Biometrika (2013), **xx**, x, *pp*. 1–17

Printed in Great Britain

Entropy testing for nonlinearity in time series.

BY SIMONE GIANNERINI
Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, 40126, Bologna, Italy.
simone.giannerini@unibo.it

ESFANDIAR MAASOUMI
Department of Economics, Emory University, 1602 Fishburne Dr, Economics, Atlanta, Georgia, USA.
esfandiar.maasoumi@emory.edu

AND ESTELA BEE DAGUM
Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, 40126, Bologna, Italy.
estellebee@dagum.us

SUMMARY

We propose a test for identification of nonlinear serial dependence in time series against the general "null" of linearity, in contrast to the more widely examined null of "independence". The approach is based on a combination of an entropy dependence metric, possessing many desirable properties and used as a test statistic, together with $i)$ a suitable extension of surrogate data methods, a class of Monte Carlo distribution-free tests for nonlinearity; $ii)$ the use of a smoothed sieve bootstrap scheme. We show how the tests can be employed to detect the lags at which a significant nonlinear relationship is expected in the same fashion as the autocorrelation function is used for linear models. We prove the asymptotic validity of the procedures proposed and of the corresponding inferences. The small sample size performance of the tests is assessed through a simulation study. Applications to real data sets of different kinds are also presented.

Some key words: Nonlinearity; Time series; Test; Entropy; Bootstrap; Surrogate data

1. INTRODUCTION

The literature on tests for nonlinearity, or nonlinear serial dependence in time series is rather extensive. Providing a unified mathematical framework that encompasses all aspects of nonlinearity has proven elusive. While the class of linear process is well defined from a mathematical point of view, the same does not hold for processes that do not belong to that class. Thus, even though departures from linearity can occur in many directions, testing for nonlinearity is often a test for specific nonlinear feature or form, making it difficult to compare existing proposals. Nonlinear features have been found in different disciplines; for instance, some of the particular notions of nonlinear dynamics and chaos theory such as initial value sensitivity, fractal dimensions, and non uniform noise amplification, motivated the introduction of new tools and tests. In other situations, nonlinearity is inferred given an inability of a linear model. Thus, the problem reduces to either a diagnostic test (usually performed on the residuals of a linear model) or a specification test between models. For a recent review on the topic see Giannerini (2012) and references therein.

Two common traits in almost all the tests for nonlinearity are, $i)$ they are based on specific moments or features of the distribution of the process, and $ii)$ they focus on the null of "no nonlinearity", or "no dependence". The latter is a rather strong "strawman" unless the process has been pre-filtered. Furthermore, many of these tests are designed to work with a restricted class of models. This fact is crucial: since in practice the true model is never known, the reported performance of such tests based on simulation studies does not reflect the real performance which can depend on the degree of misspecification.

In this paper we propose a flexible test of nonlinearity based on the whole (pairwise) distribution of the process. While our test is strictly "diagnostic", it demonstrates an impressive ability to identify areas of nonlinearity, and is able to handle null hypotheses such as general linear time series processes. The test is nonparametric in that it does not require the specification of a model; furthermore, it has power against a wide range of processes and can be used to identify the lags at which a nonlinear relationship is expected in the same fashion as the autocorrelation function is used for the identification of linear models.

Our null hypotheses are connected with the formal definition of linear processes. In particular H_0 assumes that the data generating process $\{X_t\}$ is a zero mean linear Gaussian stationary process as follows:

$$H_0 : X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t \;\text{ with } \{\varepsilon_t\} \sim \text{independent and identically distributed } N(0, \sigma_\varepsilon^2), \tag{1}$$

where $\sum_{j=1}^{\infty} \phi_j^2 < \infty$ and $E[X_t]^4 < \infty$. A related, more general, hypothesis is that $\{X_t\}$ is a zero mean linear stationary process (not necessarily Gaussian) that admits an infinite autoregressive representation: kdens

$$H_0' : X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t \;\text{ with } \{\varepsilon_t\} \sim \text{independent and identically distributed } f(0, \sigma_\varepsilon^2), \tag{2}$$

where f is the probability density function of the error process $\{\varepsilon_t\}$. The alternative hypothesis H_1 states that $\{X_t\}$ does not admit a representation as in (1) or (2). Now, we assume we are given a time series $\mathbf{x} = (x_1, \ldots, x_n)$ and we would like to test whether $\mathbf{x}$ might be operationally considered as a realization of the process of Eq. (1) or Eq. (2).

The test statistics we propose are functionals of an entropy-metric measure of dependence for time series. Such entropy statistic possesses many desirable properties and has been shown to be powerful in other settings, for example in Granger et al. (2004) and Maasoumi & Racine (2009). We will show that under the null hypothesis of a linear Gaussian process the entropy measure reduces to a nonlinear function of the linear correlation coefficient. Hence, we are able to build a test statistic from the quadratic divergence between the (parametric) estimator of the entropy measure under H_0 and the corresponding unrestricted (nonparametric) estimator. Also, the statistic based on the nonparametric estimator can be utilized for testing H_0'. Then, we derive the asymptotic distribution of the test statistics under both H_0 and H_0'. Typically, the asymptotic approximations depend upon quantities that are unknown and require large samples to be valid; hence, we propose and examine two resampling schemes. The first one is based on the method of surrogate data while the second makes use of the smoothed sieve bootstrap. We prove the asymptotic validity of the procedures proposed and of the corresponding inferences.

The finite sample performance of the tests are assessed by means of a simulation study where we show how the combination of the entropy based measure together with resampling techniques can form a powerful and flexible test for nonlinearity that can be used in many applied fields. Finally, we discuss in some detail two applications to real data sets.

The article is structured as follows. In Section 2 we define the entropy based measure and derive the properties of its estimators both under the null and in the unrestricted case. In section 3 we propose the test statistics and show the asymptotic derivations for its estimator under the two null hypotheses. In Section 4 we describe how the method of surrogate data and bootstrap techniques can be adapted in our context in order to build tests with good finite sample properties. Also, we prove the theoretical validity of the corresponding inferences. Section 5 is devoted to a simulation study in which we investigate power and size of the tests, while in section 6 we present two different applications of the tests to real time series.

2. TOWARDS A NONLINEAR AUTOCORRELATION FUNCTION

2·1. *Introduction and definition*

In literature, there exist many proposals of measures of dependence, each of them motivated by different needs and built to characterize specific aspects of the process under study. An important class of such measures is based on entropy functionals developed within information theory (see e.g. Maasoumi, 1993; Joe, 1989). For instance, Shannon mutual information and the Kullback-Leibler divergence have spread widely in the context of nonlinear dynamics. Such entropy measures have been used also in the context of time series analysis (Robinson, 1991; Granger & Lin, 1994; Tjøstheim, 1996; Hong & White, 2005). However, most of these entropies are not "metric" since either they do not obey the triangular inequality or they are not commutative operators. These shortcomings are not consequential for most tests, but impinge on our ability to assess and quantify degrees of dependence or departures from points of interest, or to search for minimum distanced/optimal solutions or models. The measure we discuss here is the metric entropy measure S_ρ, a normalized version of the Bhattacharya-Hellinger-Matusita distance, defined as follows:

$$S_\rho(k) = \frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left((f_{(X_t,X_{t+k})}(x_1,x_2))^{1/2} - (f_{X_t}(x_1)f_{X_{t+k}}(x_2))^{1/2}\right)^2 dx_1dx_2 \qquad (3)$$

where $f_{X_t}(\cdot)$ and $f_{(X_t,X_{t+k})}(\cdot,\cdot)$ denote the probability density function of X_t and of the vector (X_t, X_{t+k}) respectively. The measure is a particular member of the family of symmetrized general "relative" entropies, which includes as a special case non metric relative entropies often referred to as Shannon or Kullback-Leibler divergence. In the case of serial dependence in time it can be interpreted as a (nonlinear) autocorrelation function. $S_\rho(k)$ satisfies many desirable properties. In particular, $i)$ it is a metric and is defined for both continuous and discrete variables, $ii)$ it is normalized and takes the value 0 if X_t and X_{t+k} are independent and 1 if there is a measurable exact relationship between continuous variables, $iii)$ it reduces to a function of the linear correlation coefficient in the case of jointly Gaussian variables, and $iv)$ it is invariant with respect to continuous, strictly increasing transformations. The above mentioned properties of the metric entropy can be seen as part of a general discussion on measures of dependence as put forward in Rényi (1959) and further studied in Maasoumi (1993) and Granger et al. (2004). Also, Micheas & Zografos (2006) study the generalization to the multivariate case of a class of measures that include S_ρ as a special case; see the supplementary material for further details.

A key result under the perspective of testing for nonlinearity concerns the relationship with the correlation coefficient in the Gaussian case (point iii)); a correction to Granger et al. (2004) is in order. We find the following:

PROPOSITION 1. *Let $(X,Y) \sim N(\mathbf{0},\mathbf{1},\rho)$ be a standard Normal random vector with joint probability density function given by $f_{X,Y}(\cdot,\cdot,\rho)$, where ρ is the correlation coefficient. Then, the following relation holds:*

$$S_\rho = 1 - \frac{2\left(1-\rho^2\right)^{1/4}}{(4-\rho^2)^{1/2}} \tag{4}$$

Proof. See Appendix. □

For the sake of brevity, in the following, instead of $S_\rho(k)$ we use S_k.

2·2. *The parametric estimator under H_0*

Eq. 4 allows us to obtain an estimator for S_k under the null of a linear Gaussian process based on the sample correlation coefficient ρ_k. We denote such a parametric estimator with $\hat{S}_k^p$ where the superscript p stands for *parametric*. In the following two results we derive the asymptotic distribution of $\hat{S}_k^p$ and prove its consistency.

PROPOSITION 2. *Let $\{X_t\}$ be the zero mean stationary process under H_0 as in Eq. (1). Also, let $\hat{\rho}_k$ be the sample autocorrelation function of $\{X_t\}$ at lag k and let $\hat{S}_k^p = 1 - \frac{2\left(1-\hat{\rho}_k^2\right)^{1/4}}{(4-\hat{\rho}_k^2)^{1/2}}$ be the corresponding sample estimator of S_ρ at lag k based on Eq. (4). Lastly, define the following function $g:[-1,1]\to[0,1]$, $g(x) = 1 - \frac{2\left(1-x^2\right)^{1/4}}{(4-x^2)}$. The function g is differentiable on $(0,1)$ and the ith derivative $g^{(i)}(x) \neq 0$ for $x \neq 0$. Then, for every $k = 0,1,\ldots$ we have:*

$$n^{1/2}(\hat{S}_k^p - S_k) \to N(0,\sigma_p^2) \quad \textit{in distribution}$$

where $\sigma_p^2 = [g'(\zeta)]^2$ and $\zeta = \sum_{i=1}^{\infty}\left\{\rho_{(i+k)}\rho_{(i-k)} - 2\rho_i\rho_k\right\}^2$ is the asymptotic variance of $\hat{\rho}_k$ (see Brockwell & Davis, 1991, pp. 221-222).

Proof. See Appendix. □

In the case of no correlation ($\rho_k = 0$) we have $g'(\rho_k) = 0$. In such instance the approximation is driven by higher order derivatives, in particular the even-order ones. Now we show that $\hat{S}_k^p$ is a mean-square consistent estimator for S_k.

PROPOSITION 3. *Under the hypotheses of Proposition 2 we have:*

$$\hat{S}_k^p \to S_k$$

in L^2.

Proof. See Appendix. □

2·3. *The unrestricted nonparametric estimator*

The issue of estimation of S_k and related entropy measures under conditions that allow to build tests for serial dependence has been studied in Robinson (1991); Tjøstheim (1996); Skaug & Tjøstheim (1996); Granger et al. (2004); Hong & White (2005); Fernandes & Néri (2010). In this section we adapt the relevant theory to our case and derive the asymptotic distribution of the nonparametric estimator for S_k.

The nonparametric estimator of S_k, which we denote by $\hat{S}_k^u$, has the form

$$\hat{S}_k^u = \frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(\left\{\hat{f}_{(X_t,X_{t+k})}(x_1,x_2)\right\}^{1/2} - \left\{\hat{f}_{X_t}(x_1)\hat{f}_{X_{t+k}}(x_2)\right\}^{1/2}\right)^2 w(x_1,x_2)dx_1dx_2$$

where we use kernel density estimators for f_{X_t}, $f_{X_{t+k}}$ and $f_{(X_t,X_{t+k})}$:

$$\begin{aligned} \hat{f}_{X_t}(x) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1}K\left(\frac{x-X_i}{h_1}\right) \\ \hat{f}_{(X_t,X_{t+k})}(x_1,x_2) &= \frac{1}{n-k}\sum_{i=1}^{n-k}\frac{1}{h_1h_2}K\left(\frac{x_1-X_i}{h_1},\frac{x_2-X_{i+k}}{h_2}\right)\ . \end{aligned} \tag{5}$$

Here, K is a kernel function, h_1, h_2 are the bandwidths and $w(x_1,x_2)$ is a continuous weight function that is needed both for excluding outlying observations and for the asymptotic analysis. We assume the following regularity conditions:

(A1) $\{X_t, t\in\mathbb{N}\}$ is strictly stationary and β-mixing process with exponentially decaying coefficient.

(A2) f_{X_t},$f_{X_{t+k}}$,$f_{(X_t,X_{t+k})}$ are continuously differentiable up to order s and their derivatives are bounded and square integrable. Also, the joint density function of $(Z_{k_1},\dots,Z_{k_\varsigma})$ where $Z_t=(X_t,X_{t+k})$ and with $1\le\varsigma\le4$ is Lipschitz-continuous, i.e. $|f(Z_{k_1}+\delta,\dots,Z_{k_\varsigma}+\delta)-f(Z_{k_1},\dots,Z_{k_\varsigma})|\le\mathcal{D}\times||\delta||$, where $\mathcal{D}$ is an integrable function and $1\le\varsigma\le4$.

(A3) $K(u)$ is continuous kernel function, differentiable up to order s, that satisfies: $\int uK(u)du=0$; $\int u^2K(u)du<\infty$; $K(x)=\int\bar{K}(u)\exp\{i\eta x\}d\eta$; $e_K=\int K^2(u)du$; $v_K=\int\left(\int K(u)K(u+v)du\right)^2dv$, where $i=\sqrt{-1}$ and $\bar{K}(u)$ is a real function such that $\int|\bar{K}(u)|d\eta<\infty$

(A4) Define the bandwidths $h_1=h_1(n,X_t)$, $h_2=h-2(n,X_{t+k})$ such that $h_i\to0$ and $nh_i\to\infty$ as $n\to\infty$. Also, $h_i=o\left(n^{-1/(2s+1)}\right)$ with $i=1,2$.

(A5) $w(x_1,x_2)=\mathbb{I}\{(x_1,x_2)\in D\}$ is a non-negative, separable weigthing function (i.e. $w(x_1,x_2)=w(x_1)w(x_2)$) with $D=D_1\times D_1$ and D_1 is a closed real interval.

These assumptions lead us to the following result:

PROPOSITION 4. *Under assumptions (A1)–(A5):*

1. $\hat{S}_k^u\to S_k$, *in* L^2.
2. $n^{1/2}(\hat{S}_k^u-S_k)\to N(0,\sigma_u^2)$ *in distribution,*
   where σ_u^2 *is the asymptotic variance that depends on the weight function* $w(x_1,x_2)$.

Proof. See Appendix. □

The above assumptions can be relaxed to some extent without affecting the results. For instance, one could assume also α-mixing processes and less restrictive assumptions on the kernels. While the choice of kernel function has a limited impact on the performance of the test presented in the next section, the choice of the bandwidth plays a crucial role. In this paper, we have investigated two methods for selecting the bandwidth. The first method is the maximum likelihood cross validation (MLCV): we choose the bandwidth h that maximizes the score function $\mathrm{CV}(h)=n^{-1}\sum_{i=1}^{n}\log\hat{f}_{-i}(X_i)$, where $\hat{f}_{-i}(X_i)=(n-1)^{-1}h^{-1}\sum_{j\neq i}K(h^{-1}(X_i-X_j))$ is the leave-one-out kernel density estimate of X_i. The second method is the normal reference method for which we have either $h = 1.06\,\hat{\sigma}\,n^{-1/5}$ (univariate case) or $h_i = 1.06\,\hat{\sigma_i}\,n^{-1/6}$ with $i = 1, 2$ (bivariate case). For further details on both methods see Silverman (1986).

The practical implementation of $\hat{S}_k^u$ requires also the computation of a double integral for which adaptive quadrature methods have been employed. Details of the implementation of the software are given in the supplementary material. An alternative estimator of the measure that uses summation instead of integration can be used; however, as remarked in Granger et al. (2004), such solution is not advisable as it can cause a degradation in the performance of the tests.

3. THE TEST STATISTICS

In order to test the null hypotheses of linearity H_0 (1) and H_0' (2) we propose the following test statistics:

$$\hat{T}_k = \left[\hat{S}_k^u - \hat{S}_k^p\right]^2 \qquad \text{for } H_0 \tag{6}$$

$$\hat{S}_k^u \qquad \text{for } H_0' \tag{7}$$

$\hat{T}_k$ is the squared divergence between the unrestricted nonparametric estimator and the parametric estimator of S_k under H_0. The following theorem establishes strong convergence and the asymptotic distribution of $\hat{T}_k$ under the null.

THEOREM 1. *Under the assumptions of Propositions* 2 *and* 4*:*

1. $\hat{T}_k \to 0$ *in* L^2.
2. $\dfrac{n\hat{T}_k}{\sigma_a^2} \to \chi_1^2$ *in distribution.*

where σ_a^2 *is the asymptotic variance of* $\hat{T}_k^{1/2}$.

Proof. See Appendix. □

Theorem 1 shows that the test statistic will converge to zero in L^2 if the process is linear and Gaussian. Hence, large values of $\hat{T}_k$ will indicate departure from the hypothesis of linearity. The derivation of the asymptotic approximation for the significance level and power of the test is bound to deriving an estimator for the asymptotic variance σ_a^2, which, in turn, depends upon σ_u^2 and σ_p^2. The task is feasible only for few specific cases and the exercise is of little practical relevance. Furthermore, preliminary investigations show that very large sample sizes are required to obtain meaningful results if the asymptotic approximations are used (see supplementary material). The same problems have been reported in literature (see e.g Hjellvik & Tjøstheim, 1995; Hjellvik et al., 1998; Tjøstheim, 1996; Hong & White, 2005).

As for the general null hypothesis of linearity H_0' (2) we propose using the nonparametric estimator $\hat{S}_k^u$. Even though, unlike the case of $\hat{T}_k$, it is not possible to derive an asymptotic test, this fact has little practical relevance. In fact, Proposition 4 assures that under mild conditions that include the class of linear processes defined by H_0', the statistic $\hat{S}_k^u$ is consistent and asymptotically Gaussian. Based on the above discussion we propose and study two resampling schemes that, together with our test statistics, lead to valid inferences and deliver a good performance in finite samples. The first scheme is based on surrogate data methods and is suited to testing H_0 while the second scheme relies on a smoothed version of the sieve bootstrap and is suitable also to testing H_0'.

4. THE SURROGATE DATA APPROACH

The method of surrogate data was introduced in the context of nonlinear time series analysis motivated by chaos theory. It can be regarded as a resampling approach for building tests for nonlinearity in absence of the distribution theory. Even though the use of tests based on simulations has been common practice in the Statistics community long before 1990, in the literature on nonlinear dynamics Theiler et al. (1992) is usually indicated as the seminal paper on the subject. The main idea at the basis of the method can be summarized as follows: $i)$ a null hypothesis regarding the process that has generated the observed series is formulated; for instance, H_0 : the generating process is linear and Gaussian, $ii)$ a set of B resampled series, called *surrogate series*, consistent with H_0, are obtained through Monte Carlo methods, $iii)$ a suitable test statistic known to have discriminatory power against H_0 is computed on the surrogates, obtaining the distribution of the test statistic under H_0. Thus, the significance level of the test and p-values may be derived from the resampling distribution. Clearly, the basic principle behind surrogate data tests is the same as the bootstrap principle but the performances can be different.

In Theiler et al. (1992) and Theiler & Prichard (1996) a null hypothesis of linearity is tested by generating surrogates having the same periodogram and the same marginal distribution as the original series. It is assumed that the generating process is a linear Gaussian process as in Eq. (1). Also, it is assumed that the process admits a spectral density function that forms a Fourier pair with the autocovariance function. Given an observed series $\mathbf{x} = (x_1, \ldots, x_n)^T$ we can define its discrete Fourier transform

$$\zeta_{\mathbf{x}}(\omega) = \frac{1}{(2\pi n)^{1/2}} \sum_{t=1}^{n} \mathbf{x} \exp(-i\omega t), \quad -\pi \le \omega \le \pi$$

and the sample periodogram $I(\mathbf{x}, \omega) = |\zeta_{\mathbf{x}}(\omega)|^2$. In general, it can be shown that $\zeta_{\mathbf{x}}(\omega) = \frac{1}{(2\pi)^{1/2}} P_n \mathbf{x}$ where P_n is an orthonormal matrix. Hence, assuming n as an odd number, the series $\mathbf{x}$ can be uniquely recovered from the sample mean, the periodogram values $I(\mathbf{x}, \omega_j)$ $j = 1, \ldots, \frac{n-1}{2}$ and the phases $\theta_1, \ldots, \theta_{\frac{n-1}{2}}$ through the following formula:

$$x_t = \bar{x} + \left(\frac{2\pi}{n}\right)^{1/2} \sum_{j=1}^{\frac{n-1}{2}} 2(I(\mathbf{x}, \omega_j))^{1/2} \cos(\omega_t j + \theta_j) \tag{8}$$

Such a relation allows one to obtain a surrogate series $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)^T$ by randomizing the phases of the above equation as follows:

$$x_t^* = \bar{x} + \left(\frac{2\pi}{n}\right)^{1/2} \sum_{j=1}^{m} 2(I(\mathbf{x}, \omega_j))^{1/2} \cos(\omega_t j + \theta_j) \tag{9}$$

where $\theta_1, \ldots, \theta_m$ are i.i.d. $U[0, 2\pi]$. The surrogate series will have the same sample mean and periodogram as the original series. In Chan (1997) it is proved that the phase randomization method described above is exactly valid under the null hypothesis that the generating process is a stationary Gaussian circular process. By valid, it is meant that tests based upon it are similar (they have a Neyman structure). Also, the author proves the asymptotic validity for the null hypothesis of a stationary Gaussian process with fast-decaying autocorrelations (see also Chan & Tong (2001), Chap. 4.4). Apart from the notable exception of Chan (1997) and despite the fair amount of literature on surrogate data methods, to our knowledge, comprehensive studies on the theoretical properties of such tests are lacking.

The approach we propose in this paper is an extension of the scheme that fits within the unified framework of constrained randomization (see Schreiber (1998); Schreiber & Schmitz (2000)). Basically, the generation of surrogate time series may be seen as an optimization problem that is solved by means of simulated annealing. Also, all the test statistics produced act as pivotal due to the constrained randomization (see Theiler & Prichard (1996)). The procedure can be summarized as follows: $i)$ Define one or more constraints in terms of a cost function C. This function reaches a global minimum when the constraints are fulfilled; $ii)$ Minimize the cost function C among all the possible permutations of the series through simulated annealing. In our case we generate surrogate series having the same autocorrelation function and the same sample mean of the original series. In the following proposition we show that under the null H_0 of a linear Gaussian process the surrogate approach combined with our test statistics leads to valid inferences.

PROPOSITION 5. *Under the null hypothesis H_0 that the data generating process is linear and Gaussian the constrained randomization approach, together with $\hat{T}_k$ (or $\hat{S}_k^u$), leads to asymptotically valid inferences in that the associated p-value follows a uniform distribution on $(0, 1)$*

Proof. See Appendix. □

In Schreiber & Schmitz (1996) it is shown that the constrained randomization approach leads to tests with better power and size as compared to those derived through the phase randomization. The procedure and the implementation are described in detail in the supplementary material.

5. THE BOOTSTRAP APPROACH

The second approach considered is a smoothed version of the sieve bootstrap. The main idea of the sieve bootstrap relies upon the Wold decomposition of a stationary process. In fact, under mild assumptions, a real-valued purely stochastic stationary process admits a one-sided infinite-order autoregressive representation. Hence, the sieve approximates a possibly infinite-dimensional model through a sequence of finite dimensional autoregressive models. The non smoothed version of this approach has been investigated in a number of studies (see e.g. Kreiss & Franke (1992); Bühlmann (1997, 2002)). In particular, in Bühlmann (1997) it is shown that the scheme leads to valid inferences for smooth functions of linear statistics. Since our test statistics have a component based on kernel density estimators we need to resort to the smoothed version of the sieve bootstrap proposed in Bickel & Bühlmann (1999). Such scheme is asymptotically valid for estimators that are (compactly) differentiable functionals of empirical measures. The idea of resampling from a smooth empirical distribution ensures that the bootstrap process inherits the mixing properties needed to prove asymptotic results.

In brief, the smoothed sieve scheme can be adapted to our case as follows: $i)$ fit an autoregressive model to the data; $ii)$ resample from the kernel density estimate of the residuals of the fit; $iii)$ generate a new series by driving the fitted model with the residuals obtained at step $ii)$. The full implementation of the scheme and further details are provided in the supplementary material.

In Bühlmann (1997) it is shown that if the AIC criterion for model selection is used then consistency is achieved for the arithmetic mean and a class of nonlinear statistics. Moreover, the method adapts automatically to the decay of the dependence structure of the process. Finally, the author remarks that the performance of the method is quite insensitive to choice of the criterion used for model selection as long as the order chosen is reasonable.

In the following proposition we prove the validity of the inference based on the combination of our test statistics and the smoothed sieve bootstrap scheme:

PROPOSITION 6. *given the assumptions of Theorem 4.1 of Bickel & Bühlmann (1999)*

1. *Under* $H_0 : \sup_x |pr^*[n^{1/2}(\hat{T}_k^* - T_k^*)) \leq x] - pr[n^{1/2}(\hat{T}_k - T_k \leq x]| = o_p(1) \quad n \to \infty$
2. *Under* $H_0' : \sup_x |pr^*[n^{1/2}(\hat{S}_k^{u*} - S_k^*)) \leq x] - pr[n^{1/2}(\hat{S}_k^u - S_k \leq x]| = o_p(1) \quad n \to \infty$

Proof. See Appendix. □

Interestingly, as discussed in Bickel & Bühlmann (1997) the closure of the class of linear processes that satisfy Wold's representation theorem is surprisingly broad and can include also non ergodic Poisson sum processes. Also, empirical evidence shows that the scheme is able to perform well also for processes that do not admit a linear representation such as threshold processes. From the point of view of testing for non linearity these facts pose several important questions. In general, we expect the sieve bootstrap test to be more conservative than the surrogate based one.

6. A SIMULATION STUDY

6·1. *Introduction*

In this section we assess the performance of the tests in finite samples by means of a simulation study. We focus on the following models:

Model 1: $x_t = 0.8\, x_{t-1} + \varepsilon_t$

Model 2: $x_t = 0.6\, x_{t-1} + 0.4\, \varepsilon_{t-1} + \varepsilon_t$

Model 3: $x_t = 0.8\, x_{t-1} + \zeta_t$

Model 4: $x_t = 0.6\, x_{t-1} + 0.4\, \zeta_{t-1} + \zeta_t$

Model 5: $x_t = 0.6\, \varepsilon_{t-1}\, x_{t-2} + \varepsilon_t$

Model 6: $x_t = 0.8\, \varepsilon_{t-2}^2 + \varepsilon_t$

Model 7: $x_t = \sigma_t \varepsilon_t$
$\sigma_t^2 = 0.1 + 0.6\, \sigma_{t-1}^2 + 0.3\, x_{t-1}^2$

Model 8: $x_t = \begin{cases} -0.8\, x_{t-1} + \varepsilon_t \text{ if } x_{t-1} \leq 0 \\ \;\;0.8\, x_{t-1} + \varepsilon_t \text{ if } x_{t-1} > 0 \end{cases}$

Model 9: $x_t = 4\, x_t\, (1 - x_t) + \sigma_t \eta_t$ with $(\eta_t + \frac{1}{2}) \sim \text{Beta}(10, 10)$

where the innovation processes are independent and identically distributed with $\varepsilon_t \sim N(0, 1)$ and $\zeta_t \sim$ Student's t with 3 degrees of freedom. The null hypothesis of linearity and Gaussianity of Eq. (1) is tested by means of $\hat{T}_k$ under both the surrogate and the bootstrap schemes. The general null of linearity of Eq. (2) is tested through $\hat{S}_k^u$ coupled with the bootstrap scheme.

Models 1 and 2 are linear Gaussian processes so that the rejection percentages give an indication of the size of the test. Models 3 and 4 are linear autoregressive moving average processes but with Student's t innovations so that the tests based on T should reject the null while those based on S should not. Models 5–9 are nonlinear processes that do not admit an infinite autoregressive representation. In particular, Model 5 is a bilinear process, Model 6 is a nonlinear moving average process, Model 7 is a generalized autoregressive conditional heteroscedastic process, Model 8 is a threshold autoregressive process and Model 9 is the logistic map at chaotic regime with additive Beta noise.

If the data generating process is nonlinear we expect that the values of the statistics obtained on the original series differ significantly from those obtained under the null. In all the experiments $\alpha = 0.05$ and the number of surrogates/bootstrap replicates is set to $B = 499$. All the results are given in terms of rejection percentages of the test over 500 Monte Carlo replications. As

Table 1. $\hat{T}_k$, *Surrogates,* MLCV *criterion: rejection percentages,* $k = 1, \ldots, 5$, *at* $\alpha = 5\%$.

	$n = 120$						$n = 240$					
	1	2	3	4	5	C	1	2	3	4	5	C
Model 1	5.0	4.0	7.5	6.0	4.5	22.5	5.5	5.0	3.5	4.5	7.5	12.5
Model 2	6.0	8.5	5.0	7.5	8.5	25.5	5.0	6.5	10.0	7.0	8.5	14.5
Model 3	5.0	7.0	21.5	15.5	16.0	37.0	34.0	60.0	78.0	64.0	36.5	80.5
Model 4	8.5	18.5	15.0	16.5	14.5	41.5	19.5	81.0	37.5	18.0	11.5	87.0
Model 5	30.5	67.5	10.5	23.0	3.5	78.5	40.0	91.5	13.5	48.5	5.0	99.5
Model 6	6.0	88.0	3.5	3.0	4.5	90.5	5.0	99.0	5.5	4.5	7.5	100.0
Model 7	37.0	29.0	17.0	18.5	10.5	65.5	69.5	56.0	45.0	33.0	31.0	94.5
Model 8	38.0	20.0	11.0	10.5	9.5	60.0	42.5	27.0	14.5	8.0	8.5	73.0
Model 9	100.0	100.0	100.0	96.5	19.5	100.0	100.0	100.0	100.0	98.5	21.5	100.0

for the sample sizes, we have chosen $n = 120$ and $n = 240$. The standard error of the Monte Carlo estimates is at most 2%. Note that, in analogy with tests based on autocorrelations, our procedures depend upon the choice of the lag k. In the simulations we have set the maximum lag k_{max} to 5. Hence, we show the rejection percentages for each k as to stress the ability of our statistics to highlight the lags at which a nonlinear dependence is expected. This means that the null depends on k so that we write H_0^k. Now, as is customary, if one is interested in assessing nonlinearity no matter the lag, then a null such as $H_0^{k_{max}} : \bigcap_{k=1}^{k_{max}} H_0^k$ can be adopted, see e.g. Fernandes & Néri (2010). In other words, the null of linearity is rejected if the test rejects for at least one lag. This combined result is presented in the last columns of the tables and is denoted with a "C".

6·2. *Surrogates*

In this section we report the results regarding the test implemented through the constrained randomization method presented in Section 4 which we denote by surrogates. In Table 1 we present the empirical rejection percentages for the test based on $\hat{T}_k$ where the MLCV criterion for the bandwidth is used. The rejection percentages clearly show high power in almost every situation and even for relatively small sample sizes. Since the reference bandwidth leads to a certain amount of oversize (see supplementary material) we have adopted the cross validated bandwidth. Remarkably, the test manages to identify correctly the lags at which nonlinear dependence is known to be present for different models, see e.g. the nonlinear dependence at lag 2 of Model 6. Moreover, for $n = 240$ the test is able to discriminate between linear-Gaussian and linear-non Gaussian processes.

As with the traditional correlogram, the results of the test can be depicted in an appealing graphical fashion. For instance, in Fig. 1 we show with the black solid line the mean value of $\hat{T}_k$, $k = 1, \ldots, 5$ over 200 realizations of length $n = 120$ of Model 1 (Gaussian autoregressive process, left) and Model 6 (nonlinear moving average process, right). The green/light gray and the blue/dark gray dashed lines represent the mean quantiles of the distribution of the surrogates at the 95% and 99% levels, respectively.

6·3. *Bootstrap*

The results pertaining to the implementation of the test by means of the smoothed sieve bootstrap (denoted by bootstrap) are shown in Tables 2 and 3. In these instances the size of the test is almost always correct or smaller than the nominal 5% (especially for $\hat{S}_k^u$) even for relatively short time series. At the same time the empirical power is notable for many nonlinear processes. Interestingly, for $\hat{T}_k$ the smoothed sieve bootstrap scheme has very little power against linear non-Gaussian processes so that the surrogate scheme is the choice fo r testing against the linear

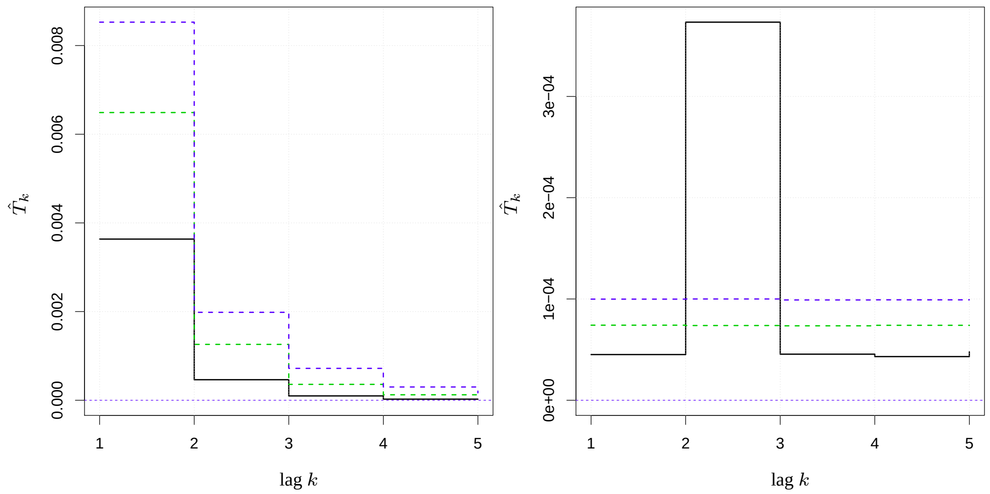

Fig. 1. (Left): mean $\hat{T}_k$, $k = 1, \ldots, 5$ for 200 realizations of Model 1 (Gaussian autoregressive process, black solid line). The rejection bands at 95% and 99% level (green/light gray and blue/dark gray dashed line, respectively) are obtained from the mean quantiles of the surrogates distribution. (Right) the same but for Model 6.

Table 2. $\hat{T}_k$, *bootstrap, reference criterion: rejection percentages,* $k = 1, \ldots, 5$ *at* $\alpha = 5\%$.

	$n = 120$						$n = 240$					
	1	2	3	4	5	C	1	2	3	4	5	C
Model 1	0.5	0.5	3.0	3.0	7.0	11.5	0.0	2.0	5.0	3.5	3.5	8.0
Model 2	0.0	1.0	1.5	4.5	5.0	10.5	0.0	3.5	1.5	4.5	7.5	9.5
Model 3	0.0	2.5	3.5	5.0	5.0	11.5	0.0	0.5	2.5	5.0	8.0	12.0
Model 4	0.0	2.5	8.0	7.5	8.0	17.0	0.0	2.5	6.5	8.0	6.0	12.5
Model 5	44.0	84.0	25.5	42.0	13.5	93.5	62.5	97.5	31.0	66.5	20.5	100.0
Model 6	8.5	94.5	5.5	6.5	3.5	96.0	9.5	99.0	7.5	5.5	11.5	100.0
Model 7	59.5	50.0	44.0	35.5	28.5	83.0	84.5	74.5	69.0	56.0	49.5	94.5
Model 8	10.0	13.0	13.0	11.0	8.0	35.0	11.0	24.0	18.5	11.0	9.0	47.0
Model 9	100.0	99.5	3.5	3.0	3.0	100.0	100.0	100.0	54.5	48.0	51.0	100.0

Gaussian hypothesis (compare the results for Models 3 and 4 in Table 1 and 2). Also, the results with the likelihood cross validation are not as good as those with the plug-in reference bandwidth (see supplementary material). This may be due to the residual based nature of the sieve bootstrap. Recall that in this method, one is centering the residuals around their means, and the likelihood cross validation is seemingly doing a better job of detecting the removal of dependence structure in the centered residuals. In the nonparametric estimation approach, cross validation is well capable of detecting irrelevant regressors; see Li & Racine (2007). The “re-centering” bootstrap is also a good method for removing estimation uncertainty, perhaps due to removal of omitted variables. In our context, “omitted variables” are indeed the nonlinear terms! Hence the lack of power to detect them with cross validation. For this reason, we do not generally recommend the centered residual based bootstrap technique in conjunction with cross validation. Overall, the results indicate that the reference criterion should be paired with the bootstrap scheme, while the MLCV criterion has to be preferred when using the surrogate based test. If we consider the results of the combined tests over the 5 lags (columns denoted with a C) we observe a power gain at the expense of increased type I error rates. In such cases a different combination function or multiple

Table 3. $\hat{S}_k^u$, *bootstrap, reference criterion: rejection percentages,* $k = 1, \ldots, 5$ *at* $\alpha = 5\%$.

	$n = 120$						$n = 240$					
	1	2	3	4	5	C	1	2	3	4	5	C
Model 1	0.0	0.0	0.5	0.5	2.5	3.0	0.0	0.0	0.0	0.0	1.0	1.0
Model 2	0.0	0.0	0.5	1.0	2.0	3.0	0.0	0.0	0.5	1.0	2.5	4.0
Model 3	2.0	1.5	2.5	4.5	3.0	6.5	0.5	1.0	1.0	2.0	3.0	4.0
Model 4	0.5	2.0	5.0	2.5	3.5	9.0	1.0	1.5	1.0	3.5	3.0	3.0
Model 5	35.0	76.5	19.5	37.0	8.0	85.5	52.0	96.5	28.0	61.0	17.5	99.5
Model 6	5.5	89.5	3.0	3.0	0.5	90.5	4.0	100.0	5.0	2.0	8.0	99.0
Model 7	47.0	42.0	35.0	28.5	22.5	74.5	72.5	64.5	56.0	46.0	41.0	92.0
Model 8	24.5	6.5	5.5	6.5	3.5	34.0	45.5	12.5	10.0	4.0	5.0	53.5
Model 9	100.0	90.0	1.0	3.0	5.5	100.0	100.0	100.0	23.0	26.0	24.0	100.0

testing control might be required. The case of $\hat{S}_k^u$ (see Table 3) is remarkable: the power gain has no side effects on the type I error so that the combined test can be used with confidence.d

Finally, a note on the computational complexity of the above mentioned procedures. It can be shown that the computational complexity of the bootstrap-reference implementation is linear with respect to the sample size n. The surrogate-MLCV implementation has a complexity which is quadratic with respect to n, instead. For more details see the supplementary material.

The results shown above confirm that the tests proposed can be successfully applied in many fields. In contrast to the above performance of the metric entropy Granger et al. (2004), among many others, have shown the failure of correlation based methods to detect dependencies that are identified above for these same models and a number of others. In these other contexts, no other nonlinear method has shown more power than the metric entropy method, though some have matched its performance only in the case of some models. This suggests that the use of the metric entropy is a risk averse approach.

7. REAL DATA

In this section we show the results of the application of our tests to real time series. The two series analyzed are described in detail in Tsay (2005) and were taken from the companion R package FinTS. In both cases we have applied two tests: $i)$ the surrogate test with the MLCV bandwidth criterion; $i)$ the bootstrap test with the reference bandwidth criterion. The first series contains the monthly log returns in percentages of IBM stock from January 1960 to December 1998, for overall $n = 468$ observations. The series has a white noise type ACF and PACF. The time plot is shown in Fig. 2(left) while the plot of $\hat{T}_k$ at lags 1:12 is shown in Fig. 2(right).

The second series concerns the daily exchange rate between U.S. dollar and Japanese yen from 2000-01-03 to 2004-03-26. The series has $n = 1063$ observations and has been differenced and log-transformed. Such series has a white noise type ACF, while the PACF results significant at lag 1 (not shown here). The time plot is shown in Fig. 3(left) while the plot of $\hat{T}_k$ at lags 1:24 is shown in Fig. 3(right).

The evidence against linearity is clear in the two series as the two tests provide the same outcome. In particular, for the IBM data there are possible non-linear effects at lags 3 and 5 (see Fig. 2(right)); as for the daily USD-YEN the tests suggest a significant effect at lag 1. If we compare qualitatively the plots of $\hat{T}_k$ for the two series with those obtained from the simulation study we notice similarities with the bilinear process for the IBM series and with a nonlinear moving average for the USD-YEN series. Even if in principle it would be unfeasible to perform a model identification solely on the base of such plots, the information conveyed by our test can

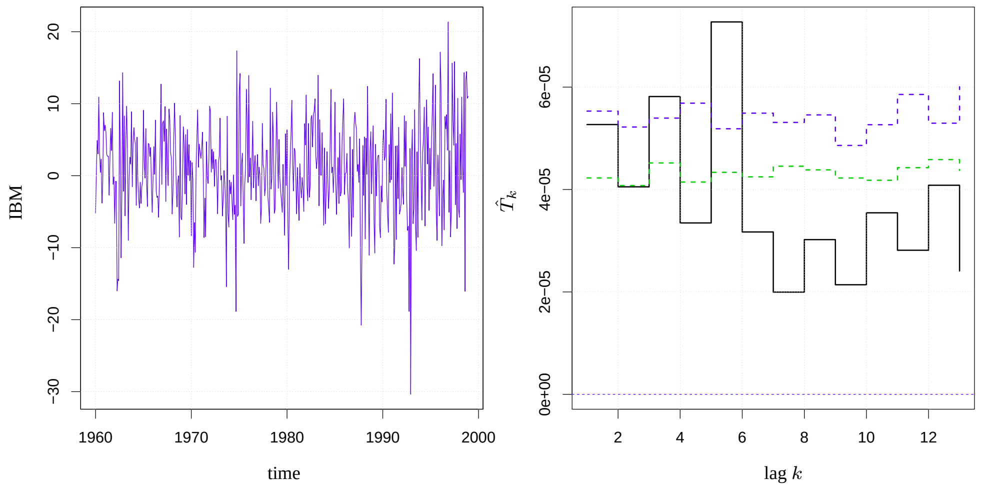

Fig. 2. (Left): time plot of the monthly log returns in percentages of IBM stock from 1960 to 1999. (Right) plot of $\hat{T}_k$ for the IBM series at lags 1:12. The dashed lines indicate the rejection bands at 95% (green/light gray) and at 99% (blue/dark gray).

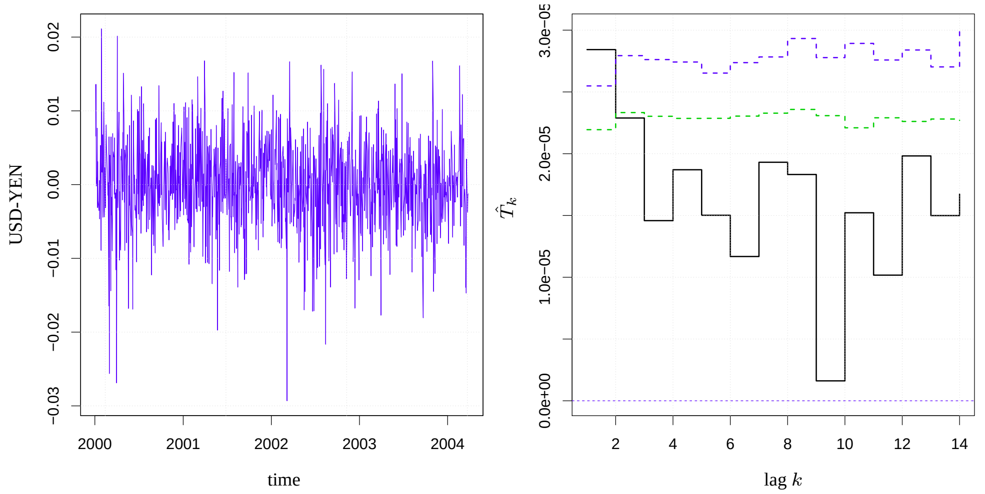

Fig. 3. (Left): time plot of the (differenced and logged) daily exchange rate between U.S. dollar and Japanese yen from 2000-01-03 to 2004-03-26. (Right) plot of $\hat{T}_k$ for the series at lags 1:24. The dashed lines indicate the rejection bands at 95% (green/light gray) and at 99% (blue/dark gray).

help considerably. In this instance, the results point to a complex dependence upon past shocks that is consistent with the findings reported in literature.

8. Conclusions

In this paper we have shown the great potential and flexibility of the test for the detection of nonlinear dependence in time series based upon the combination of the entropy measure S_ρ together with resampling methods. The test proposed turns out to be powerful also in those situation when several other tests may fail. For instance, high frequency time series might show periodicites at distant lags due to the sampling rate. In such a case it would be unfeasible to apply tests that require the building of a nonlinear model or a Volterra series expansions that involve many lags. On the contrary, our test, being based on pairwise comparisons, can be applied at no additional costs. Moreover, S_ρ is a measure of dependence that involves the whole joint distribution function and this gives the test a potential advantage over all those statistics based upon specific moments or aspects of such distributions. A pleasant bonus is that our test, while diagnostic in nature, is able to correctly pin point departures from linearity due to tail effects or threshold phenomena, even for small to moderate sample sizes. Being based on resampling techniques, our tests have a high computational burden. For these reasons we have created a R package that implements a parallel version of all the routines. The package can be found at www2.stat.unibo.it/giannerini/software.html and is forthcoming on CRAN.

In conclusion, we think that our proposal might be fruitfully employed in many applied fields. Future investigations will include the extension of the test to discrete/categorical variables, the derivation of a portmanteau version of the test and the study of different bandwidth selection methods.

ACKNOWLEDGEMENTS

This work has been partially supported by MIUR funds. Simone Giannerini acknowledges the support of the Institute for Mathematical Sciences of the National University of Singapore.

APPENDIX 1

Proofs

Proof of Proposition 1

Given a bivariate standard gaussian density function:

$$f_{X,Y}(x,y) = \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{\frac{x^2-2\rho xy+y^2}{2\left(-1+\rho^2\right)}}$$

with marginal densities $f_X(x) = \frac{1}{(2\pi)^{1/2}} e^{\frac{-x^2}{2}}$ and $f_Y(y) = \frac{1}{(2\pi)^{1/2}} e^{\frac{-y^2}{2}}$, then we may write:

$$\begin{aligned} S_\rho &= 1 - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(f_X(x)\, f_Y(y)\, f_{X,Y}(x,y)\right)^{1/2} dx\, dy \\ &= 1 - \frac{1}{4\pi^2(1-\rho^2)^{1/2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(e^{-\frac{(x^2+y^2)(2-\rho^2)-2\rho xy}{2(1-\rho^2)}} \right)^{\frac{1}{2}} dx\, dy\ . \end{aligned}$$

Now if we operate the polar coordinate transformation $x = r\cos\theta$ and $y = r\sin\theta$ we obtain

$$\begin{aligned} S_\rho &= 1 - \frac{1}{2\pi(1-\rho^2)^{1/4}} \int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2\left(2-\rho^2-\rho\sin 2\theta\right)}{4\left(1-\rho^2\right)}} d\theta\, dr \\ &= 1 - \frac{1}{\pi}\int_0^{2\pi} \frac{(1-\rho^2)^{\frac{3}{4}}}{(\rho^2-2+\rho\sin 2\theta)} d\theta \\ &= 1 - \frac{2\left(1-\rho^2\right)^{3/4}}{(4-5\rho^2+\rho^4)^{1/2}} = 1 - \frac{2\left(1-\rho^2\right)^{1/4}}{(4-\rho^2)^{1/2}} \end{aligned}$$

Proof of Proposition 2

The result follows directly from applying the delta method to $\hat{S}_k^p = g(\hat{\rho}_k)$.

Proof of Proposition 3

Let $\hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$ be the sample estimator of ρ_k, where $\hat{\gamma}_k = n^{-1}\sum_{t=1}^{n-k} X_t\, X_{t+k}$. From Theorem 6.3.5 of Fuller (1996) $\hat{\rho}_k \to \rho_k$ in probability. Now, since $g : [-1,1] \to [0,1]$, $g(x) = 1 - \frac{2\left(1-x^2\right)^{1/4}}{(4-x^2)^{1/2}}$ is a continuous bounded function, then from Theorem 5.1.4 of Fuller (1996) it follows that:

$$\hat{S}_k^p = g(\hat{\rho}_k) \to g(\rho_k) = S_k \quad \text{in probability}$$

Furthermore, since $0 < \hat{S}_k < 1$ almost surely for all k, then, from Theorem 6.2.4 of Sen et al. (2009) it follows that

$$\hat{S}_k^p \to S_k \quad \text{in } L^2.$$

Proof of Proposition 4

1. The proof follows the lines of that of Tjøstheim (1996) since Assumptions (A1)–(A5) allow to apply the same arguments. The quantity to be estimated can also be written as

$$S_k = 1 - \int\int B[u(x_1,x_2)]w(x_1,x_2)dF(x_1,x_2)$$

where $u(x_1,x_2) = \{f_{X_t},(x_1)f_{X_{t+k}}(x_2), f_{X_t,X_{t+k}}(x_1,x_2)\}$ and $B[u(x_1,x_2)] = \left[\frac{f_{X_t}(x_1)f_{X_{t+k}}(x_2)}{f_{X_t,X_{t+k}}(x_1,x_2)}\right]^{\frac{1}{2}}$. so that the nonparametric estimator results:

$$\hat{S}_k^u = 1 - \int\int B[\hat{u}(x_1,x_2)]w(x_1,x_2)d\hat{F}(x_1,x_2)$$

Now we have:

$$\hat{S}_k^u - S_k = \int\int B[u(x_1,x_2)]w(x_1,x_2)\{dF(x_1,x_2) - d\hat{F}(x_1,x_2)\} + \qquad \text{(A1)}$$

$$\int\int (B[u(x_1,x_2)] - B[\hat{u}(x_1,x_2)])w(x_1,x_2)d\hat{F}(x_1,x_2) \qquad \text{(A2)}$$

Due to the ergodic theorem $(A1) \to 0$ in L^2. In order to prove that $(A2) \to 0$ in L^2 note that there exists an integer N such that for $n \geq N$ we have $K_n = \text{pr}[u(x_1,x_2) \in A] = 1$ where A is an open set that includes the support of $u(x_1,x_2)$. Now, by the mean value theorem, there exists a random function $u'(x_1,x_2)$ such that

$$K_n|B[u(x_1,x_2)] - B[\hat{u}(x_1,x_2)]| \leq$$

$$\sum_{i=1}^{3} K_n \left|\frac{\partial B[u'(x_1,x_2)]}{\partial u_i}\right| |u_i(x_1,x_2) - \hat{u}_i(x_1,x_2)| \,.$$

The result follows directly from the boundedness of $\left|\frac{\partial B[u'(x_1,x_2)]}{\partial u_i}\right|$ and from the strong consistency of the kernel density estimators.

2. The regularity conditions (A1)–(A5) allow to apply the theoretical framework outlined in Tjøstheim (1996). The proof follows from that of Tjøstheim (1996) by posing $u(x_1,x_2) = \{f_{X_t},(x_1)f_{X_{t+k}}(x_2), f_{X_t,X_{t+k}}(x_1,x_2)\}$ and $B[u(x_1,x_2)] = \left[\frac{f_{X_t}(x_1)f_{X_{t+k}}(x_2)}{f_{X_t,X_{t+k}}(x_1,x_2)}\right]^{\frac{1}{2}}$.

Proof of Theorem 1

1. The results follows directly from Propositions 3 and 4 and from the algebra of convergence in L^2.
2. From Propositions 3 and 4 and from the algebra of convergence in distribution it follows that $n^{1/2}(S_k^u - S_k) - n^{1/2}(S_k^p - S_k) = n^{1/2}(S_k^u - S_k^p) \to N(0,\sigma_a^2)$ in distribution, where $\sigma_a^2 = \sigma_p^2 + \sigma_u^2$. Hence,

$$\frac{n^{1/2}(S_k^u - S_k^p)}{\sigma_a} = \frac{(n\hat{T}_k)^{1/2}}{\sigma_a} \to N(0,1) \quad \text{in distribution, and} \quad \frac{n\hat{T}_k}{\sigma_a^2} \to \chi_1^2 \quad \text{in distribution.}$$

Proof of Proposition 5

The rationale behind the method lies in the relationship between the sample periodogram $I(\mathbf{x},\omega)$ and the autocovariance function $\hat{\gamma}_k$ of $\mathbf{x}$ (where k is the lag):

$$I(\mathbf{x},\omega) = \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1} \hat{\gamma}_k \exp(-ik\omega)$$

Now, since $\hat{\gamma}_k$ and $I(\mathbf{x}, \omega)$ are related through an invertible function, the preservation of the sample autocorrelation in the surrogate series is equivalent to the preservation of the sample periodogram. In fact, $V = (\bar{x}, \hat{\gamma}_k, k = 1, \dots)$ is a (joint) sufficient statistic for a linear Gaussian process. Moreover, in order to apply the results of Chan (1997) it is straightforward to show that the test statistics $\hat{S}^u_k$ and $\hat{T}_k$ are asymptotically independent of any finite set of X_t where $t \in \mathbb{N}$. To this aim, consider the statistic $\hat{T}_k = \hat{S}^u_k - \hat{S}^p_k$ and let $\mathcal{I} = i_1, \dots, i_N$ a subset of indices of $\mathbb{N}$. We can write:

$$\hat{S}^u_k =$$

$$= \frac{1}{2} \int\int \left(\left\{ \frac{1}{n-k} \sum_{i \in \mathcal{I}} \frac{1}{h_1 h_2} K\left(\frac{x_1 - X_i}{h_1}, \frac{x_2 - X_{i+k}}{h_2}\right) \right\}^{\frac{1}{2}} - \left\{ \frac{1}{n} \sum_{i \in \mathcal{I}} \frac{1}{h_1} K\left(\frac{x - X_i}{h_1}\right) \frac{1}{n} \sum_{i \in \mathcal{I}} \frac{1}{h_2} K\left(\frac{x - X_i}{h_2}\right) \right\}^{\frac{1}{2}} \right)^2 +$$

$$\left(\left\{ \frac{1}{n-k} \sum_{i \in \{\mathbb{N} - \mathcal{I}\}} \frac{1}{h_1 h_2} K\left(\frac{x_1 - X_i}{h_1}, \frac{x_2 - X_{i+k}}{h_2}\right) \right\}^{\frac{1}{2}} - \left\{ \frac{1}{n} \sum_{i \in \{\mathbb{N} - \mathcal{I}\}} \frac{1}{h_1} K\left(\frac{x - X_i}{h_1}\right) \frac{1}{n} \sum_{i \in \{\mathbb{N} - \mathcal{I}\}} \frac{1}{h_2} K\left(\frac{x - X_i}{h_2}\right) \right\}^{\frac{1}{2}} \right)^2 w(x_1, x_2) dx_1 dx_2 \,.$$

Now, since the first term of the integrand function vanishes as $n \to \infty$ and since the estimator $\hat{S}^u_k$ is asymptotically Gaussian with limiting variance that does not depend upon any finite subset of observations then the result follows immediately. The same argument holds for $\hat{S}^u_k$. In fact, let $\hat{\rho}_k$ be the sample autocorrelation function of $\{X_t\}$ at lag k and let $\hat{S}^p_k = 1 - \frac{2(1-\hat{\rho}_k^2)^{1/4}}{(4-\hat{\rho}_k^2)^{1/2}}$. We have that

$$\hat{\rho}_k = \frac{1}{n} \sum_{i \in \{\mathcal{I}\}} X_i X_{i+k} + \frac{1}{n} \sum_{i \in \{\mathbb{N} - \mathcal{I}\}} X_i X_{i+k} \,.$$

Again, since the first of the two terms of the sum vanishes as the sample size diverges and since $\hat{\rho}_k$ is asymptotically Gaussian with limiting variance $\zeta = \sum_{i=1}^{\infty} \left\{ \rho_{(i+k)} \rho_{(i-k)} - 2\rho_i \rho_k \right\}^2$ then the asymptotic independence holds. In turn, since $\hat{S}^p_k$ is a piecewise monotone function of $\hat{\rho}_k$ the result follows.

In conclusion, the assumptions of Chan (1997) are fulfilled so that inference based on the constrained randomization approach is equivalent to that implied by the phase randomization scheme and is asymptotically valid under H_0.

Proof of Proposition 6

In Bickel & Bühlmann (1999) it is shown that the sieve scheme is valid under the assumption of an underlying infinite-order autoregressive process; this covers both H_0 and H_0'.

$\hat{T}_k$ has two components: $\hat{T}_k = \hat{S}^p_k - \hat{S}^u_k$. The parametric component can be written as

$$\hat{S}^p_k = g_1 \left\{ \frac{1}{n-k} \sum_{t=k+1}^{n} h(X_t, X_{t-k}) \right\}$$

namely, it is a nonlinear differentiable function of the linear statistic $\hat{\rho}_k$ where $h(X_1, X_2) = X_1 X_2$ and g_1 is the function of Eq. (4).

The second component $\hat{S}^u_k$, being based on kernel density estimators, can be seen as a functional of the distribution of (X_t, X_{t+k}). This component can be written as

$$\hat{S}^u_k = 1 - \int\int (f_1(x_1) \times f_2(x_2) \times f_{12}(x_1, x_2))^{1/2} \, dx_1 dx_2 =$$

$$1 - \frac{\text{const}}{(n-k)^{1/2}} \sum_{\mathbf{x} \in \mathbb{R}^2} g_2(X_t, X_{t+k})$$

where $f_1 = \hat{f}_{X_t}$, $f_2 = \hat{f}_{X_{t+k}}$ and $f_{12} = \hat{f}_{(X_t, X_{t+k})}$ are the kernel density estimators defined in Eq 5, const is a real constant that depends on n, k, h_1, h_2 and $g_2(x_1, x_2) = (f_1(x_1) f_2(x_2) f_{12}(x_1, x_2))^{1/2}$. Now, from assumptions (A2) and (A3) g_2 is a continuous bounded function and has bounded first derivative on the open interval $(0, \infty)$. Hence, assumption 3.1 of Bickel & Bühlmann (1999) is satisfied and the functional $\hat{T}_k$ fulfils the assumptions of Theorem 4.1 of Bickel & Bühlmann (1999) so that the result follows directly from the consistency of the smoothed sieve bootstrap process. The parametric estimator $\hat{S}^p_k$ fulfils also the conditions of Theorem 3.3 of Bühlmann (1997) so that also a non-smoothed version of the sieve bootstrap would lead to valid inferences for this component.

REFERENCES

BICKEL, P. & BÜHLMANN, P. (1997). Closure of linear processes. *J. Theoret. Probab.* **10**, 445–479. Dedicated to Murray Rosenblatt.

BICKEL, P. & BÜHLMANN, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5**, 413–446.

BROCKWELL, P. & DAVIS, R. (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.

BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3**, 123–148.

BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* **17**, 52–72.

CHAN, K. (1997). On the validity of the method of surrogate data. In *Nonlinear Dynamics and Time Series*, C. D. Cutler & D. T. Kaplan, eds., vol. 11 of *Fields Inst. Communications*. American Math. Soc., pp. 77–97.

CHAN, K. & TONG, H. (2001). *Chaos: a statistical perspective*. New York: Springer Verlag.

FERNANDES, M. & NÉRI, B. (2010). Nonparametric entropy-based tests of independence between stochastic processes. *Econometric Rev.* **29**, 276–306.

FULLER, W. (1996). *Introduction to Statistical Time Series, II ed.* New York: John Wiley & Sons.

GIANNERINI, S. (2012). The quest for nonlinearity in time series. In *Handbook of Statistics: Time Series*, T. Subba Rao & C. Rao, eds., vol. 30 of *Handbook of Statistics*. Elsevier, pp. 43–63.

GRANGER, C. & LIN, J. (1994). Using the mutual information coefficient to identify lags in nonlinear models. *J. Time Ser. Anal.* **15**, 371–384.

GRANGER, C., MAASOUMI, E. & RACINE, J. (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* **25**, 649–669.

HJELLVIK, V. & TJØSTHEIM, D. (1995). Nonparametric tests of linearity for time series. *Biometrika* **82**, 351–368.

HJELLVIK, V., YAO, Q. & TJØSTHEIM, D. (1998). Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* **68**, 295 – 321.

HONG, Y. & WHITE, H. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* **73**, 837–901.

JOE, H. (1989). Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association* **84**, 157–164.

KREISS, J.-P. & FRANKE, J. (1992). Bootstrapping stationary autoregressive moving-average models. *J. Time Ser. Anal.* **13**, 297–317.

LI, Q. & RACINE, J. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

MAASOUMI, E. (1993). A compendium to information theory in economics and econometrics. *Econometric Rev.* **12**, 137–181.

MAASOUMI, E. & RACINE, J. (2009). A robust entropy-based test of asymmetry for discrete and continuous processes. *Econometric Rev.* **28**, 246–261.

MICHEAS, A. & ZOGRAFOS, K. (2006). Measuring stochastic dependence using φ-divergence. *Journal of Multivariate Analysis* **97**, 765 – 784.

RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10**, 441–451.

ROBINSON, P. (1991). Consistent nonparametric entropy-based testing. *The Review of Economic Studies* **58**, pp. 437–453.

SCHREIBER, T. (1998). Constrained randomization of time series data. *Phys. Rev. Lett.* **90**, 2105–2108.

SCHREIBER, T. & SCHMITZ, A. (1996). Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.* **77**, 635–638.

SCHREIBER, T. & SCHMITZ, A. (2000). Surrogate time series. *Physica D* **142**, 346–382.

SEN, P., PEDROSO DE LIMA, A. & SINGER, J. M. (2009). *From finite sample to asymptotic methods in statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Leiden: Cambridge Univ. Press.

SILVERMAN, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

SKAUG, H. & TJØSTHEIM, D. (1996). Testing for serial independence using measures of distance between densities. In *Athens Conference on Applied Probability and Time Series*, P. Robinson & M. Rosenblatt, eds., vol. 115 of *Lecture Notes*. Springer, New York, pp. 363–377.

THEILER, J., EUBANK, S., LONGTIN, A., GALDRIKIAN, B. & FARMER, J. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D* **58**, 77–94.

THEILER, J. & PRICHARD, D. (1996). Constrained-realization monte-carlo method for hypothesis testing. *Physica D* **94**, 221–235.

TJØSTHEIM, D. (1996). Measures of dependence and tests of independence. *Statistics: A Journal of Theoretical and Applied Statistics* **28**, 249–284.

TSAY, R. (2005). *Analysis of Financial Time Series (Wiley Series in Probability and Statistics)*. Wiley-Interscience.