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 random disturbance component 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 usefulness 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

Inferences about time series structures have 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.\(^1,8\) As a classification problem, group of series can be discriminated in different clusters according to the degree of similarity.\(^10,11\) 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. As a financial application, Maharaj compared interest rates patterns among various countries.

Most of the existing techniques are applicable to stationary time series or to non-stationary time series that can be transformed to stationary by some transformation as differencing. On the other hand, these statistical tests are basically based on comparing either the spectral densities of pair of time series, or the coefficients of an autoregressive (AR) model.

Since many phenomena in applied sciences show a non-stationary behavior, for example, the second order structure changes over time, these processes are not easily transformable to stationarity. In order to compare series that are non-stationary in variance, Maharaj proposed two non-parametric tests which compare the evolutionary spectra and the sums of squared wavelet coefficients at different times or scales, respectively.

On the other hand, as certain class of non-stationary series (locally stationary series) can be represented by time-varying AR models, in this paper we propose a new approach to compare the autoregressive structures of the locally stationary time series. Further, this new approach also includes the case where the variance and covariances of the random disturbances change over time.

In section 2, a brief description about both locally stationary processes and approximation functions theory 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. Further, the asymptotic properties of autoregressive estimators are well understood. A more general class of models is the time-varying autoregressive class, in the context of locally stationary processes. It is given by the following difference equation:

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

(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.
2.1. Locally Stationary Processes

Stationarity has always been the 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 considered processes having a time-varying spectral representation

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

with an orthogonal increment process \( \xi(\omega) \) and a time-varying transfer function \( A_t(\omega) \). Nevertheless, 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 \( a_t(\cdot) \) at each time point.

Dahlhaus defined the following more general 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 if there is a representation

\[ X_{t,T} = \mu \left( \frac{t}{T} \right) + \int_{-\pi}^{\pi} e^{i\omega t} A_x \left( \frac{t}{T}, \omega \right) d\xi(\omega), \quad (2.2) \]

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.
  \[ \text{Cov}[d\xi(\omega), d\xi(\omega')] = \delta(\omega - \omega') \]
  such that
  \[ \text{Cum}\{d\xi(\omega_1), \ldots, d\xi(\omega_k)\} = \eta \left( \sum_{j=1}^{k} \omega_j \right) g_k(\omega_1, \ldots, \omega_{k-1}) d\omega_1 \ldots d\omega_k, \]
  where \( \text{Cum}\{\ldots\} \) denotes the cumulant of \( k - \text{th} \) order, \( g_1(\omega) = 0 \), \( g_2(\omega) = 1 \), \( |g_k(\omega_1, \ldots, \omega_{k-1})| \leq \text{const}_k \) for all \( k \) and \( \eta(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j) \) is the period \( 2\pi \) extension of the Dirac delta function.

- \( A_x(u, \omega) \) is a function on \([0, 1] \times [-\pi, \pi]\) which is \( 2\pi \)-periodic in \( \omega \), with \( A_x(u, -\omega) = A_x(u, \omega) \).

The functions \( A_x(u, \omega) \) and \( \mu(\omega) \) are assumed to be smooth in \( u \), in order to guarantee that the process has a locally stationary behavior. The evolutionary spectrum of \( X_{t,T} \) is defined as \( f_x(u, \omega) = |A_x(u, \omega)|^2 \).
It is not difficult to prove that in (2.1), $X_{t,T}$ has the representation (2.2) with
\[
A_x(u,\omega) = \frac{\sigma(u)}{\sqrt{2\pi}} \left( 1 + \sum_{j=1}^{p} a_j(u)e^{-ij\omega} \right)^{-1},
\]
where
\[
\varepsilon_t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} d\xi(\omega), \quad t \in \mathbb{Z}.
\]
For simplicity, in the sequel we assume that $\mu(u) = 0$.

### 2.2. Function Expansions

In most of the approximation functions theory, 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 $L^2(\mathbb{R})$.

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 $\chi$ with knot points $x_0 < x_1 < \ldots < x_{M+1}$, where $x_0$ and $x_{M+1}$ are the two end points of $\chi$, 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.

Thus, if $f(x)$ is a smooth function, it can be well approximated by a spline function $f^*(x)$ in the sense that $\sup_{x \in \chi} |f(x) - f^*(x)| \to 0$ 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 $c_k, \ k = 1, \ldots, K$, such that
\[
f(x) \approx f^*(x) = \sum_{k=1}^{K} c_k \psi_k(x), \quad (2.3)
\]
where $K$ depends on the number of knots and the order of the B-splines.\(^{16}\)

### 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,

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

we can obtain the collection \{$(\tilde{\phi}_{j,k}) \cup \{\tilde{\psi}_{j,k}\}$\}_j \geq l; k \in \mathbb{Z} which forms an orthonormal basis of $L^2(\mathbb{R})$, for some coarse scale $l$. Daubechies\(^7\) describes the way to construct compactly supported $\phi$ and $\psi$ that generate an orthonormal system and have space-frequency localization.

Since we are interested in functions that are defined in 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 $L^2(\mathbb{R})$ as given by Ref. 3 where

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

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 $\psi_{j,k}$, respectively.

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

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

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

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

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 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.
3. Hypothesis Testing Procedure

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

\[
X_{t,T} = \sum_{i=1}^{p} a_i \left( \frac{t}{T} \right) X_{t-i,T} + \varepsilon^x_t,
\]

\[
Y_{t,T} = \sum_{i=1}^{p} b_i \left( \frac{t}{T} \right) Y_{t-i,T} + \varepsilon^y_t,
\]

(3.1)

with the corresponding assumptions for \( \{ \varepsilon^x_t \}, \{ \varepsilon^y_t \}, a_i(u), b_i(u) \) with \( u = \frac{t}{T}, i = 0, 1, \ldots, p \), described in section 2.1. The processes \( \{ X_{t,T}, t = 1, \ldots, T \} \) and \( \{ Y_{t,T}, t = 1, \ldots, T \} \) 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

\[
H_0 : a_i(u) = b_i(u) \text{ for all } i = 1, \ldots, p, \ u \in (0, 1)
\]

\[
H_1 : a_i(u) \neq b_i(u) \text{ for either some } i = 1, \ldots, p, \text{ or some } u \in (0, 1).
\]

(3.2)

The procedure proposed by Ref. 11 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 \( b_i(u) \), \( i = 1, \ldots, p \), from (3.1) using (2.3) or (2.5). To illustrate the procedure, we will use wavelets. We obtain the following equations,

\[
X_{t,T} = \sum_{i=1}^{p} \left( \alpha^{(i)}_{0,0} \phi \left( \frac{t}{T} \right) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \epsilon^{(i)}_{j,k} \psi^{(i)}_{j,k} \left( \frac{t}{T} \right) \right) X_{t-i,T} + \varepsilon^x_t + s^x_t,
\]

\[
Y_{t,T} = \sum_{i=1}^{p} \left( \delta^{(i)}_{0,0} \phi \left( \frac{t}{T} \right) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d^{(i)}_{j,k} \psi^{(i)}_{j,k} \left( \frac{t}{T} \right) \right) Y_{t-i,T} + \varepsilon^y_t + s^y_t
\]

(3.3)

where \( \{ \alpha^{(i)}_{0,0}, \epsilon^{(i)}_{j,k}, \delta^{(i)}_{0,0}, d^{(i)}_{j,k} \} \) \( 0 \leq j \leq J, 0 \leq k \leq 2^j - 1 \), are the coefficients for the corresponding functions \( a_i(u) \) and \( b_i(u) \), \( i = 1, \ldots, p \); \( \varepsilon^x_t, \varepsilon^y_t \) as before and \( s^x_t, s^y_t \) are errors due to the functional approximation (truncated expansions). However, these errors decay rapidly to zero as the time series length increases.

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

\[
X = \Psi_x c + \varepsilon_x + s_x,
\]

(3.4)
where

\[
\begin{align*}
X &= (X_{p+1,T}, X_{p+2,T}, \ldots, X_{T-1,T}, X_{T,T})', \\
c &= (c_0, c_0, \ldots, c_0, c_{J,2}^{(1)}, c_{J,2}^{(2)}, \ldots, c_{J,2}^{(p)}, c_{J,2}^{(p)}, \ldots, c_{J,2}^{(p)})', \\
\varepsilon_x &= (\varepsilon_x^{p+1}, \varepsilon_x^{p+2}, \ldots, \varepsilon_x^T)', \\
S_x &= (s_x^{p+1}, s_x^{p+2}, \ldots, s_x^T)',
\end{align*}
\]

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

\[
\Psi_x^{(i)} = \begin{bmatrix}
\phi_x \left( \frac{p+1}{T} \right) X_{p+1-i,T} & \psi_x^{(i)} \left( \frac{p+1}{T} \right) X_{p+1-intercept, \psi_x^{(i)} \left( \frac{p+1}{T} \right) X_{p+1-i,T} \\
\phi_x \left( \frac{p+2}{T} \right) X_{p+2-i,T} & \psi_x^{(i)} \left( \frac{p+2}{T} \right) X_{p+2-i,T} \\
\vdots & \ddots \\
\phi_x \left( \frac{T-1}{T} \right) X_{T-1-i,T} & \psi_x^{(i)} \left( \frac{T-1}{T} \right) X_{T-1-i,T} \\
\phi_x \left( \frac{T}{T} \right) X_{T-i,T} & \psi_x^{(i)} \left( \frac{T}{T} \right) X_{T-i,T}
\end{bmatrix},
\]

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

\[
Y = \Psi_y d + \varepsilon_y + s_y, \quad (3.5)
\]

where the quantities \( Y, \Psi_y, d, \varepsilon_y \) and \( s_y \) are similarly defined. Furthermore,

\[
\begin{align*}
E(\varepsilon_x) &= E(\varepsilon_y) = 0, \\
E(\varepsilon_x, \varepsilon_y') &= \text{Diag} \begin{bmatrix}
\sigma_x^2 \left( \frac{p+1}{T} \right), \sigma_x^2 \left( \frac{p+2}{T} \right), \ldots, \sigma_x^2 \left( \frac{T}{T} \right)
\end{bmatrix} = \Sigma_{xx}, \\
E(\varepsilon_y, \varepsilon_y') &= \text{Diag} \begin{bmatrix}
\sigma_y^2 \left( \frac{p+1}{T} \right), \sigma_y^2 \left( \frac{p+2}{T} \right), \ldots, \sigma_y^2 \left( \frac{T}{T} \right)
\end{bmatrix} = \Sigma_{yy}.
\end{align*}
\]

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

\[
E(\varepsilon_x, \varepsilon_y') = \text{Diag} \begin{bmatrix}
\sigma_{xy} \left( \frac{p+1}{T} \right), \sigma_{xy} \left( \frac{p+2}{T} \right), \ldots, \sigma_{xy} \left( \frac{T}{T} \right)
\end{bmatrix} = \Sigma_{xy}.
\]

The dimensions of \( X, \varepsilon_x, s_x, Y, \varepsilon_y \) and \( s_y \) are \( (T-p) \times 1 \), of \( c \) and \( d \) are \( pL \times 1 \), of \( \Psi_x \) and \( \Psi_y \) are \( (T-p) \times pL \) and of each matrix \( \Sigma_i \) is \( (T-p) \times (T-p) \), where \( L = 2^{J+1} \) corresponds to the total number of wavelet coefficients used in each approximation.

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

\[
Z = \Psi \beta + \varepsilon + s,
\]

where

\[
Z = \begin{bmatrix} X \\ Y \end{bmatrix}, \quad \Psi = \begin{bmatrix} \Psi_x & 0 \\ 0 & \Psi_y \end{bmatrix}, \quad \beta = \begin{bmatrix} c \\ d \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} \quad \text{and} \quad s = \begin{bmatrix} s_x \\ s_y \end{bmatrix}.
\]
with \( E(\varepsilon) = 0 \) and
\[
E(\varepsilon' \varepsilon) = \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.
\]

Thus, the generalized least squares estimator of \( \beta \) is given by
\[
\hat{\beta} = \left( \Psi' \Sigma^{-1} \Psi \right)^{-1} \Psi' \Sigma^{-1} Z.
\]

The dimensions of \( Z, s \) and \( \varepsilon \) are \( 2(T - p) \times 1 \), of \( \beta \) is \( 2pL \times 1 \) and of \( \Psi \) is \( 2(T - p) \times 2pL \).

Assuming that the process \( \varepsilon \) follows a multivariate Normal distribution with mean \( 0 \) and matrix of variances and covariances \( \Sigma \), it can be shown similarly to Ref. 5 and Ref. 9 that asymptotically
\[
\sqrt{T}(\hat{\beta} - \beta) \sim N(0, \Sigma^*).
\]

Hence, the hypotheses given in (3.2) are equivalent to
\[
H_0 : c = d \\
H_1 : c \neq d
\]
or equivalently to
\[
H_0 : C\beta = 0 \\
H_1 : C\beta \neq 0
\]
with \( C = [I_{pL} - I_{pL}] \). It follows that \( \sqrt{T}(C\hat{\beta} - C\beta) \) is asymptotically \( N_{pL}(0, C\Sigma^* C') \).

Defining now \( v \) by
\[
v = [C(\Psi' \Sigma^{-1} \Psi)^{-1} C']^{-1} \hat{v} (C\hat{\beta} - C\beta),
\]

it can be shown that under \( H_0 \), \( v \) has asymptotically a \( N(0, I_{pL}) \). Consequently under \( H_0 \), the statistic
\[
W = v' v = (C\hat{\beta})' [C(\Psi' \Sigma^{-1} \Psi)^{-1} C']^{-1} (C\hat{\beta})
\]
has asymptotically a \( \chi^2_{pL} \) distribution. Hence, for a given significance level \( \alpha \), 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 \( \varepsilon_x \) and \( \varepsilon_y \) are known. In practice, these structures have also to be estimated.

Considering the residuals \( r_x(t) \) and \( r_y(t) \) obtained in the least square estimation procedure we have,
\[
E[r_x^2(\frac{t}{T})] = \sigma_x^2(\frac{t}{T}), \quad (3.6)
\]
\[
E[r_y^2(\frac{t}{T})] = \sigma_y^2(\frac{t}{T}), \quad (3.7)
\]
\[
E[r_x(\frac{t}{T})r_y(\frac{t}{T})] = \sigma_{xy}(\frac{t}{T}), \quad (3.8)
\]
Comparing time-varying autoregressive structures

hence, applying the wavelets expansion, we obtain

\[ \sigma^2_x(tT) = \alpha_{x,0} \phi(tT) + \sum_{j=0}^{2^J-1} \sum_{k=0}^{2^J-1} c_{j,k}^x \psi_{j,k}(tT), \]

\[ \sigma^2_y(tT) = \alpha_{y,0} \phi(tT) + \sum_{j=0}^{2^J-1} \sum_{k=0}^{2^J-1} c_{j,k}^y \psi_{j,k}(tT), \]

\[ \sigma_{xy}(tT) = \alpha_{xy,0} \phi(tT) + \sum_{j=0}^{2^J-1} \sum_{k=0}^{2^J-1} c_{j,k}^{xy} \psi_{j,k}(tT). \]

In practice, truncated expansions are used. The criterion for choosing the maximum scale \( J \) is the same described previously. Finally, considering the ordinary least square smoothing to the series \( r^2_x(tT) \) and \( r^2_y(tT) \), it is easy to obtain consistent estimates of \( \sigma^2_x(tT) \), \( \sigma^2_y(tT) \) and \( \sigma_{xy}(tT) \).

We propose a generalization of the Cochrane and Orcutt procedure given by the following iterative estimation algorithm:

1) Apply the generalized least square estimation considering \( \Sigma \) equal to the identity matrix;
2) obtain residuals;
3) smooth the squared and cross residuals considering the ordinary least square, obtaining an estimate of \( \Sigma \);
4) apply the generalized least square estimation considering the estimated matrix \( \Sigma \);
5) return to 2 until numerical convergence of parameters or a maximum number of iterations.

4. Simulations

In this section, some simulation results are presented in order to evaluate the estimation procedure considering B-splines and wavelets, the asymptotic null distribution and the performance of the proposed test. All the 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

\[ b_1(tT) = a_1(tT) = \sin \left( \frac{2\pi tT}{T} \right) / 3. \]  

(4.1)

Hence, the processes generated in the simulation are given by

\[ X_{t,T} = a_1(tT) X_{t-1,T} + \varepsilon^x_t, \]

\[ Y_{t,T} = b_1(tT) Y_{t-1,T} + \varepsilon^y_t. \]
where $\varepsilon^x_t$ and $\varepsilon^y_t$ are Gaussian white noises, and

$$\sigma^2_x \left( \frac{t}{T} \right) = \left( 1 + \frac{\cos \left( \frac{2\pi t}{4} \right)}{4} \right)^2,$$

$$\sigma^2_y \left( \frac{t}{T} \right) = 0.25 \left( 1 + \frac{\cos \left( \frac{2\pi t}{4} \right)}{4} \right)^2 + 0.25 \left( 1 + \frac{\cos \left( \frac{2\pi t}{4} \right)}{4} \right)^6,$$

$$\sigma_{xy} \left( \frac{t}{T} \right) = 0.5 \left( 1 + \frac{\cos \left( \frac{2\pi t}{4} \right)}{4} \right)^3.$$

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. Notice that 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 estimates using B-splines show a variability higher than the others at the boundaries.

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.

In order to evaluate the asymptotic null distribution of the proposed statistic, one thousand of time-varying autoregressive series of order one were simulated,
Comparing time-varying autoregressive structures

considering coefficients $b_1(t) = a_1(t)$, where $a_1(t)$, $\epsilon_x^t$ and $\epsilon_y^t$ were the same as in the previous simulations. The proposed Wald test was applied for each generated series ($L = 4$, $T = 128$, and the $D_8$ wavelet) and the histogram, kernel density estimates and theoretical distribution are presented in Fig. 5.

Focusing the evaluation of suggested Wald test power, we generate time-varying autoregressive series of order one, with coefficients $b_1(t) = (1 + \lambda)a_1(t)$, where $\lambda = (0, 0.2, 0.4, ..., 1.8)$, and $a_1(t)$, $\epsilon_x^t$ and $\epsilon_y^t$ as before. Hence, the difference between the two autoregressive structures increases as the value of $\lambda$ 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 $T = 64, 128, 256, L = 4, 8, 16$ 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 statistics 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 in 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 sampling rate. However, the test based on wavelet bases seems to have a better performance for small samples.

5. Application to Real Data

In order to evaluate the performance of the proposed test in real applications, we consider two financial time series from Germany and USA. The volatility structure of financial assets is very useful to quantify risks. Furthermore, their predictions are frequently used to determine the value at risk of portfolios. We define the log-volatility time series as

$$z_t = \log(r_t^2 + \gamma),$$  

(5.1)

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

<table>
<thead>
<tr>
<th>T</th>
<th>L</th>
<th>Wavelet</th>
<th>0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td>Haar</td>
<td>0.049</td>
<td>0.053</td>
<td>0.081</td>
<td>0.129</td>
<td>0.227</td>
<td>0.371</td>
<td>0.486</td>
<td>0.674</td>
<td>0.792</td>
<td>0.889</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>0.057</td>
<td>0.064</td>
<td>0.090</td>
<td>0.140</td>
<td>0.233</td>
<td>0.368</td>
<td>0.529</td>
<td>0.680</td>
<td>0.787</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8</td>
<td>0.047</td>
<td>0.073</td>
<td>0.092</td>
<td>0.148</td>
<td>0.260</td>
<td>0.429</td>
<td>0.604</td>
<td>0.739</td>
<td>0.858</td>
<td>0.925</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BS</td>
<td>0.070</td>
<td>0.066</td>
<td>0.107</td>
<td>0.191</td>
<td>0.278</td>
<td>0.415</td>
<td>0.522</td>
<td>0.690</td>
<td>0.819</td>
<td>0.870</td>
</tr>
<tr>
<td>64</td>
<td></td>
<td>Haar</td>
<td>0.051</td>
<td>0.072</td>
<td>0.087</td>
<td>0.116</td>
<td>0.189</td>
<td>0.255</td>
<td>0.385</td>
<td>0.509</td>
<td>0.646</td>
<td>0.793</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>0.105</td>
<td>0.110</td>
<td>0.132</td>
<td>0.158</td>
<td>0.232</td>
<td>0.316</td>
<td>0.404</td>
<td>0.510</td>
<td>0.678</td>
<td>0.733</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8</td>
<td>0.110</td>
<td>0.106</td>
<td>0.128</td>
<td>0.172</td>
<td>0.221</td>
<td>0.308</td>
<td>0.409</td>
<td>0.506</td>
<td>0.677</td>
<td>0.722</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BS</td>
<td>0.125</td>
<td>0.132</td>
<td>0.160</td>
<td>0.195</td>
<td>0.249</td>
<td>0.322</td>
<td>0.430</td>
<td>0.523</td>
<td>0.657</td>
<td>0.764</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>Haar</td>
<td>0.152</td>
<td>0.194</td>
<td>0.182</td>
<td>0.196</td>
<td>0.249</td>
<td>0.298</td>
<td>0.362</td>
<td>0.437</td>
<td>0.529</td>
<td>0.621</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>0.166</td>
<td>0.142</td>
<td>0.185</td>
<td>0.186</td>
<td>0.215</td>
<td>0.252</td>
<td>0.299</td>
<td>0.398</td>
<td>0.459</td>
<td>0.531</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8</td>
<td>0.181</td>
<td>0.194</td>
<td>0.206</td>
<td>0.206</td>
<td>0.256</td>
<td>0.273</td>
<td>0.323</td>
<td>0.391</td>
<td>0.472</td>
<td>0.546</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BS</td>
<td>0.203</td>
<td>0.213</td>
<td>0.201</td>
<td>0.238</td>
<td>0.277</td>
<td>0.308</td>
<td>0.335</td>
<td>0.440</td>
<td>0.434</td>
<td>0.595</td>
</tr>
<tr>
<td>128</td>
<td></td>
<td>Haar</td>
<td>0.038</td>
<td>0.059</td>
<td>0.093</td>
<td>0.172</td>
<td>0.338</td>
<td>0.549</td>
<td>0.725</td>
<td>0.896</td>
<td>0.973</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>0.063</td>
<td>0.056</td>
<td>0.118</td>
<td>0.190</td>
<td>0.352</td>
<td>0.520</td>
<td>0.744</td>
<td>0.891</td>
<td>0.976</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8</td>
<td>0.049</td>
<td>0.049</td>
<td>0.117</td>
<td>0.184</td>
<td>0.327</td>
<td>0.538</td>
<td>0.773</td>
<td>0.896</td>
<td>0.953</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BS</td>
<td>0.083</td>
<td>0.093</td>
<td>0.139</td>
<td>0.226</td>
<td>0.391</td>
<td>0.593</td>
<td>0.756</td>
<td>0.887</td>
<td>0.949</td>
<td>0.976</td>
</tr>
<tr>
<td>256</td>
<td></td>
<td>Haar</td>
<td>0.069</td>
<td>0.056</td>
<td>0.082</td>
<td>0.175</td>
<td>0.245</td>
<td>0.377</td>
<td>0.540</td>
<td>0.750</td>
<td>0.852</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>0.145</td>
<td>0.138</td>
<td>0.182</td>
<td>0.208</td>
<td>0.286</td>
<td>0.406</td>
<td>0.583</td>
<td>0.709</td>
<td>0.825</td>
<td>0.908</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8</td>
<td>0.138</td>
<td>0.130</td>
<td>0.187</td>
<td>0.239</td>
<td>0.286</td>
<td>0.426</td>
<td>0.560</td>
<td>0.753</td>
<td>0.816</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BS</td>
<td>0.147</td>
<td>0.162</td>
<td>0.201</td>
<td>0.238</td>
<td>0.323</td>
<td>0.459</td>
<td>0.588</td>
<td>0.713</td>
<td>0.856</td>
<td>0.912</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>Haar</td>
<td>0.042</td>
<td>0.052</td>
<td>0.198</td>
<td>0.190</td>
<td>0.490</td>
<td>0.801</td>
<td>0.946</td>
<td>0.990</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>0.039</td>
<td>0.055</td>
<td>0.265</td>
<td>0.531</td>
<td>0.851</td>
<td>0.958</td>
<td>0.997</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8</td>
<td>0.039</td>
<td>0.072</td>
<td>0.251</td>
<td>0.591</td>
<td>0.869</td>
<td>0.982</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BS</td>
<td>0.043</td>
<td>0.086</td>
<td>0.218</td>
<td>0.534</td>
<td>0.822</td>
<td>0.977</td>
<td>0.995</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Haar</td>
<td>0.028</td>
<td>0.031</td>
<td>0.136</td>
<td>0.371</td>
<td>0.699</td>
<td>0.904</td>
<td>0.987</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>0.023</td>
<td>0.053</td>
<td>0.163</td>
<td>0.394</td>
<td>0.675</td>
<td>0.912</td>
<td>0.984</td>
<td>0.997</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8</td>
<td>0.013</td>
<td>0.032</td>
<td>0.139</td>
<td>0.377</td>
<td>0.688</td>
<td>0.928</td>
<td>0.982</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BS</td>
<td>0.091</td>
<td>0.178</td>
<td>0.407</td>
<td>0.738</td>
<td>0.912</td>
<td>0.983</td>
<td>0.996</td>
<td>0.994</td>
<td>0.997</td>
<td>0.997</td>
</tr>
</tbody>
</table>

This illustrative application is based on the log-volatilities of the national stock market indexes of Germany and USA. We considered $\gamma = 0.01$, and daily log-returns from September first 1999 to August twelve 2003, resulting on a time series of length 1024. The log-volatilities were standardized to zero mean and unit variance, and they are presented in Fig. 9.

The time-varying autoregressive estimates are shown in Fig. 10. The autoregressive coefficients seem to be higher in the first half of the time series, and the values are reduced in the second half. We considered the wavelets Haar, D2, D8 and...
the B-splines basis, with $L = 8$ functions. The p-values of the test to compare if the two log-volatilities have the same time-varying autoregressive structure are $0.080$, $0.296$, $0.077$ and $0.256$, respectively. Hence, considering a test size of $\alpha = 0.05$, we do not reject the hypothesis that the two time series have the same autoregressive structure. As a possible interpretation of this result, we can say that both stock markets have a large number of common risk factors. However, some additional macroeconomic analysis should be performed.

6. Conclusions

In the context of non-stationary time series there exist series which are not easily transformable to stationary series, since their parameters vary on time. Assuming that these processes follow a time-varying autoregressive model, we proposed a statistical test for comparing their function 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.
Fig. 7. Number of basis functions ($L$) effect on the Wald test power, using $D_8$ and $T = 128$.

References

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

Fig. 9. Log-volatility time series of the German and North American Stocks.

Fig. 10. Estimated time-varying autoregressive structures to the stock market log-volatilities.