A powerful entropy test for “linearity” against nonlinearity in time series.

Simone Giannerini, Esfandiar Maasoumi, and Estela Bee Dagum

Abstract. In this paper we investigate a test for the identification of nonlinear dependence in time series against more general “nulls” than mere “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 a suitable extension of surrogate data methods, a class of Monte Carlo based tests introduced with the aim of building consistent distribution-free tests for nonlinearity. The use of bootstrap methods is also investigated. In this paper we show how the test can be employed in order to detect the lags at which a significant nonlinear relationship is expected in the same fashion as the autocorrelation function is used for linear processes. We discuss some theoretical aspects related to the entropy measure and its estimators. The power and size of the test is assessed through simulation studies; applications to well known real data are also presented.

1. Introduction

The literature on tests for nonlinearity, or nonlinear serial dependence in time series is vast. Among the various proposals, we may distinguish between i) “specification” tests that aim at assessing the appropriateness of a specific non-linear model, (e.g. TAR, GARCH, etc., see Fan and Yao [2003], Cryer and Chan [2008] and references therein) and ii) “diagnostic” tests that assess the general non-linear character of the process under study (see e.g. Teräsvirta et al. [1993], Lee et al. [1993], Hjellvik and Tjøstheim [1993], Hjellvik et al. [1998], Rusticelli et al. [2009]). Almost all tests in both categories are based on moments of the distribution of the process, and focus on the null of “no nonlinearity”, or “no dependence”. In this paper we propose a test that is based on the whole distribution. 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 latter are accommodated by the use of both “surrogate data” and sieve type bootstrap resampling techniques. We investigate the use of an entropy based metric as a measure of (nonlinear) dependence for time series as an appropriate test statistic. This statistic has been shown to be powerful in other settings, for example in Granger et al. [2004] and Maasoumi and Racine [2009]. In the absence of reliable exact or asymptotic distribution theory for these tests, we examine the use of resampling methods under the null hypothesis of linearity. We are able to show how the combination of the entropy based measure together with resampling techniques can constitute a powerful and flexible test for nonlinearity that can be used in many applied fields.

This work has been partially supported by PRIN funds. The authors would like to thank Qiwei Yao and Francesco Battaglia for useful discussions and comments.
Some further refinements to and deeper understanding of the metric entropy measure of dependence is provided in this paper, including corrections and refinements to properties of such entropies given in Granger et al. [2004]. These include a discovery of a lower upper limit for the metric entropy measure for discrete variables which is dependent on the distribution function, and a normalized version that accommodates both this distribution dependence and draws a closer parallel to how ACF is employed for linear dependence analysis. The normalization is also responsive to another discovery, the non-existence of the density for the case of exactly related continuous variables. Also, a technical correction is offered for the Gaussian case. These findings represent a significant collection of results that are important to note for an increasing large number of practitioners who are adopting these powerful distribution based tests of hypothesis.

In the following section we define the entropy based test statistic and its estimators; also, we discuss some relevant theoretical aspects of the measure. In sections 3 and 4 we describe how the method of surrogate data and bootstrap techniques can be adapted in our context. Section 5 is devoted to a simulation study in which we investigate extensively power and size of the test, while in section 6 we apply the test to real time series.

2. Towards a nonlinear autocorrelation function

There exist many proposals for a 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] and ref. therein). For instance, Shannon mutual information (Kullback-Leibler, AIC) has spread widely in the context of linear and nonlinear dynamics as well as time series analysis. However, most of these entropies do not allow to define a metric since either they do not obey the triangular inequality or they are not commutative operators. These shortcomings are not consequential for some 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. Recent studies describe the properties that an ideal measure of dependence should possess (see e.g. Micheas and Zografos [2006] or Granger et al. [2004] and ref. therein). The measure we study here is the metric entropy measure $S_\rho$, a normalized version of the Bhattacharya-Hellinger-Matusita distance, defined as follows:

$$ S_\rho(k) = \frac{1}{2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left( \sqrt{f_{X_t,X_{t+k}}(x_1,x_2)} - \sqrt{f_{X_t}(x_1)f_{X_{t+k}}(x_2)} \right)^2 dx_1 dx_2 $$

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 symmetrised general "relative" entropy for a particular and unique value, 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, 

1) it is a metric and is defined for both continuous and discrete variables, 
2) 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 (nonlinear) relationship between continuous variables, 

it reduces to a function of the linear correlation coefficient in the case of Gaussian variables, and iv) it is invariant with respect to continuous, strictly increasing transformations.

**Statement of the problem:** The central problem we wish to address in this paper may be exemplified as follows:

Let the Null hypothesis be that \( \{X_t\} \) is a zero mean linear stationary process:

\[
X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{with} \quad \{\varepsilon_t\} \sim N(0, \sigma^2).
\]

where \( \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \) and \( E[X_t] \) is a time series \( x = (x_1, \ldots, x_n) \) and we would like to test whether \( x \) might be operationally considered as a realization of the process of eq. (2.2). Also, let \( \hat{\rho}_k \) be the sample autocorrelation function of \( x \) at lag \( k \) and let \( \hat{S}_\rho(k) \) be the corresponding sample estimator of \( S_\rho(k) \) at lag \( k \) based on Eq. (2.1). The latter estimator is intended to be the nonparametric implementation of [Granger et al. 2004]. The motivation for the use of \( S_\rho \) lies in its good theoretical properties as a measure of dependence. In fact, we can think it as a powerful tool for discriminating between linear and nonlinear dependence. Hence, the alternative hypothesis is a set of nonlinear processes which are identified by nonlinear dependence at various lags, when the nonparametric entropy measure is significant. The challenge is to generate the null distribution of the test statistic under the null hypothesis of Eq. (2.2). In the following, we will show how to address the issue by means of resampling methods that are adapted to accommodate this type of problem.

**Theoretical Properties and a normalization of the metric entropy measure.** 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]. Rényi proposed a set of seven axioms that an ideal measure of bivariate dependence should fulfill. In [Micheas and Zografos 2006], the discussion has been extended to measures of multivariate dependence based on the \( \varphi \)-divergence and three more axioms were introduced. The authors discuss also the properties of the Matusita distance \( M = 2S_\rho \) and notice that such measure fulfills all the axioms considered. Also, it might be useful to restate Axioms (A5) and (A5)' of their paper:

(A5): The measure takes its maximum value if and only if there exists a measurable exact relationship between the variables.

(A5)': The measure takes its maximum value if and only if the probability measures \( P \) and \( Q \) are singular; here \( P \) is the measure associated to the joint distribution of the variables, whereas \( Q \) is the probability measure associated to the product of the marginal distributions.

The authors assert the equivalence of these two axioms, but, as a matter of fact, such equivalence holds only for discrete random variables. In fact, in the case of perfect dependence, the joint probability measure does not admit a density function as shown in the following. Let \( X \) and \( Y \) be random variables absolutely continuous with respect to the Lebesgue measure, with density functions \( h(x) \) and \( g(y) \), respectively. Assume further that there exist their joint density function
\[ f(x, y) = \lim_{\partial x \to 0} \lim_{\partial y \to 0} \frac{Pr(x \leq X \leq x + \partial x, y \leq y + \partial y)}{\partial x \partial y}. \]

Now, if we let \( X \to Y \) almost surely, then we have

\[ f(x, y) = \lim_{\partial x \to 0} \frac{Pr(x \leq X \leq x + \partial x, x \leq X \leq x + \partial x)}{\partial x \partial x} \neq h(x). \]

Hence, the joint density does not reduce to the marginal density so that the measure is not defined. Such remark applies to all other measures defined in terms of the density functions. Notice that, as a matter of facts, this results does not pose practical problems; in fact, in the unlikely situations of perfect dependence, a small quantity of noise can be added to the series.

In the case of discrete random variables the problem is not present. Nevertheless, in such case it is worth considering another issue that, apparently, has not been discussed before. As stated in [Granger et al. 2004] and reported above in point ii), \( S_\rho \) ranges in [0, 1]. However, this appears to be true only for continuous random variables. In case of discrete processes, the maximum value that \( S_\rho \) can take is not 1 and will depend on the kind of distributions involved. We illustrate this point with an example involving Bernoulli variables.

**Proposition 1.** Let \( X \sim Ber(p_x) \) and \( Y \sim Ber(p_y) \) be Bernoulli random variables with parameters \( p_x \) and \( p_y \), respectively. Let also \( p_{xy} = P(X = 1, Y = 1) \) be the joint probability. Clearly, in the case of perfect dependence we have \( p_{xy} = p_x = p_y = p \) so that:

\[ S_\rho = 1 - p^{3/2} - (1 - p)^{3/2} \quad (2.3) \]

As also shown in Figure [i] left), such a function takes its maximum when \( p = 1/2 \) for which we have \( S_\rho = 1 - \sqrt{\frac{1}{2}} \approx 0.293 \) which is far from 1. Hence, besides the strength of the dependence, the maximum possible value of \( S_\rho \) will typically depend on the kind of marginal distributions and their associated parameters. This problem makes unfeasible the comparison between different sequences on the basis of \( S_\rho \) as is. Also, it is clear from Figure [i] left), that two series can be perfectly correlated but \( S_\rho \) tends to 0 as \( p \) approaches 0 or 1.

A possible solution to this issue is to rescale along the lines of the autocorrelation function:

\[ S_\rho^*(k) = \frac{S_\rho(k)}{S_\rho(0)} \quad k = 1, \ldots \]

Once again, note this issue is shared with all those measure defined in terms of density functions, including those belonging to the class of the \( \phi \)-divergence discussed in [Micheas and Zografos 2006]. As 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 relation:

**Proposition 2.** Let \( (X, Y) \sim N(0, 1, \rho) \) be a standard Normal random vector with joint probability density function given by \( f_{X,Y}(\cdot, \cdot, \rho) \), where \( \rho \) is the correlation
Figure 1. (Left) Relationship between $S_\rho$ and the parameter $p$ of a Bernoulli random variable in the perfect dependence scenario, see Eq. (2.3). (Right) Relationship between $S_\rho$ and the correlation coefficient $\rho$ in the Gaussian case, see Eq. (2.4).

Coefficient, and with marginal pdfs $f_X(\cdot)$ and $f_Y(\cdot)$, respectively. Then, the following relation holds:

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

Proof. See Appendix.

In Figure 1(right) we present the picture of such relation. The inspection of the plot reveals a remarkable steepness when $\rho$ is close to $|1|$. In other words, $S_\rho$ appears to be particularly sensitive for detecting and distinguishing between high levels of dependence. This property might be useful for characterizing nearly integrated Gaussian processes. In the following we derive the asymptotic distribution of the estimator for $S_\rho$ based on the sample correlation coefficient (Eq. (2.4)) under the hypothesis of a linear Gaussian process.

Theorem 3. Let $\{X_t\}$ be the zero mean stationary process:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \text{ with } \{\varepsilon_t\} \sim N(0, \sigma^2)$$

where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $E[X_t]^4 < \infty$. Also, let $\hat{\rho}_k$ be the sample autocorrelation function of $\{X_t\}$ at lag $k$ and let $\hat{S}_\rho(k)$ be the corresponding sample estimator of $S_\rho(k)$ at lag $k$ based on Eq. (2.4). Lastly, define the following function $g : [-1, 1] \to \mathbb{R}^+$. $g(x) = 1 - \frac{2(1 - x^2)^{1/4}}{\sqrt{4 - x^2}}$. The function $g$ is differentiable and $g'(x) \neq 0$ for $x \neq 0$. Then, for every $k = 0, 1, \ldots$ we have:

$$\sqrt{n}(\hat{S}_\rho(k) - S_\rho(k)) \xrightarrow{d} N(0, [g'(w)]^2)$$

where $w = \sum_{i=1}^{\infty} \left\{ \rho_{i+k} \rho_{i-k} - 2 \rho_i \rho_k \right\}^2$ is the asymptotic variance of $\hat{\rho}_k$ [see [Brockwell and Davis, 1991, pp. 221-222]].
Proof. See Appendix.

Notice that in the case of no correlation ($\rho_k = 0$) we have $g'(\rho_k) = 0$. In such a case the approximation is driven by higher order derivatives, in particular the even-order ones. The issue of the estimation of $S_{\rho}(k)$ under conditions that allow its use in tests for serial dependence has been studied in Granger et al. [2004], Fernandes and Néri [2010], see also Hong and White [2005]. Both articles make use of nonparametric kernel estimation and build tests for independence but under different conditions. In particular, in Fernandes and Néri [2010] estimators for $\beta$-mixing processes are investigated. In the present paper we extend the use of $S_{\rho}(k)$ to the case where the null hypothesis is that of linearity; to this aim we investigate different methods in order to obtain the distribution of $S_{\rho}(k)$ under $H_0$. Note that since $S_{\rho}(k)$ is a measure of dependence that involves the whole bivariate distribution of the process under study, a test based upon it can fully account for possible non-linear effects. This feature clearly distinguishes our proposal from those that employ statistics built on specific moments of the distribution of $(X_t, X_{t+k})$. In other applications, it has been shown to be at least as powerful as even the existing characteristic function tests, such as tests of time reversibility [Racine and Maasoumi, 2007]. The practical implementation of $S_{\rho}$ requires (i) the estimation of two univariate and one bivariate densities, for which kernel methods have been used, and (ii) the computation of a double integral for which adaptive quadrature methods have been employed. In general, while the latter issue does not pose particular problems, the performance of the test will be influenced by the choice of the method for bandwidth selection. As also mentioned in Granger et al. [2004], maximum likelihood cross validation (MLCV) is suggested in this context even though we have investigated the use of the normal reference method as well.

3. Surrogate data methods

The method of surrogate data was originally introduced in the context of nonlinear time series analysis motivated by chaos theory and can be regarded as a resampling approach for building tests for nonlinearity in the absence of reliable distribution theory for the statistics employed in that field. 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. Indeed, some of the ideas that can be found in the literature on surrogate data analysis are original and can be more generally useful. 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 (DGP) is formulated; for instance, $H_0$: the DGP 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 $H_0$, for instance, the nonlinear prediction error, is computed on the surrogates obtaining the distribution of the test statistic under $H_0$. Thus, the significance level of the test can be derived in the usual way by comparing such distribution with the value that the statistic takes on the original series. One may notice that the basic principle behind surrogate data tests is the same as the bootstrap principle, so that the two approaches should be unified. Nevertheless, in this paper we will treat them separately as to make clear the historical distinctions between them and the contributions coming from different fields.
In Theiler et al. [1992] and Theiler and 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. In brief, it is assumed that the DGP under the null is a linear Gaussian process as in Eq. (2.2). Also, it is assumed that the process admits a spectral density function that forms a Fourier pair with the autocovariance function. Now, given an observed series \( x = (x_1, \ldots, x_n)^T \) we can define the discrete Fourier transform

\[
\zeta_x(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} x_t \exp(-i\omega t), \quad -\pi \leq \omega \leq \pi
\]

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

\[
x_t = \bar{x} + \sqrt{\frac{2\pi}{n}} \sum_{j=1}^{\frac{n-1}{2}} 2 \sqrt{I(x, \omega_j)} \cos(\omega_j t + \theta_j)
\]

Such a relation allows to obtain a surrogate series \( y = (y_1, \ldots, y_n)^T \) by randomizing the phases of the above equation as to obtain

\[
y_t = \bar{x} + \sqrt{\frac{2\pi}{n}} \sum_{j=1}^{m} 2 \sqrt{I(x, \omega_j)} \cos(\omega_j t + \theta_j)
\]

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 of 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 DGP is a stationary Gaussian circular process. By valid, it is meant that test based upon it are similar (they have a Neyman structure). Also, the author gives some arguments for which the method might be asymptotically valid for the null hypothesis that the DGP is a stationary Gaussian process with fast-decaying autocorrelations (see also Chan and Tong [2001], Chap. 4.4).

After the first contributions, several works that extend the method in various directions have been presented. In general, such extensions involve the introduction of either new test statistics or ad hoc algorithms for generating surrogates in order to test specific (not necessarily linear) hypotheses. For instance, Small and Judd [1998], Small et al. [2001] Small [2005] discuss the issue of obtaining pivotal statistics, in particular, they study a class of such statistics based on the correlation integral. For interesting reviews on the topic see Schreiber and Schmitz [2000] and Kugiumtzis [2001, 2008]. Note that, apart from the notable exception of Chan [1997], despite the fair amount of literature on surrogate data methods, to our knowledge, comprehensive studies on the theoretical properties of such tests are lacking.

3.1. The constrained randomization approach. The approach we have adopted in this paper is an extension of the method that allows to treat the problem within a general unified framework (see Schreiber [1998], Schreiber and Schmitz [2000]). Basically, the generation of surrogate time series is seen as a constrained optimization problem that is solved by means of simulated annealing (see e.g. Vidal...
The method has shown to overcome some of the problems linked to phase randomization techniques (see e.g. Schreiber and Schmitz [1996]). Also, all the test statistics produced act as pivotal due to the constrained randomization (see Theiler and 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. Usually, the constraints are defined by the null hypothesis to be tested. The simulated annealing is a useful tool for combinatorial minimization of functions with false minima. The cost function $C$ is interpreted as an energy in a thermodynamic system. Minimizing $C$ is then equivalent to finding the ground state of a system. The annealing is a procedure used to bring a glassy solid close to the optimal state by first heating it and then cooling it. The simulation of this tempering procedure exploits the fact that in thermodynamic equilibrium at some finite temperature $T$, the possible configurations of the system are visited with a probability according to the Boltzmann distribution of the canonical ensemble $p = e^{-\frac{C}{T}}$. In the simulation, this is attained by accepting changes of the configuration with a probability $p = 1$ if the energy is decreased ($\Delta C < 0$) and $p = e^{-\frac{\Delta C}{T}}$ if the energy is increased ($\Delta C \geq 0$). This rule is often referred to as the Metropolis step. In the minimization problem, the temperature is the parameter in the Boltzmann distribution that determines the probability of accepting the unfavourable changes that are needed to get out of false minima.

In the following we will describe in detail the algorithm for testing the hypothesis that the DGP is a linear gaussian process. To this aim we will generate surrogate series having the same autocorrelation function (ACF) and the same sample mean of the original series. The rationale behind the method lies in the relationship between the sample periodogram $I(x, \omega) = |\tilde{\zeta}_x(\omega)|^2$ and the autocovariance function $\hat{\gamma}_k$ of $x$ (where $k$ is the lag):

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

Clearly, as shown in the above discussion, since from $\hat{\gamma}_k$ it possible to obtain the periodogram $I(x, \omega)$ and vice versa the preservation of the sample autocorrelation in the surrogate series is equivalent to the preservation of the sample periodogram.

In Schreiber and Schmitz [1996] it is shown that such an approach leads to tests with better power and size with respect to those derived through the phase randomization. The procedure is the following. Let $x$ be an observed series of length $n$ being $\hat{\rho}_k$ its ACF at lag $k$ and let $x^*$ be the candidate surrogate with ACF at lag $k$ denoted by $\hat{\rho}_k^*$. The cost function we have adopted here is

$$C(x, x^*) = \max_{k=1}^{n/4} |\hat{\rho}_k - \hat{\rho}_k^*|.$$  

Notice that other cost functions based on a discrepancy between two ACFs can be employed. First, denote with $x^*$ a random permutation of the original series $x$; also, start with a temperature $T$. For each temperature value $T$, the algorithm is the following:

1. swap two observations of $x^*$ and obtain the series $x^{*(s)}$;
2. compute $\Delta C = |C(x, x^{*(s)}) - C(x, x^*)|$. 

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

Clearly, as shown in the above discussion, since from $\hat{\gamma}_k$ it possible to obtain the periodogram $I(x, \omega)$ and vice versa the preservation of the sample autocorrelation in the surrogate series is equivalent to the preservation of the sample periodogram.
if $\Delta C < 0$ accept the swap, that is, $x^* = x^{* (s)}$ 
if $\Delta C \geq 0$ accept the swap with probability $p = e^{-\Delta C}$;

(4) repeat step (1) – (3) until either the number of accepted swaps or the number of trials reach specified thresholds;

(5) lower the temperature $T$, for instance by setting $T = \alpha T$ where $\alpha < 1$;

(6) repeat the whole procedure until the cost function reaches a specified threshold;

In general, the choice of the parameters for the algorithm is problem-specific and a certain amount of experimentation and tuning are expected in order to obtain good results. In our instance, we have found parameters’ settings, dependent on the series’ length $n$, that can be employed almost automatically. It is important to notice that the accuracy of the results is a user defined parameter. In our case we have set 0.01 as the threshold for the cost function $C$, namely, for the first $n/4$ lags, the maximum discrepancy between the ACF of the original series and that of the surrogates does not exceed 0.01. While this requirement appears reasonable for our purposes, more stringent values can be chosen if needed at the cost of longer computing times. Clearly, the great flexibility of this method is that different hypotheses can be tested simply by changing the cost function. However, note that the computational cost is a problem that has to be taken into account. For a more detailed discussion on different cooling schemes, cost functions and computational issues related to the algorithm see Schreiber and Schmitz [2000].

4. The bootstrap approach

The second approach considered is based on bootstrap methods and involves modelling procedures. The main idea relies upon the Wold decomposition of a stationary stochastic process. A stationary stochastic process $X_t$ can be decomposed into the sum of two uncorrelated processes: $X_t = Z_t + V_t$ where $Z_t$ is the MA($\infty$) component:

\[(4.1) \quad Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1 \quad \text{with} \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty, \{\varepsilon_t\} \sim WN(0, \sigma^2_{\varepsilon}).\]

$V_t$ is a deterministic component $V_t = \mu + \sum_{j=1}^{\infty} [\alpha_j \sin(\omega_j t) + \beta_j \cos(\omega_j t)]$. If $\{V_t\} = 0$ (purely non-deterministic process) $X_t$ admits a MA($\infty$) representation. This fact was at the basis of the extension of classical residual based bootstrap methods to linear processes (see e.g. Berkowitz and Kilian [2000], Politis [2003]). Such methods can be adapted as follows in order to obtain surrogates (see also Hinich et al. [2005]):

(1) Given a time series $(x_1, \ldots, x_n)$ fit a AR($p$) model

\[X_t = \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_t\]

and obtain the estimates $(\hat{a}_{1,n}(p), \ldots, \hat{a}_{p,n}(p))$

(2) Obtain the residuals.

\[\hat{\varepsilon}_{t,n} = x_t - \sum_{i=1}^{p} \hat{a}_{i,n}(p)x_{t-i}, \quad t = p + 1, \ldots, n\]

Under the null hypothesis of linearity such residuals will form a i.i.d. process.
(3) Resample the residuals with replacement from the (centered) empirical cdf:

\[ \hat{F}_n(x) = \frac{1}{n-p} \sum_{t=p+1}^{n} I(\hat{\varepsilon}_{t,n} - \hat{\varepsilon}_{\cdot,n} \leq x), \quad \text{with} \quad \hat{\varepsilon}_{\cdot,n} = \frac{1}{n-p} \sum_{t=p+1}^{n} \hat{\varepsilon}_{t,n} \]

denote the sample obtained with \((e^*_Q, \ldots, e^*_{-n})\), \(Q \in \mathbb{N}\).

(4) Obtain the surrogate time series according to:

\[ x^*_{t,n} = \sum_{i=1}^{p} \hat{a}_{i,n}(p)x^*_{t-i,n} + e^*_{t}, \quad t = -Q, \ldots, n \]

where the initial values are \(x^*_{-Q-1,n} = \cdots = x^*_{-Q-p,n} = 0\).

(5) Repeat steps (3) and (4) \(B\) times.

The method we have chosen for obtaining the transient set is not the only possible solution. Notice that in our instance we do not condition every surrogate realization to the same initial values so that we avoid problems related to such a choice, see also Berkowitz and Kilian [2000] for a discussion. The validity of the scheme underlying this approach has been investigated in a number of studies and is sometimes referred to as the sieve bootstrap (see e.g. Kreiss and Franke [1992], Bühlmann [1997, 2002]). The surrogates obtained through this scheme will be realizations of linear processes having the same correlation structure as the original data.

One of the main issues related to this method is the choice of the order of the fitted process \(p\). In practice a finite order must be chosen, nevertheless, several authors have stressed that if we require \(p = p(n) \to \infty\) as \(n \to \infty\) but \(\frac{p(n)}{n} \to 0\), where \(k\) is a small integer, then such method becomes non parametric in that the procedure is valid for the whole class of AR(\(\infty\)) processes. In particular, in Bühlmann [1997] it is shown that if the AIC criterion for order 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.

As mentioned above, this scheme is valid for processes that admit a linear representation as in Eq. (4.1). Interestingly, as discussed in Bickel and 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 belong to the MA(\(\infty\)) class such as threshold processes. From the point of view of testing for non linearity these facts pose several important questions that we will try to investigate. First of all, it is clear that tests based upon this scheme will cover a hypothesis which is somehow different to that built upon surrogate data methods as previously described. The alternative hypothesis here is that the process under study does not admit a linear representation as in Eq. (4.1). In view of the above discussion we expect this test to be more conservative than the surrogate based one. Second, since the version of the sieve bootstrap employed here might be no longer valid under the alternative hypothesis the crucial issue we investigate is whether this fact can give discriminatory power to our test statistic. Third, it would be worth investigating the performance of the test when
a fixed order model (e.g. an ARMA(1,1)) that does not depend on the length of the series \( n \) is used.

5. Simulation Results

In this section we show how the entropy dependence metric \( S_\rho \) can serve as a useful tool for characterizing nonlinear processes and identifying the lags at which nonlinear effects might be expected. In this respect, \( S_\rho \) can somehow be employed as the nonlinear counterpart of the ACF. In the following we report the results of a simulation study performed on the following models:

- **Model 1: AR(1)**
  \[ x_t = 0.8 x_{t-1} + \varepsilon_t \]

- **Model 2: ARMA(1,1)**
  \[ x_t = 0.6 x_{t-1} + 0.4 \varepsilon_{t-1} + \varepsilon_t \]

- **Model 3: Bilinear**
  \[ x_t = 0.6 \varepsilon_{t-1} x_{t-2} + \varepsilon_t \]

- **Model 4: NLMA**
  \[ x_t = 0.8 \varepsilon_{t-2}^2 + \varepsilon_t \]

- **Model 5: GARCH(1,1)**
  \[ x_t = \sigma_t \varepsilon_t \]
  \[ \sigma_t^2 = 0.1 + 0.6 \sigma_{t-1}^2 + 0.3 x_{t-1}^2 \]

- **Model 8: SETAR**
  \[ x_t = \begin{cases} 
  -0.8 x_{t-1} + \varepsilon_t & \text{if } x_{t-1} \leq 0 \\
  0.8 x_{t-1} + \varepsilon_t & \text{if } x_{t-1} > 0 
  \end{cases} \]

- **Model 9: NLAR**
  \[ x_t = 4 x_t (1 - x_t) + \sigma_t \eta_t \text{ with } (\eta_t + \frac{1}{2}) \sim \text{Beta}(10, 10) \]
  \[ \sigma_t = \min[4 x_{t-1} (1 - x_{t-1}), 1 - 4 x_{t-1} (1 - x_{t-1})] \]

where \( \varepsilon_t \sim \text{i.i.d. N}(0, 1) \). The null hypothesis tested here is that the DGP is a stationary linear stochastic process so that its dependence structure is fully captured through the ACF. The surrogates were obtained under the scheme presented in the previous sections. If the DGP is nonlinear we expect that the value of \( S_\rho \) will differ significantly from those obtained under \( H_0 \). Notice that Model 9 is the logistic map of a chaotic regime with additive noise. In prior studies, for example by Granger et al (2004), the null hypothesis was one of iid, or (implicitly) iid Gaussian, with no dependence at all lags mitigating the need for the surrogate and Seive approaches of the current paper. Note that the results pertaining to the first two models will give an indication of the size of the test while the results for the other models will reflect its discriminating power.

5.1. SA surrogates. In this section we report the results regarding the test implemented through the constrained randomization method which we denote by SA surrogates. Each surrogate is a random permutation of the original series possessing exactly its ACF (up to the max. tolerance of 0.01). In Table 1 we present the empirical rejection frequencies at the 95% level for the test where the “reference” (plug-in) bandwidth have been used and for lengths of the sample sizes set to \( n = 120 \) and \( n = 240 \), respectively.

The rejection frequencies of Table 1 clearly show high power in almost every situation and even for a limited series’ length. Note that, with plug in bandwidth we obtain some false rejection rates higher than expected. This highlights the need for cross validated bandwidth, and the problem is not present if the MLCV criterion is used instead, see Table 2. Remarkably, the test manages to identify correctly the lags at which nonlinear dependence is known to be present for different models. As with the traditional correlogram the results of the test can be depicted in an appealing graphical fashion.

For instance, in Figure 2 we show with the black solid line the mean value of \( S_\rho(k), k = 1, \ldots, 10 \) over 200 realizations of length \( n = 120 \) of Model 1 (AR(1))
Table 1. SA surrogates, Plug-in: Monte Carlo rejection frequencies for $S_p(k)$, $k = 1,\ldots,10$ at level 95% for the processes considered with 200 replications, $B = 200$ surrogates.

<table>
<thead>
<tr>
<th>$n = 120$</th>
<th>lag $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.210</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.200</td>
</tr>
<tr>
<td>Bilinear</td>
<td>0.365</td>
</tr>
<tr>
<td>NLMA</td>
<td>0.960</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.580</td>
</tr>
<tr>
<td>SETAR</td>
<td>0.810</td>
</tr>
<tr>
<td>NLAR</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 240$</th>
<th>lag $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.280</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.300</td>
</tr>
<tr>
<td>Bilinear</td>
<td>0.445</td>
</tr>
<tr>
<td>NLMA</td>
<td>0.090</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.735</td>
</tr>
<tr>
<td>SETAR</td>
<td>0.915</td>
</tr>
<tr>
<td>NLAR</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 2. SA surrogates, MLCV criterion: Monte Carlo rejection frequencies for $S_p(k)$, $k = 1,\ldots,10$ at level 95% for the processes considered with 200 replications, $B = 200$ surrogates.

<table>
<thead>
<tr>
<th>$n = 120$</th>
<th>lag $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.105</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.865</td>
</tr>
<tr>
<td>Bilinear</td>
<td>0.215</td>
</tr>
<tr>
<td>NLMA</td>
<td>0.065</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.395</td>
</tr>
<tr>
<td>SETAR</td>
<td>0.575</td>
</tr>
<tr>
<td>NLAR</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 240$</th>
<th>lag $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.165</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.135</td>
</tr>
<tr>
<td>Bilinear</td>
<td>0.375</td>
</tr>
<tr>
<td>NLMA</td>
<td>0.100</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.670</td>
</tr>
<tr>
<td>SETAR</td>
<td>0.885</td>
</tr>
<tr>
<td>NLAR</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Figure 2. (Left): $S_\rho(k)$, $k = 1, \ldots, 10$ for 200 realizations of an AR(1) process (Model 1) (black solid line). The confidence bands at the 95% level (green dashed line) and at 99% level (blue dashed line) were obtained from the mean quantiles of the surrogates distribution. (Right) the same but for the NLMA Model (Model 4).

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.

5.2. AR surrogates-bootstrap. The results pertaining to the implementation of the test by means of bootstrap methods (denoted by AR surrogates) are shown in Table 3 and 4; the simulation setting is the same as the previous section. Interestingly, in this case the size of the test is almost always correct or slightly smaller than the nominal 5% even for short series. As for the plug-in/reference bandwidth (Table 3) the empirical power for $n = 240$ is remarkable for many processes indicating that the test can be successfully employed in many fields.

Interestingly, the results with ML cross validation are not as good as those with the plug-in reference bandwidth. 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 ML cross validation is seemingly doing a better job of detecting the removal of dependence structure in the centered residuals. It is worth noting that in the NP estimation approach, cross validation is well capable of detecting irrelevant regressors; see Li and Racine (their 2008 textbook on NP). 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 the nonlinear terms! Hence the lack of power to detect them. For this reason,
Table 3. AR surrogates, Plug-in bandwidth: Monte Carlo rejection frequencies for $S_p(k)$, $k = 1, \ldots, 10$ at level 95% for the processes considered with 200 replications, $B = 200$ surrogates.

<table>
<thead>
<tr>
<th>$n = 120$</th>
<th>lag $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1   2   3   4   5   6   7   8   9   10</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.00 0.02 0.015 0.02 0.035 0.025 0.03 0.045 0.055 0.065</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.00 0.00 0.01 0.01 0.02 0.02 0.015 0.03 0.04 0.055</td>
</tr>
<tr>
<td>Bilinear</td>
<td>0.32 0.73 0.12 0.33 0.08 0.14 0.065 0.105 0.075 0.065</td>
</tr>
<tr>
<td>NLMA</td>
<td>0.05 0.865 0.05 0.06 0.05 0.04 0.075 0.055 0.065 0.065</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.39 0.345 0.29 0.235 0.24 0.22 0.135 0.140 0.140 0.155</td>
</tr>
<tr>
<td>SETAR</td>
<td>0.275 0.12 0.085 0.055 0.035 0.04 0.06 0.065 0.055 0.055</td>
</tr>
<tr>
<td>NLAR</td>
<td>1.00 0.935 0.015 0.045 0.05 0.045 0.055 0.065 0.075 0.065</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 240$</th>
<th>lag $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1   2   3   4   5   6   7   8   9   10</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.00 0.00 0.00 0.00 0.005 0.005 0.015 0.025 0.035 0.025</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.00 0.00 0.01 0.015 0.015 0.015 0.035 0.02 0.04 0.05</td>
</tr>
<tr>
<td>Bilinear</td>
<td>0.42 0.935 0.22 0.555 0.16 0.28 0.13 0.20 0.09 0.14</td>
</tr>
<tr>
<td>NLMA</td>
<td>0.075 0.99 0.075 0.04 0.055 0.065 0.075 0.055 0.04 0.115</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.695 0.60 0.57 0.43 0.39 0.345 0.275 0.285 0.205 0.215</td>
</tr>
<tr>
<td>SETAR</td>
<td>0.465 0.15 0.09 0.065 0.075 0.035 0.045 0.04 0.055 0.075</td>
</tr>
<tr>
<td>NLAR</td>
<td>1.00 1.00 0.045 0.04 0.035 0.075 0.075 0.115 0.08 0.075</td>
</tr>
</tbody>
</table>

Table 4. AR surrogates, MLCV criterion: Monte Carlo rejection frequencies for $S_p(k)$, $k = 1, \ldots, 10$ at level 95% for the processes considered with 200 replications, $B = 200$ surrogates.

<table>
<thead>
<tr>
<th>$n = 120$</th>
<th>lag $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1   2   3   4   5   6   7   8   9   10</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.01 0.02 0.025 0.055 0.055 0.05 0.035 0.06 0.055 0.075</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.01 0.03 0.025 0.025 0.035 0.05 0.035 0.025 0.02 0.05</td>
</tr>
<tr>
<td>Bilinear</td>
<td>0.025 0.055 0.015 0.03 0.025 0.035 0.015 0.02 0.025 0.025</td>
</tr>
<tr>
<td>NLMA</td>
<td>0.01 0.05 0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.005</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.035 0.04 0.03 0.045 0.03 0.025 0.045 0.035 0.05 0.03</td>
</tr>
<tr>
<td>SETAR</td>
<td>0.205 0.075 0.04 0.065 0.045 0.045 0.045 0.05 0.03 0.04</td>
</tr>
<tr>
<td>NLAR</td>
<td>1.00 0.99 0.95 0.61 0.21 0.22 0.235 0.225 0.23 0.225</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 240$</th>
<th>lag $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1   2   3   4   5   6   7   8   9   10</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.01 0.01 0.045 0.035 0.02 0.045 0.03 0.035 0.03 0.035</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>0.02 0.00 0.02 0.02 0.03 0.04 0.045 0.045 0.035 0.04</td>
</tr>
<tr>
<td>Bilinear</td>
<td>0.03 0.19 0.02 0.045 0.015 0.05 0.02 0.02 0.02 0.02</td>
</tr>
<tr>
<td>NLMA</td>
<td>0.005 0.175 0.00 0.00 0.00 0.005 0.005 0.00 0.01 0.005</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.075 0.06 0.055 0.055 0.035 0.025 0.025 0.02 0.03 0.02</td>
</tr>
<tr>
<td>SETAR</td>
<td>0.415 0.11 0.07 0.045 0.055 0.055 0.055 0.06 0.045 0.06</td>
</tr>
<tr>
<td>NLAR</td>
<td>1.00 1.00 1.00 1.00 0.71 0.285 0.285 0.285 0.28 0.28</td>
</tr>
</tbody>
</table>

we do not generally recommend the centered residual based bootstrap technique, especially with cross validation.
Figure 3. (Left): $S_ρ(k)$, $k = 1, \ldots, 10$ for 200 realizations of a Bilinear process (Model 3) (black solid line). The confidence bands at the 95% level (green dashed line) and at 99% level (blue dashed line) were obtained from the mean quantiles of the surrogates distribution. (Right) the same but for the SETAR process (Model 8).

Analogously to Figure 2 we show in Figure 3 the results pertaining to the Bilinear and to the SETAR processes.

The results shown above somehow confirm our expectations deriving from theoretical arguments; the AR based test is more conservative than that based on surrogate data. In particular we notice a loss of discriminatory power for those processes that do not admit a MA($\infty$) representation but for which the sieve bootstrap appears to deliver good results. These facts suggest that it might be worth to investigate in future works the performance of the test when a misspecified or a fixed order model is used.

6. Real Data

In this section we show the results of the application of our tests to real time series. The first 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 SA surrogate test with the MLCV bandwidth criterion; $i)$ the AR surrogate 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 Figure 4(left) while the plot of $S_ρ$ at lags 1:12 is shown in Figure 4(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 Figure 5(left) while the plot of $S_ρ$ at lags 1:24 is shown in Figure 5(right).

The results are summarized in Table 5. Briefly, the evidence against linearity is clear in the IBM series as both the SA and the AR tests indicate a non-linear effect at lags 3 and 5 (see also Figure 4(right)). As for the daily USD-YEN series the
Figure 4. (Left): time plot of the monthly log returns in percentages of IBM stock from 1960 to 1999. (Right) plot of $S_\rho$ for the IBM series at lags 1:12. Confidence bands at the 95% level (green dashed line) and at 99% level (blue dashed line) are given.

Figure 5. (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 $S_\rho$ for the series at lags 1:24. Confidence bands at the 95% level (green dashed line) and at 99% level (blue dashed line) are given.

Table 5. Significant lags at level 95% for the two series. SA indicates the SA test with the MLCV bandwidth criterion; AR indicates the AR test with the reference bandwidth criterion.

<table>
<thead>
<tr>
<th></th>
<th>SA</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>USD-YEN</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

two tests indicate different significant lags, even if from the inspection of Figure 5 (right) one may notice that $S_\rho$ at lag 1 is very close to the rejection band at 95%. Hence, we tend to support the hypothesis of nonlinearity at lag 1.
7. 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 $S_\rho$ together with surrogate resampling methods. The test pro-
posed turns out to be powerful also in those situation when several other tests may
fail. For instance, high frequency time series might show periodicitces 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 expansion up to, say,
lag 24 or more. On the contrary, our test, being based on pairwise comparisons, can
be applied at no additional costs. Moreover, $S_\rho$ is a measure of dependence that
involves the whole joint distribution function and this gives the test a potential ad-
vantage over all those test 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 phen-
omena, even for small to moderate sample sizes. Being based on resampling tech-
niques, our tests have a high computational burden. For these reasons we have built
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](http://www2.stat.unibo.it/giannerini/software.html). The library has been
implemented by writing Fortran 90 and R routines [R Development Core Team, 2010]. In particular, we made use of the R packages *tsseries, tseriesChaos, adapt*.
The code for estimating $S_\rho$ was initially written by Jeff Racine and modified by
Simone Giannerini.

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

Appendix A. Proofs

Proof of Proposition 2

Given a bivariate standard gaussian density function:

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

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

$$S_\rho = 1 - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{f_X(x)f_Y(y)f_{X,Y}(x,y)} \, dx \, dy$$

$$= 1 - \frac{1}{4\pi^2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(e^{\frac{-(x^2+y^2)(2-\rho^2)-2\rho xy}{2(1-\rho^2)}}\right)^{\frac{1}{2}} \, dx \, dy.$$
Now if we operate the polar coordinate transformation $x = r \cos \theta$ and $y = r \sin \theta$ we obtain

$$S_\rho = 1 - \frac{1}{2\pi \sqrt{1 - \rho^2}} \int_0^{2\pi} \int_0^\infty e^{-\frac{r^2 (2 - \rho^2 - \rho \sin 2\theta)}{4(1 - \rho^2)}} \, d\theta \, dr$$

$$= 1 - \frac{1}{\pi} \int_0^{2\pi} \left(1 - \rho^2\right) 4d\theta$$

$$= 1 - \frac{2 (1 - \rho^2)^{3/4}}{\sqrt{4 - 5\rho^2 + \rho^4}} = 1 - \frac{2 (1 - \rho^2)^{1/4}}{\sqrt{4 - \rho^2}}$$

**Proof of Theorem** The result follows directly from applying the delta method to the function $S_\rho(k) = g(\rho_k)$.

**References**


Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, 40126, Bologna, Italy.
E-mail address: simone.giannerini@unibo.it

Department of Economics, Emory University, Atlanta, GA, USA.
E-mail address: esfandiar.maasoumi@emory.edu

Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, 40126, Bologna, Italy.
E-mail address: estellebee@dagum.us