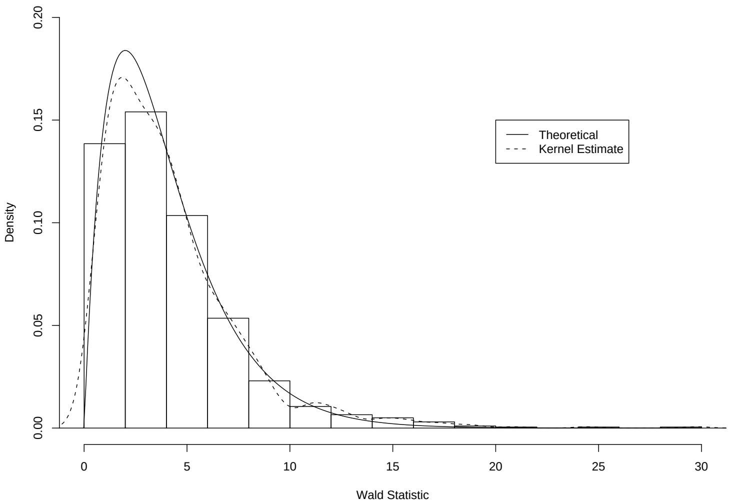

Fig. 5. Histogram, kernel density estimates and theoretical distribution of the Wald statistics under the null hypothesis.

Table 1. Power estimates ($\alpha = 0.05$) of the Wald test for common autoregressive function.

T	L	Wavelet	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8
64	4	Haar	0.049	0.053	0.081	0.129	0.227	0.371	0.486	0.674	0.792	0.889
		D2	0.057	0.064	0.090	0.140	0.233	0.368	0.529	0.680	0.787	0.892
		D8	0.047	0.073	0.092	0.148	0.260	0.429	0.604	0.739	0.858	0.925
		BS	0.070	0.066	0.107	0.191	0.278	0.415	0.522	0.690	0.819	0.870
	8	Haar	0.057	0.072	0.087	0.116	0.189	0.255	0.385	0.509	0.646	0.793
		D2	0.105	0.110	0.132	0.158	0.232	0.316	0.404	0.510	0.678	0.733
		D8	0.110	0.106	0.128	0.172	0.221	0.308	0.409	0.506	0.677	0.722
		BS	0.125	0.132	0.160	0.195	0.249	0.322	0.430	0.523	0.657	0.764
	16	Haar	0.152	0.194	0.182	0.196	0.249	0.298	0.362	0.437	0.529	0.621
		D2	0.166	0.142	0.185	0.186	0.215	0.252	0.299	0.398	0.459	0.531
		D8	0.181	0.194	0.206	0.206	0.256	0.273	0.323	0.391	0.472	0.546
		BS	0.203	0.213	0.201	0.238	0.277	0.308	0.335	0.440	0.434	0.595
128	4	Haar	0.042	0.048	0.130	0.239	0.452	0.679	0.869	0.956	0.991	0.998
		D2	0.032	0.051	0.127	0.235	0.464	0.692	0.873	0.966	0.984	0.995
		D8	0.038	0.052	0.138	0.296	0.550	0.758	0.914	0.975	0.994	1.000
		BS	0.065	0.090	0.135	0.276	0.473	0.711	0.886	0.966	0.986	0.998
	8	Haar	0.038	0.059	0.093	0.172	0.338	0.549	0.752	0.896	0.973	0.995
		D2	0.063	0.056	0.118	0.190	0.352	0.520	0.744	0.891	0.976	0.983
		D8	0.049	0.049	0.117	0.184	0.327	0.538	0.773	0.896	0.953	0.991
		BS	0.083	0.093	0.139	0.226	0.391	0.593	0.756	0.887	0.949	0.976
	16	Haar	0.069	0.056	0.082	0.175	0.245	0.377	0.540	0.750	0.852	0.941
		D2	0.145	0.138	0.182	0.208	0.286	0.406	0.583	0.709	0.825	0.908
		D8	0.138	0.130	0.187	0.239	0.286	0.426	0.560	0.753	0.816	0.900
		BS	0.147	0.162	0.190	0.243	0.332	0.459	0.588	0.713	0.856	0.912
256	4	Haar	0.042	0.052	0.198	0.490	0.801	0.946	0.990	1.000	1.000	1.000
		D2	0.039	0.055	0.205	0.531	0.851	0.958	0.997	1.000	1.000	1.000
		D8	0.039	0.072	0.251	0.591	0.869	0.982	0.999	1.000	1.000	1.000
		BS	0.043	0.086	0.218	0.534	0.822	0.977	0.995	0.999	0.999	1.000
	8	Haar	0.028	0.051	0.136	0.371	0.699	0.904	0.987	1.000	1.000	1.000
		D2	0.023	0.053	0.163	0.394	0.675	0.912	0.984	0.997	0.999	1.000
		D8	0.013	0.032	0.139	0.377	0.688	0.928	0.982	1.000	1.000	1.000
		BS	0.091	0.178	0.407	0.738	0.912	0.983	0.996	0.994	0.997	0.997
	16	Haar	0.022	0.050	0.108	0.226	0.482	0.752	0.944	0.987	1.000	1.000
		D2	0.046	0.079	0.132	0.317	0.526	0.776	0.924	0.979	0.985	0.989
		D8	0.043	0.045	0.126	0.262	0.510	0.771	0.945	0.988	0.998	1.000
		BS	0.076	0.086	0.157	0.312	0.524	0.784	0.919	0.977	0.990	0.991

Table 2. Power estimates ($\alpha = 0.05$) of the Wald test in the AR(2) Model.

T	L	Basis	0	0.1	0.2	0.3	0.4
128	4	D1	0.023	0.155	0.655	0.958	0.999
		D2	0.033	0.168	0.668	0.976	0.999
		D8	0.035	0.162	0.675	0.974	1.000
		BS	0.077	0.218	0.696	0.957	0.985
	8	D1	0.023	0.093	0.456	0.863	0.996
		D2	0.044	0.131	0.430	0.851	0.984
		D8	0.082	0.167	0.468	0.855	0.969
		BS	0.087	0.162	0.476	0.846	0.966
256	4	D1	0.019	0.325	0.969	1.000	1.000
		D2	0.032	0.349	0.979	1.000	1.000
		D8	0.020	0.379	0.969	0.999	1.000
		BS	0.072	0.411	0.974	0.994	0.997
	8	D1	0.019	0.198	0.857	0.999	1.000
		D2	0.026	0.183	0.896	1.000	1.000
		D8	0.047	0.247	0.879	0.988	0.996
		BS	0.061	0.252	0.889	0.995	0.996

time series of extension 1025. Before the analysis, the log-volatilities were standardized to mean zero and unit variance by subtracting the average and dividing by the

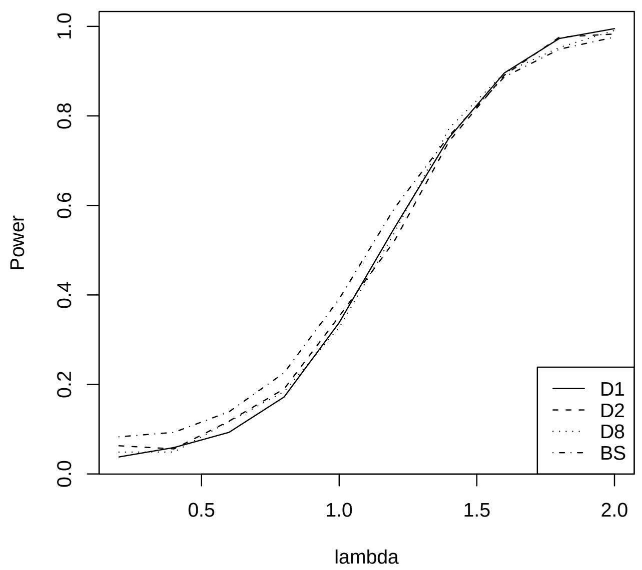

Fig. 6. Basis functions effect on the Wald test power, using $T = 128$ and $L = 8$.

standard deviation. The standardized log-volatilities are shown in Figure 9.

The time-varying autoregressive estimates (AR(1)) are shown in Figures 10 and 11. Assuming $L = 4$, the p-value of test using the wavelets Haar, D2, D8 and the B-splines basis are 0.972, 0.004, 0.040 and < 0.001, respectively. Considering the case when $L = 8$, the respective p-values are 0.050, 0.123, 0.005 and 0.015.

It is important to notice that for this application, the decision of the test depends on the basis chosen and on the number of functions considered in the expansion. Anyway, analyzing all the results, it is reasonable to conclude that the two autoregressive structures are different. The test using the Haar wavelet for $L = 4$ was the only one suggesting that the AR structures were not different ($p = 0.972$). However the p-value of the test decreased to $p = 0.050$ for $L = 8$. This result suggests that the structures are different but $L = 4$ is too small for Haar. On the other hand, using $L = 8$ may lead to a decrease in the power of the test considering D2 and B-splines, as the number of parameters to be estimated increased.

In conclusion, considering a test size of $\alpha = 0.05$, we reject the hypothesis that the two time series have the same autoregressive structure.

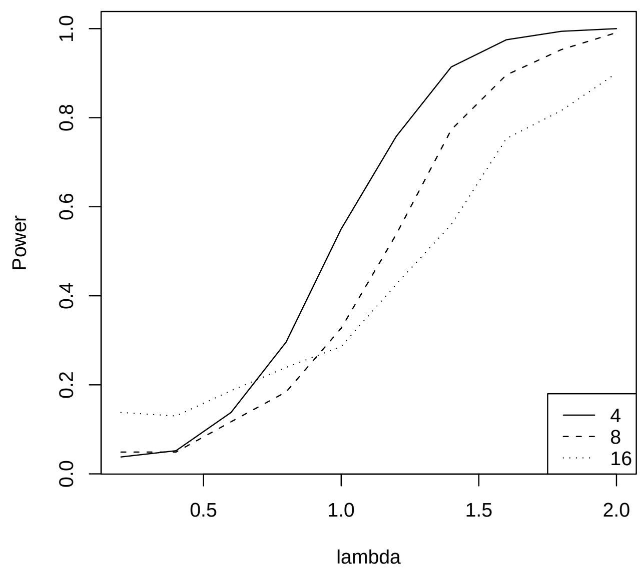

Fig. 7. Number of basis functions (L) effect on the Wald test power, using $D8$ and $T = 128$.

5.2. *Application 2*

In this second example, the data was acquired in an experiment of functional magnetic resonance imaging (fMRI). The BOLD signal (blood oxigenation level dependent) observed in fMRI sessions can be considered as a indirect measure of local neural activity. In this experiment, faces were randomly presented to the subjects, who should decide if they were neutral, sad or very sad faces.

During this experiment, both subjects activated a brain region named Insula. We are interested in verify if the Insulas' BOLD signals of the two participants have the same autoregressive structure, as they were submitted to the same sequence of faces' presentation (stimulation). Figure 12 shows the BOLD time series.

In this application, the time series extension was 129 points. The signals were normalized to mean zero and variance one. The test of same time-variant autoregressive structure (AR(1)) was applied to these series, considering the Haar, D2, D8 wavelets, B-splines for $L = 4$ and $L = 8$. The estimated structures are shown in Figure 13 and 14. Considering $L = 4$, the p-values for the test using Haar, D2, D8 and B-splines are 0.013, 0.014, 0.003, and 0.001, respectively. Assuming $L = 8$, the p-values are 0.081, 0.006, 0.028 and 0.017, respectively. Note that in this case, the results suggest that autoregressive structures are not equal. In fact, it seems that they have a similar behaviour, but the autoregressive coefficient is greater in

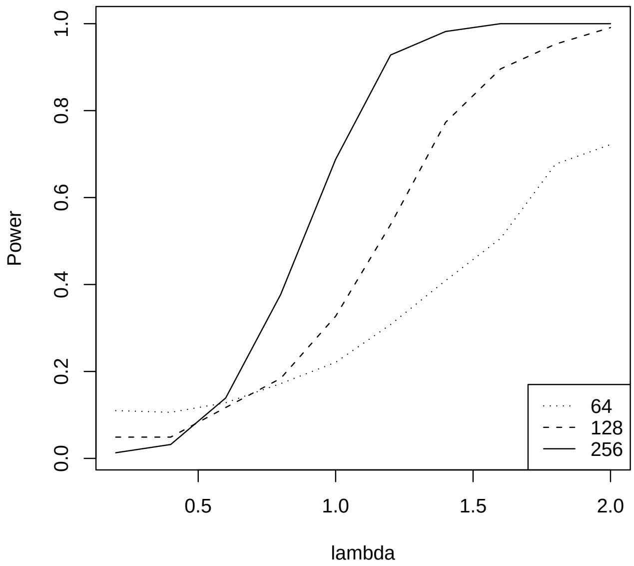

Fig. 8. Sampling rate ($1/T$) effect on the Wald test power, using $D8$ and $L = 8$.

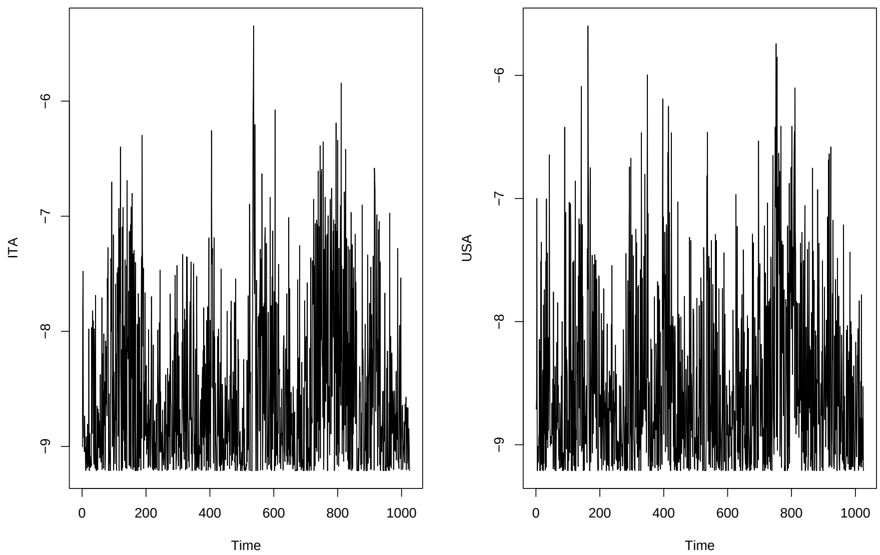

Fig. 9. Log-volatility time series from Italy and USA stock indexes.

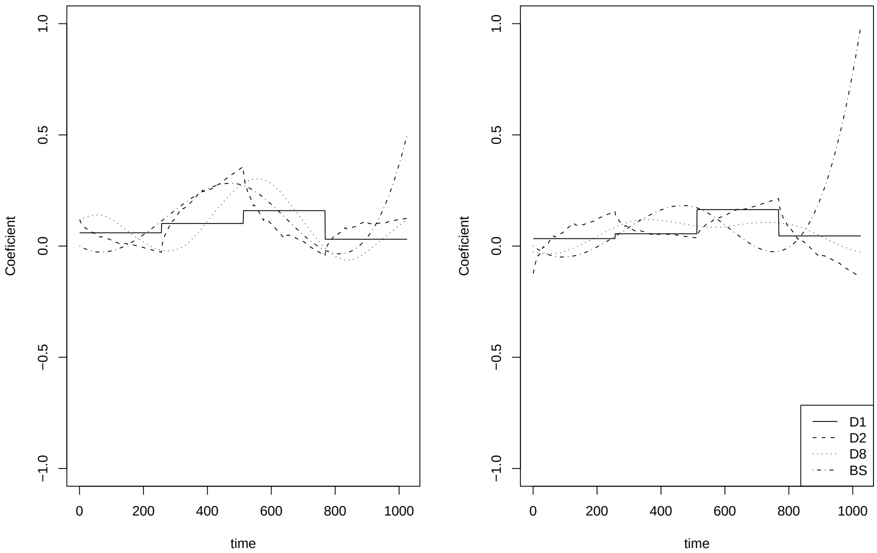

Fig. 10. Estimated time-varying autoregressive structures (AR(1)) of the log-volatilities using $L = 4$.

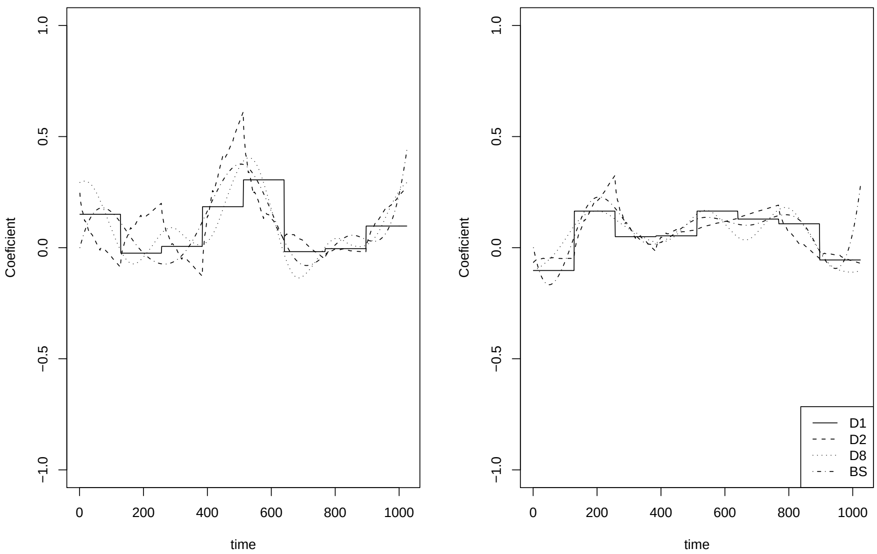

Fig. 11. Estimated time-varying autoregressive structures (AR(1)) of the log-volatilities using $L = 8$.

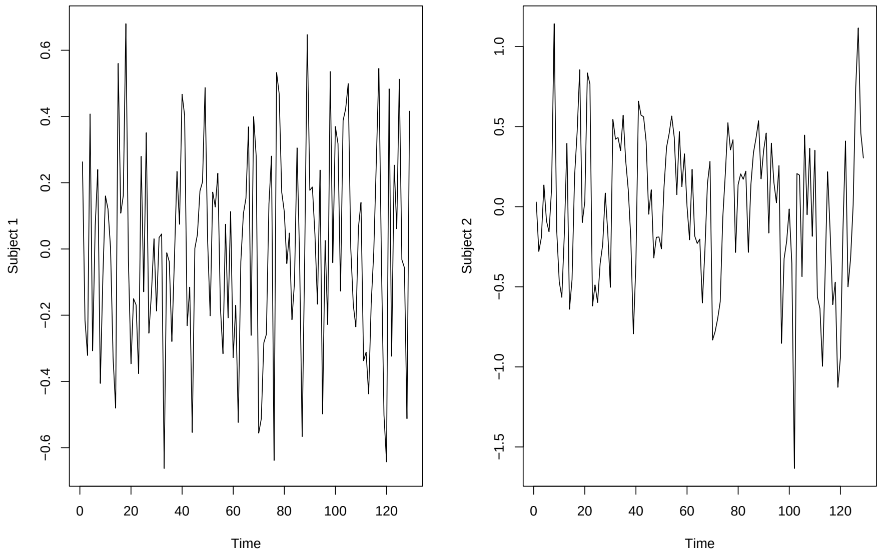

Fig. 12. Insulas' BOLD signal from the two subjects.

subject 2. Further, using $L = 8$ we observe a decrease in the power, indicating that a larger expansion is not necessary.

In summary, we conclude that the time-variant autoregressive structure in Insula is not the same for the two subjects, even if they were submitted to the same stimulus at the same time intervals.

6. Conclusions

In the context of locally stationary time series with autoregressive representation and time-varying parameters, we proposed a statistical test to evaluate the similarity of two locally stationary series comparing their functional autoregressive parameters. Simulation studies provide evidence that the proposed test performs well and has a better performance if we use the wavelet expansion for the estimation procedure of the autoregressive functions. The application to real data demonstrates that this test can be successfully applied in cases of locally stationary time series. Furthermore, the test can be easily extended to the multivariate case.

Acknowledgments

GES was supported by a CNPq grant and University of Quindío; PAM and CMCT were partially supported by FAPESP grant 03/10105-2. JRS was supported by a

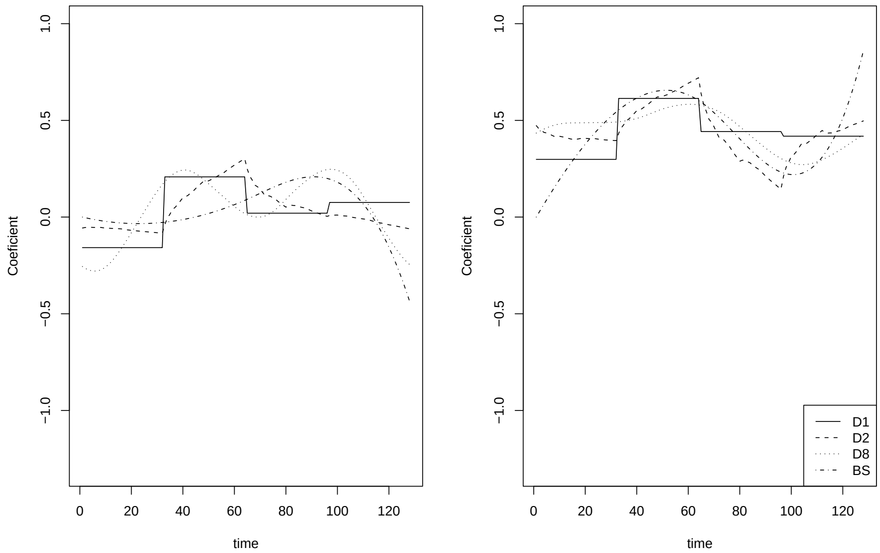

Fig. 13. Estimated time-varying autoregressive structures (AR(1)) of the BOLD signals, using $L = 4$.

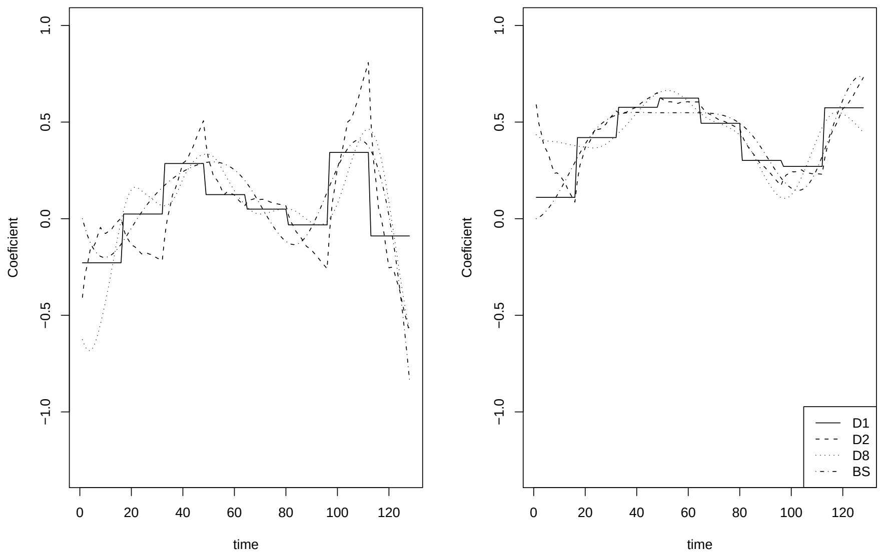

Fig. 14. Estimated time-varying autoregressive structures (AR(1)) of the BOLD signals, using $L = 8$.

CNPq grant. We are grateful to Dr. Elisson Fernando Cardoso for providing the fMRI dataset.

References

1. I.V. Basawa, L. Billard and R. Srinivasan, Large-sample tests of homogeneity for time series models, *Biometrika*, **71** (1984) 203–206.
2. J. Caiado, N. Crato and D. Peña, A periodogram-based metric for time series classification, *Computational Statistics and Data Analysis*, **50** (2006) 2668–2684.
3. D.S. Coates and P.J. Diggle, Test for comparing two estimated spectral densities, *Journal Time Series Analysis*, **7** (1986) 7–20.
4. D. Cochrane and G.H. Orcutt, Application of least squares regression to relationships containing autocorrelated error terms, *Journal of American Statistiscal Association*, **44** (1949) 32–61.
5. A. Cohen, I. Daubechies and P. Vial, Wavelets on the interval and fast wavelet transforms, *Applied and Computational Harmonic Analysis*, **1** (1993) 54–81.
6. R. Dahlhaus, On the Kullback-Leibler information divergence of locally stationary processes, *Stochastic Processes and Applications*, **62** (1996) 139–168.
7. R. Dahlhaus, Fitting time series models to nonstationary processes, *The Annals of Statistics*, **25** (1997) 1–37.
8. R. Dahlhaus, M.H. Newmann and von R. Sachs, Nonlinear wavelet estimation of time-varying autoregressive processes, *Bernoulli*, **5** (1999) 873–906.
9. P.J. Diggle and N.I. Fisher, Nonparametric comparison of cumulative periodograms, *Applied Statistics*, **40** (1991) 423–434.
10. I. Daubechies, *Ten lectures on wavelets*, (CBMS Lecture Notes Series, SIAM, 1992).
11. J.H. Guo, A non-parametric test for the parallelism of two first-order autoregressive processes, *Australian and New Zeland Journal of Statistics*, **41** (1999) 59–65.
12. W. Hardle, G. Kerkyacharian, D. Picard and A. Tsybakov, *Wavelets, Aproximation and Statistical Applications*, (Springer, 1998).
13. H.Y. Huang, H. Ombao and D.S. Stoffer, Discrimination and classification of Nonstationary time series using the SLEX model, *Journal of American Statistical Association*, **99** (2004) 763–774.
14. E.A. Maharaj, A significance test for classifying ARMA models, *Journal of Statistical Computation and Simulation*, **54** (1996) 305–331.
15. E.A. Maharaj, Comparison and classification of stationary multivariate time series, *Pattern Recognition*, **32** (1999) 1129–1138.
16. E.A. Maharaj, Clusters of time series, *Journal of Classification*, (2000) 297–314.
17. E.A. Maharaj, Comparison of non-stationary time series in the frequency domain, *Computational Statistics and Data Analysis*, **40** (2002) 131–141.
18. E.A. Maharaj, Using wavelets to compare time series patterns, *International Journal of Wavelets Multiresolution and Information Processing*, **3** (2005).
19. D. Piccolo, A distance measure for classifying ARIMA models, *Journal of Time Series Analysis*, **11** (1990) 652–164.
20. M.B. Priestley, *Spectral analysis and time series*, (Academic Press, 1981).
21. K. Sakiyama and M. Taniguchi, Discriminant analysis for locally stationary processes, *Journal of Multivariate Analysis*, **90** (2004) 282–300.
22. G.E. Salcedo and C.M.C. Toloi, Tests for comparing time series: Applications to water temperature and salinity measure at different depths, *Brazilian Journal of Statistics*, **61** (2000) 51–80.
23. J.R. Sato, P.A. Morettin, P.R. Arantes and E. Amaro Jr, Wavelet based time-varying

vector autoregressive models, *Computational Statistics and Data Analysis*. In Press, online publication Jan/2007.

24. R.H. Shumway, Time-frequency clustering and discriminant analysis, *Statistics and Probability Letters*, **63** (2003) 307–314.
25. L.L Schumaker, *Spline functions: Basic Theory*, (Wiley, 1981).