Entropy testing for nonlinear serial dependence in time series

Published Version:
Entropy testing for nonlinear serial dependence in time series / Simone Giannerini; Esfandiar Maasoumi; Estela Bee Dagum. - In: BIOMETRIKA. - ISSN 0006-3444. - STAMPA. - 102:3(2015), pp. 661-675. [10.1093/biomet/asv007]

Availability:
This version is available at: https://hdl.handle.net/11585/506410 since: 2015-11-04

Published:
DOI: http://doi.org/10.1093/biomet/asv007

Entropy testing for nonlinear serial dependence in time series

BY SIMONE GIANNERINI

Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, Bologna 40126, Italy

simone.giannerini@unibo.it

ESFANDIAR MAASOUMI

Department of Economics, Emory University, 1602 Fishburne Dr., Atlanta, Georgia 30322, U.S.A.

esfandiar.maasoumi@emory.edu

AND ESTELA BEE DAGUM

Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, Bologna 40126, Italy

estela.beedagum@unibo.it

SUMMARY

We propose tests for nonlinear serial dependence in time series under the null hypothesis of general linear dependence, in contrast to the more widely studied null hypothesis of independence. The approach is based on combining an entropy dependence metric, which possesses many desirable properties and is used as a test statistic, with a suitable extension of surrogate data methods, a class of Monte Carlo distribution-free tests for nonlinearity, and a smoothed sieve bootstrap scheme. We show how, in the same way as the autocorrelation function is used for linear models, our tests can in principle be employed to detect the lags at which a significant nonlinear relationship is present. We prove the asymptotic validity of the proposed procedures and the corresponding inferences. The small-sample performance of the tests in terms of power and size is assessed through a simulation study. Applications to real datasets of different kinds are also presented.

Some key words: Entropy; Nonlinearity; Smoothed sieve bootstrap; Surrogate data; Test; Time series.

1. INTRODUCTION

The literature on tests for nonlinear serial dependence in time series is extensive, but the establishment of a unified mathematical framework that encompasses all aspects of nonlinearity has proven elusive. Even though departures from linearity can occur in many directions, testing for nonlinearity is often a test for a specific nonlinear feature or form, making it difficult to compare existing proposals. Nonlinear features have arisen in many different areas; for instance, some concepts from nonlinear dynamics and chaos theory, such as initial value sensitivity, fractal dimensions and nonuniform noise amplification, have motivated the introduction of new tools and tests. In other situations, nonlinearity is inferred from the failure of a linear model. Thus, the problem of assessing the nonlinear character of a series reduces to a diagnostic test, usually performed on the residuals of a linear model, or a specification test between models. For a recent review, see Giannerini (2012).

Almost all tests for nonlinearity are based on specific moments or features of the distribution of the process, and focus on the null hypotheses of linearity, or of no dependence. The latter is a rather big straw man unless the process has been filtered. Furthermore, many such tests are designed to work with a restricted class of models. Since the true model is never known, the reported performance of such tests may not reflect the real performance, which depends on the degree of modelling misspecification.

In this paper we address the above-mentioned issues by introducing a general purpose test for nonlinear serial dependence based on the whole pairwise distribution of the process through its entropy. Previously, this type of test had been advocated for the null hypothesis of independence, for which it is easier to derive the asymptotic and resampling distributions. While our test is diagnostic, it is designed to identify different aspects of nonlinearity. Furthermore, it does not require the specification of a specific model and, in principle, can help to identify the lags at which a nonlinear relationship is expected, similarly to the autocorrelation function for linear models.

Our null hypotheses are consistent 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,

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

where $\sum_{j=1}^{\infty} \phi_j^2 < \infty$ and $E(X_t^4) < \infty$. The second null hypothesis is that $\{X_t\}$ is a zero-mean linear stationary process,

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

where the error process $\{\varepsilon_t\}$ has mean zero and variance σ_ε^2. The alternative hypothesis H_1 states that $\{X_t\}$ does not admit an infinite autoregressive representation as in (1) or (2). As discussed in Tong (1990, p. 202), any stationary process with a continuous spectrum admits a linear two-sided moving average representation with uncorrelated error terms. The one-sided moving average representation requires additional integrability conditions on the spectral density function of the process. In turn, under the assumption of invertibility of the moving average terms, we obtain the one-sided autoregressive representation adopted here. This narrows the range of processes implied by the moving average representation only slightly. Indeed, even though different authors implicitly define linear processes as being infinite autoregressive (see, e.g., Hjellvik & Tjøstheim, 1995), the closure of the class of linear processes that satisfy Wold's representation theorem is surprisingly broad and can include also nonergodic Poisson sum processes (Bickel & Bühlmann, 1997). Now, assume we are given a time series $\mathrm{x} = (x_1, \ldots, x_n)$ and we would like to test whether x might be operationally considered as a realization of the process (1) or (2). The test statistics we propose are functionals of a metric-entropy measure of dependence for time series. This measure possesses many desirable properties and has been shown to be powerful in other settings (see, e.g., Granger et al., 2004; Maasoumi & Racine, 2009). We will show that under the null hypothesis (1), the entropy measure reduces to a nonlinear function of the correlation coefficient. Hence, we construct a test statistic from the quadratic divergence between the parametric estimator of the entropy measure under H_0 and the corresponding unrestricted nonparametric estimator. The same metric-entropy statistic, estimated nonparametrically, is used to test the null hypothesis of generic linearity, i.e., H_0'. We derive the asymptotic distributions of the test statistics under H_0 and H_0'. Typically, these approximations depend on unknown quantities and rely upon the specification of a model, a shortcoming that we explicitly wish to avoid; also, they require large samples in order to be valid. To overcome these issues, we propose two resampling schemes and prove the asymptotic validity of the proposed procedures and the corresponding inferences. The first scheme is based on surrogate data methods, while the second uses the smoothed sieve bootstrap.

2. A NONLINEAR AUTOCORRELATION FUNCTION

2·1. *Introduction and definition*

There are many proposed measures of dependence, which were motivated by different needs and designed to characterize specific aspects of the process under study. An important class of such measures is based on entropy functionals (see, e.g., Joe, 1989; Maasoumi, 1993). For instance, Shannon mutual information and the Kullback–Leibler divergence became popular in nonlinear dynamics. Such measures have also been used in time series analysis (Robinson, 1991; Granger & Lin, 1994; Tjøstheim, 1996; Hong & White, 2005). However, most of these entropies are not metrics, because they either do not obey the triangle inequality or are not commutative operators. While these shortcomings may not seem immediately consequential for most tests, they have been shown to have an impact on their performance; moreover, they affect our ability to assess and quantify degrees of dependence or departures from points of interest, or to search for minimum-distance/optimal solutions or models (Granger et al., 2004). The measure we discuss here is the metric entropy S_ρ, a normalized version of the Bhattacharya–Hellinger–Matusita distance:

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

where $f_{X_t}(\cdot)$ and $f_{(X_t,X_{t+k})}(\cdot\,,\cdot)$ denote the probability density functions of X_t and the vector (X_t, X_{t+k}), respectively. This measure is a particular member of the family of symmetrized general relative entropies, which includes as a special case the nonmetric relative entropies often referred to as Shannon information or Kullback–Leibler divergence. The measure $S_\rho(k)$ can be interpreted as a nonlinear autocorrelation function and possesses many desirable properties. In particular, it is a metric and is defined for both continuous and discrete variables; it is normalized and takes the values zero if X_t and X_{t+k} are independent and unity if there is a measurable exact relationship between continuous variables; it reduces to a function of the linear correlation coefficient in the case of jointly Gaussian variables; and 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 regarding measures of dependence given by Rényi (1959) and further studied in Maasoumi (1993) and Granger et al. (2004); see the Supplementary Material. A key result from the perspective of testing for nonlinearity concerns the relationship with the correlation coefficient in the Gaussian case; the following correction to Granger et al. (2004) is in order.

PROPOSITION 1. *Let $(X_t, X_{t+k}) \sim N(0, 1, \rho)$ be a standard normal random vector with joint probability density function $f_{(X_t,X_{t+k})}(\cdot\,,\cdot\,,\rho)$ where ρ is the correlation coefficient at lag k.*

Then

$$S_\rho(k) = 1 - \frac{2\left(1-\rho^2\right)^{1/4}}{(4-\rho^2)^{1/2}}. \quad (3)$$

For the sake of brevity, below we write S_k in place of $S_\rho(k)$.

2·2. *The parametric estimator under H_0*

Equation (3) allows us to obtain an estimator for S_k based on the sample autocorrelation $\hat\rho_k$ under the null hypothesis (1) of a linear Gaussian process. We denote such a parametric estimator by $\hat S_k^{\mathrm{p}}$, where the superscript stands for parametric. In the next two results we derive the asymptotic distribution of $\hat S_k^{\mathrm{p}}$ and prove its consistency. To this end, define the function $g:[-1,1]\to[0,1]$ by $g(x) = 1 - 2(1-x^2)^{1/4}(4-x^2)^{-1/2}$. The function g is differentiable on $(0,1)$ and its ith derivative $g^{(i)}(x)$ is not equal to zero for $x \neq 0$.

PROPOSITION 2. *Let $\{X_t\}$ be the zero-mean stationary process under H_0 as in (1). Also, let $\hat\rho_k$ be the sample autocorrelation function of $\{X_t\}$ at lag k, and let $\hat S_k^{\mathrm{p}} = 1 - 2(1-\hat\rho_k^2)^{1/4}(4-\hat\rho_k^2)^{-1/2}$ be the corresponding sample estimator of S_ρ at lag k based on (3). Then, for every $k = 0, 1, \ldots$ we have $n^{1/2}(\hat S_k^{\mathrm{p}} - S_k) \to N(0, \sigma_{\mathrm{p}}^2)$ in distribution, where $\sigma_{\mathrm{p}}^2 = [g'(\zeta)]^2$, with $\zeta = \sum_{i=1}^{\infty}\{\rho_{(i+k)}\rho_{(i-k)} - 2\rho_i\rho_k\}^2$ being the asymptotic variance of $\hat\rho_k$ (Brockwell & Davis, 2009, pp. 221–2).*

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

PROPOSITION 3. *Under the hypotheses of Proposition 2, $\hat S_k^{\mathrm{p}} \to S_k$ in L^2 as $n \to \infty$.*

2·3. *Unrestricted nonparametric estimator*

Nonparametric estimation of S_k and related entropy measures, under conditions that allow us to construct tests for the null hypothesis of serial independence, has been studied by Robinson (1991), Skaug & Tjøstheim (1996), Tjøstheim (1996), Granger et al. (2004), Hong & White (2005) and Fernandes & Néri (2010). Here we adapt the relevant theory to test the null hypothesis (2) of linear serial dependence and derive the asymptotic distribution of the nonparametric estimator for S_k:

$$\hat S_k^{\mathrm{u}} = \frac{1}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\left\{\hat f_{(X_t,X_{t+k})}(x_1,x_2)\right\}^{1/2} - \left\{\hat f_{X_t}(x_1)\hat f_{X_{t+k}}(x_2)\right\}^{1/2}\right]^2 w(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

We use kernel density estimators for f_{X_t}, $f_{X_{t+k}}$ and $f_{(X_t,X_{t+k})}$, namely

$$\hat f_{X_t}(x) = n^{-1}\sum_{i=1}^{n} h_1^{-1}K\left\{(x-X_i)/h_1\right\},$$

$$\hat f_{(X_t,X_{t+k})}(x_1,x_2) = (n-k)^{-1}\sum_{i=1}^{n-k}(h_1h_2)^{-1}K\left\{(x_1-X_i)/h_1, (x_2-X_{i+k})/h_2\right\},$$

where K is a kernel function and h_1 and h_2 are bandwidths. In the expression for $\hat{S}^{\mathrm{u}}_k$, $w(x_1, x_2)$ is a continuous weight function that is needed both to exclude outlying observations and to facilitate the asymptotic analysis. We assume the following regularity conditions.

Condition 1. The process $\{X_t\}$ is strictly stationary and β-mixing with exponentially decaying coefficients.

Condition 2. The densities f_{X_t}, $f_{X_{t+k}}$ and $f_{(X_t, X_{t+k})}$ are continuously differentiable up to order s, and their derivatives are bounded and square-integrable. Also, the joint density function of $(Z_{k_1}, \ldots, Z_{k_\varsigma})$, where $Z_t = (X_t, X_{t+k})$, is Lipschitz-continuous, i.e., $|f(Z_{k_1} + \delta, \ldots, Z_{k_\varsigma} + \delta) - f(Z_{k_1}, \ldots, Z_{k_\varsigma})| \leqslant \mathcal{D}(Z_{k_1}, \ldots, Z_{k_\varsigma})\, \|\delta\|$, where $\mathcal{D}$ is an integrable function and $1 \leqslant \varsigma \leqslant 4$.

Condition 3. The kernel function $K(u)$ is continuous, differentiable up to order s, and such that $\int uK(u)\,\mathrm{d}u = 0$, $\int u^2 K(u)\,\mathrm{d}u < \infty$, $K(x) = \int \bar{K}(u)\exp(\mathrm{i}\eta x)\,\mathrm{d}\eta$, $e_K = \int K^2(u)\,\mathrm{d}u$ and $v_K = \int \{\int K(u)K(u+v)\,\mathrm{d}u\}^2\,\mathrm{d}v$, where $\mathrm{i}^2 = -1$ and $\bar{K}(u)$ is a real function satisfying $\int |\bar{K}(u)|\,\mathrm{d}\eta < \infty$.

Condition 4. The bandwidths $h_1 = h_1(n, X_t)$ and $h_2 = h_2(n, X_{t+k})$ satisfy $h_i \to 0$ and $nh_i \to \infty$ as $n \to \infty$. Also, $h_i = o(n^{-1/(2s+1)})$ for $i = 1, 2$.

Condition 5. The weight function $w(x_1, x_2) = \mathbb{I}\{(x_1, x_2) \in D\}$, where $\mathbb{I}$ denotes the indicator function, is nonnegative and separable, i.e., $w(x_1, x_2) = w(x_1)w(x_2)$, for $D = D_1 \times D_1$ with D_1 a closed real interval.

These conditions lead us to the following result.

Proposition 4. *Under Conditions* 1–5*, as* $n \to \infty$, $\hat{S}^{\mathrm{u}}_k \to S_k$ *in* L^2 *and* $n^{1/2}(\hat{S}^{\mathrm{u}}_k - S_k) \to N(0, \sigma^2_{\mathrm{u}})$ *in distribution, where* σ^2_{u} *is the asymptotic variance that depends on* $w(x_1, x_2)$.

Conditions 1–5 can be relaxed to some extent without affecting the results. For instance, one could assume α-mixing processes and less restrictive conditions 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 bandwidth plays a crucial role. In this paper we investigate two methods for selecting the bandwidth. The first method is maximum likelihood crossvalidation: we choose the bandwidth h that maximizes the score function $\textsc{cv}(h) = n^{-1}\sum_{i=1}^n \log \hat{f}_{-i}(X_i)$, 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 take either $h = 1{\cdot}06\,\hat{\sigma}\,n^{-1/5}$ in the univariate case, or $h_i = 1{\cdot}06\,\hat{\sigma}_i\,n^{-1/6}$ for $i = 1, 2$ in the bivariate case. For further details on these two methods, see Silverman (1986).

The implementation of $\hat{S}^{\mathrm{u}}_k$ requires the computation of a double integral, for which adaptive quadrature methods have been employed. Details of the software implementation 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), this can lead to degradation in the performance of the tests.

3. THE TEST STATISTICS

To test the null hypotheses of linearity, H_0 and H_0', we propose the following test statistics:

$$\hat{T}_k = \left(\hat{S}_k^{\mathrm{u}} - \hat{S}_k^{\mathrm{p}}\right)^2 \text{ for } H_0, \quad \hat{S}_k^{\mathrm{u}} \text{ for } H_0'.$$

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

THEOREM 1. *Under H_0 and the assumptions of Propositions 2 and 4, $\hat{T}_k \to 0$ in L^2 as $n \to \infty$. Moreover, $(\sigma_{\mathrm{a}}^2)^{-1} n \hat{T}_k \to \chi_1^2$ in distribution, where σ_{a}^2 is the asymptotic variance of $\hat{T}_k^{1/2}$.*

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 H_0. The derivation of the asymptotic approximation for the significance level and power of the test depends on the estimator of the asymptotic variance σ_{a}^2, which in turn depends on σ_{u}^2 and σ_{p}^2. Such approximation is feasible only for a few specific cases and is of little practical relevance since it requires the specification of a model. Furthermore, preliminary investigations show that very large sample sizes are required to obtain meaningful results. The same problems have been reported previously (see Hjellvik & Tjøstheim, 1995; Tjøstheim, 1996; Hjellvik et al., 1998; Hong & White, 2005).

As regards the general null hypothesis of linearity H_0', we propose using the nonparametric estimator $\hat{S}_k^{\mathrm{u}}$. Proposition 4 ensures that under mild conditions, which include the class of linear processes defined by H_0', the statistic $\hat{S}_k^{\mathrm{u}}$ is consistent for S_k and asymptotically Gaussian. In this case also, the issues relating to the asymptotic approximations remain, so we study two resampling schemes which, when used together with our test statistics, lead to valid inferences and deliver good performance for 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 for testing the null hypothesis of generic linearity, H_0'.

4. SURROGATE DATA APPROACH

The method of surrogate data, introduced in the context of nonlinear time series analysis, motivated by chaos theory, can be regarded as a resampling approach to building tests for nonlinearity in the absence of distribution theory. Although the use of tests based on simulations was common long before 1990, in the literature on nonlinear dynamics Theiler et al. (1992) is usually viewed as the seminal paper on the subject. The main idea can be summarized as follows: a null hypothesis regarding the process that has generated the observed series is formulated, e.g., H_0: the generating process is linear and Gaussian; a set of resampled series consistent with H_0, called surrogate series, is obtained through Monte Carlo methods; then, a suitable test statistic known to have discriminatory power against H_0 is computed on the surrogates, yielding the distribution of the test statistic under H_0, from which the significance level and p-values can be derived.

In Theiler et al. (1992), a null hypothesis of linearity is tested by generating surrogates having the same periodogram and same marginal distribution as the original series. It is assumed that the generating process is a linear Gaussian process as in (1) and that the process admits a spectral density function that forms a Fourier pair with the autocovariance function. Given an observed series $\mathrm{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$, we can define its discrete Fourier transform $\zeta_{\mathrm{x}}(\omega) = (2\pi n)^{-1/2} \sum_{t=1}^{n} x_t \exp(-\mathrm{i}\omega t)$ $(-\pi \leqslant \omega \leqslant \pi)$ and sample periodogram $I(\mathrm{x}, \omega) = |\zeta_{\mathrm{x}}(\omega)|^2$. In general, it can be shown that $\zeta_{\mathrm{x}}(\omega) = (2\pi)^{-1/2} P_n \mathrm{x}$, where P_n is an orthonormal matrix. Hence, assuming n is odd, the series x can be uniquely recovered from the sample mean, the periodogram values $I(\mathrm{x}, \omega_j)$ $(j = 1, \ldots, (n-1)/2)$ and the phases $\theta_1, \ldots, \theta_{(n-1)/2}$ through the formula $x_t = \bar{x} + (2\pi/n)^{1/2} \sum_{j=1}^{(n-1)/2} 2I(\mathrm{x}, \omega_j)^{1/2} \cos(\omega_t j + \theta_j)$. This allows one to obtain a surrogate series $\mathrm{x}^* = (x_1^*, \ldots, x_n^*)^{\mathrm{T}}$ by randomizing the phases as follows:

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

where $\theta_1, \ldots, \theta_m$ are independent and identically distributed as $\mathrm{Un}[0, 2\pi]$. The surrogate series will have the same sample mean and periodogram as the original series. Chan (1997) 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 on the method are similar, i.e., they have a Neyman structure. Chan also proved the asymptotic validity of the tests for the null hypothesis of a stationary Gaussian process with fast-decaying autocorrelations (Chan & Tong, 2001, § 4.4). With the exception of Chan (1997), and despite the large literature on surrogate data methods, to our knowledge comprehensive studies on the theoretical properties of such tests are still lacking.

The approach we propose in this paper is an extension of the scheme that fits within the unified framework of an optimization problem solved by means of simulated annealing (Schreiber, 1998). The procedure can be summarized as follows: (i) define one or more constraints in terms of a cost function C, which reaches a global minimum when the constraints are fulfilled; (ii) minimize the cost function C among all 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 as the original series. In the following proposition we show that under H_0, the surrogate approach combined with our test statistics yields 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 $\hat{S}_k^{\mathrm{u}}$, leads to asymptotically valid inferences in that the associated p-value follows a uniform distribution on* (0, 1).

The procedure and implementation are described in the Supplementary Material.

5. The bootstrap approach

The second approach we consider is a smoothed version of the sieve bootstrap. The sieve bootstrap relies on the Wold decomposition of a stationary process. In fact, under mild assumptions, a real-valued purely nondeterministic stationary process admits a one-sided infinite-order autoregressive representation. The sieve approximates a possibly infinite-dimensional model through a sequence of finite-dimensional autoregressive models. The nonsmoothed version of this approach has been investigated in a number of studies (see, e.g., Kreiss & Franke, 1992; Bühlmann, 1997, 2002). In particular, Bühlmann (1997) shows that the scheme leads to valid inferences for smooth functions of linear statistics. Since our test statistics have components based on kernel density estimators, we use the smoothed sieve bootstrap proposed in Bickel & Bühlmann (1999). Such a 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 situation in the following way: (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 in step (ii). The full implementation of the scheme and further details are provided in the Supplementary Material.

Bühlmann (1997) showed that if AIC is used for model selection, 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. The performance of the method is quite insensitive to the choice of criterion used for model selection, as long as the order chosen is reasonable. In the following proposition, we prove the validity of inference based on combining our test statistics with the smoothed sieve bootstrap scheme.

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

(i) *under* H_0, $\sup_x |\mathrm{pr}^*\{n^{1/2}(\hat{T}_k^* - T_k^*) \leqslant x\} - \mathrm{pr}\{n^{1/2}(\hat{T}_k - T_k \leqslant x)\}| = o_{\mathrm{p}}(1)$ *as* $n \to \infty$;

(ii) *under* H_0', $\sup_x |\mathrm{pr}^*\{n^{1/2}(\hat{S}_k^{\mathrm{u}*} - S_k^*) \leqslant x\} - \mathrm{pr}\{n^{1/2}(\hat{S}_k^{\mathrm{u}} - S_k \leqslant x)\}| = o_{\mathrm{p}}(1)$ *as* $n \to \infty$.

6. FINITE-SAMPLE PERFORMANCE: A SIMULATION STUDY

In this section we assess by simulation the performance of the tests in finite samples. The 24 models used are listed in Table 1, where the innovation processes are independent and identically distributed with $\varepsilon_t \sim N(0, 1)$ and ζ_t following Student's t with three degrees of freedom.

Models 1–6 are linear Gaussian processes, so the rejection percentages give an indication of the sizes of the tests. Models 7–12 are the same processes but with Student's t innovations, so the tests based on T should reject the null while those based on S should not. Models 13–24 are nonlinear processes that do not admit an infinite-order linear autoregressive representation. In particular, Models 13 and 14 are bilinear processes, Models 15 and 16 are nonlinear moving average processes, Models 17 and 18 are generalized autoregressive conditional heteroscedastic processes, Models 19–21 are threshold autoregressive processes, and Models 22 and 23 are exponential autoregressive processes; finally, Model 24 is the logistic map at a chaotic regime. The parameters of Models 13 and 14 are taken from Hjellvik et al. (1998), those of Models 17, 21 and 24 from Rusticelli et al. (2009), and those of Models 19, 22 and 23 from Tsay (2000), so that comparisons are possible.

The null hypothesis of linearity and Gaussianity in (1) is tested by means of $\hat{T}_k$ coupled with the surrogate approach and the crossvalidation criterion, whereas the general null hypothesis of linearity in (2) is tested through $\hat{S}_k^{\mathrm{u}}$ coupled with the bootstrap scheme and the reference criterion. In all the experiments the number of surrogates or bootstrap replicates is set to $B = 999$. The results are given in terms of rejection percentages of the tests at $\alpha = 0{\cdot}05$ over 1000 Monte Carlo replications. We have chosen $n = 50$, 100 and 200. In analogy with tests based on autocorrelations and those proposed in Hjellvik & Tjøstheim (1995) and Hjellvik et al. (1998), our procedures depend on the choice of the lag k. This means that the null depends on k, so we adopt the combination function $H_0^{k_{\max}} = \bigcap_{k=1}^{k_{\max}} H_0^k$; see also Fernandes & Néri (2010). In other words, the null of linearity is rejected if the test rejects for at least one in $k_{\max}$ lags, where $k_{\max} = 5$. Since this approach mirrors what is usually done with correlograms in time series analysis, and because the results below confirm that for $\hat{S}_k^{\mathrm{u}}$ this approach can indeed be followed, we have chosen to retain it and report the results here. However, a more rigorous approach would require a correction for multiple testing; see the Supplementary Material. Measuring the performance

Table 1. *Time series models used in the simulation study*

Model		Model	
1:	$x_t = 0{\cdot}8\, x_{t-1} + \varepsilon_t$	13:	$x_t = 0{\cdot}7\, \varepsilon_{t-1}\, x_{t-2} + \varepsilon_t$
2:	$x_t = -0{\cdot}8\, x_{t-1} + \varepsilon_t$	14:	$x_t = 0{\cdot}5 - 0{\cdot}4\, x_{t-1} + 0{\cdot}4\, \varepsilon_{t-1}\, x_{t-1} + \varepsilon_t$
3:	$x_t = 0{\cdot}8\, \varepsilon_{t-1} + \varepsilon_t$	15:	$x_t = 0{\cdot}8\, \varepsilon_{t-2}^2 + \varepsilon_t$
4:	$x_t = -0{\cdot}8\, \varepsilon_{t-1} + \varepsilon_t$	16:	$x_t = 0{\cdot}5\, \varepsilon_{t-2}^2 + \varepsilon_t$
5:	$x_t = 0{\cdot}6\, x_{t-1} + 0{\cdot}4\, \varepsilon_{t-1} + \varepsilon_t$	17:	$x_t = \sigma_t \varepsilon_t$ $\sigma_t^2 = 0{\cdot}0108 + 0{\cdot}8516\, \sigma_{t-1}^2 + 0{\cdot}1244\, x_{t-1}^2$
6:	$x_t = -0{\cdot}6\, x_{t-1} - 0{\cdot}4\, \varepsilon_{t-1} + \varepsilon_t$	18:	$x_t = \sigma_t \varepsilon_t$ $\sigma_t^2 = 0{\cdot}1 + 0{\cdot}6\, x_{t-1}^2$
7:	$x_t = 0{\cdot}8\, x_{t-1} + \zeta_t$	19:	$x_t = \begin{cases} 1 - 0{\cdot}5\, x_{t-1} + \varepsilon_t & \text{if } x_{t-1} \leqslant 0 \\ -1 - 0{\cdot}5\, x_{t-1} + \varepsilon_t & \text{if } x_{t-1} > 0 \end{cases}$
8:	$x_t = -0{\cdot}8\, x_{t-1} + \zeta_t$	20:	$x_t = \begin{cases} 0{\cdot}8\, x_{t-1} + \varepsilon_t & \text{if } x_{t-1} \leqslant -1 \\ -0{\cdot}8\, x_{t-1} + \varepsilon_t & \text{if } x_{t-1} > -1 \end{cases}$
9:	$x_t = 0{\cdot}8\, \zeta_{t-1} + \zeta_t$	21:	$x_t = \begin{cases} -0{\cdot}5\, x_{t-1} + \varepsilon_t & \text{if } x_{t-1} \leqslant 1 \\ 0{\cdot}4\, x_{t-1} + \varepsilon_t & \text{if } x_{t-1} > 1 \end{cases}$
10:	$x_t = -0{\cdot}8\, \zeta_{t-1} + \zeta_t$	22:	$x_t = 0{\cdot}3 + 10\, \exp\{-x_{t-1}^2\}\, x_{t-1} + \varepsilon_t$
11:	$x_t = 0{\cdot}6\, x_{t-1} + 0{\cdot}4\, \zeta_{t-1} + \zeta_t$	23:	$x_t = 0{\cdot}3 + 100\, \exp\{-x_{t-1}^2\}\, x_{t-1} + \varepsilon_t$
12:	$x_t = -0{\cdot}6\, x_{t-1} - 0{\cdot}4\, \zeta_{t-1} + \zeta_t$	24:	$x_t = 4\, x_t\, (1 - x_t)$

of the tests for $\alpha \leqslant 0{\cdot}01$ may require either more bootstrap replicates or adaptive nonparametric density estimators (Silverman, 1986).

Table 2 shows the results; the left columns refer to $\hat{T}_k$ and the null H_0 of a linear Gaussian process, and the right columns refer to $\hat{S}_k^{\mathrm{u}}$ and the null H_0' of a generic linear process. The standard error of the Monte Carlo estimates is at most of the order of 0·7% for Models 1–12 and 1·7% for Models 13–24. The rejection percentages show high power in almost every situation, even for $n = 50$. Compared with other proposals (Tsay, 2000; Rusticelli et al., 2009), our tests are almost invariably an improvement. The test based on surrogates and $\hat{T}_k$ is oversized and needs at least 200 observations to perform sensibly and also detect linear non-Gaussian processes. We managed to reduce this problem to some extent by fine-tuning the cost function of the annealing algorithm. Our conjecture is that the different convergence rates of the estimators of the two components of T_k play a role. The test based on $\hat{S}_k^{\mathrm{u}}$ coupled with the smoothed sieve bootstrap has small size even for short series when it comes to linear Gaussian processes. In some instances of linear non-Gaussian processes, the test tends to over-reject the null, especially for models 7, 9 and 11, which are characterized by positive parameters. As mentioned above, this is due to the multiple testing approach, and the test does not show any over-rejection once the significance level has been corrected; see the Supplementary Material. The test based on $\hat{S}_k^{\mathrm{u}}$ appears to be rather conservative and would lead to sensible decisions without any correction. The results shown are fairly robust with respect to the parameter values of the 24 processes. The case of threshold nonlinearity is partly an exception, since the results are more variable for different parameter settings. As pointed out by a referee, this could be due to the discontinuity of the autoregression at the threshold, so that its value would seem to exert an influence over the performance of the tests. Investigations involving smooth threshold processes might shed further light on this issue.

As also reported in Hjellvik et al. (1998), the choice of bandwidth plays an important role in the performance of the tests. Overall, our experiments indicate that the reference criterion should be paired with the bootstrap scheme, while the crossvalidation criterion should be preferred when

Table 2. *Rejection percentages at the nominal level $\alpha = 5\%$: the left columns refer to $\hat{T}_k$ and the null H_0 of a linear Gaussian process; the right columns refer to $\hat{S}_k^{\mathrm{u}}$ and the null H_0' of a generic linear process*

		$\hat{T}_k$				$\hat{S}_k^{\mathrm{u}}$		
	n	50	100	200	n	50	100	200
	Model 1	74·8	14·0	7·8	Model 1	10·4	4·4	0·8
	Model 2	70·4	11·6	6·8	Model 2	2·2	1·0	0·2
	Model 3	60·2	17·0	8·2	Model 3	1·8	2·2	4·4
Linear	Model 4	56·8	15·0	8·4	Model 4	1·6	1·8	2·2
Gaussian	Model 5	76·6	12·2	5·2	Model 5	5·2	1·8	2·2
	Model 6	81·3	11·6	6·7	Model 6	1·6	0·8	1·4
	Model 7	86·3	35·0	52·6	Model 7	19·2	11·0	6·6
	Model 8	83·4	42·3	58·4	Model 8	6·8	1·8	2·2
Linear	Model 9	80·4	32·9	45·2	Model 9	11·4	12·0	15·0
non-Gaussian	Model 10	78·7	44·3	57·9	Model 10	6·6	8·6	10·2
	Model 11	86·3	34·6	48·8	Model 11	10·6	10·6	9·4
	Model 12	80·9	36·1	46·3	Model 12	5·0	3·6	4·0
	Model 13	68·8	57·6	83·0	Model 13	36·0	62·6	87·6
	Model 14	80·4	70·8	94·6	Model 14	35·2	57·2	89·8
	Model 15	81·3	83·2	96·6	Model 15	52·4	88·0	99·0
	Model 16	86·3	73·8	97·2	Model 16	51·2	85·4	98·6
	Model 17	83·4	33·6	51·0	Model 17	18·6	37·4	62·8
Nonlinear	Model 18	84·4	66·6	86·4	Model 18	44·6	70·6	92·6
	Model 19	78·7	83·0	98·0	Model 19	29·0	51·0	91·2
	Model 20	86·3	68·0	81·4	Model 20	29·8	47·0	73·2
	Model 21	89·9	67·0	96·0	Model 21	31·2	69·2	96·2
	Model 22	98·8	100·0	100·0	Model 22	84·2	100·0	99·6
	Model 23	90·5	99·2	100·0	Model 23	75·6	99·6	100·0
	Model 24	97·4	100·0	100·0	Model 24	100·0	100·0	100·0

using the surrogate-based test. In particular, the reference bandwidth leads to severe oversize when used with surrogate data, whereas the bootstrap approach has low power when paired with likelihood crossvalidation, perhaps due to the residual-based nature of the sieve bootstrap. The computational complexity of the bootstrap-reference implementation is linear with respect to the sample size n. The surrogate-crossvalidation implementation has a complexity which is quadratic with respect to n.

7. REAL-DATA APPLICATION

The two series analysed here are described in detail in Tsay (2010), and were taken from the companion R package FinTS (Graves, 2014; R Development Core Team, 2015). In both cases we have applied the surrogate test with the crossvalidated bandwidth criterion and 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, consisting of $n = 468$ observations. The series has white noise type ACF and PACF. The data are shown in Fig. 1(a), while the plot of $\hat{T}_k$ at lags 1 to 12 is shown in Fig. 1(b).

The second series consists of daily exchange rates between the U.S. dollar and Japanese yen from 3 January 2000 to 26 March 2004. The series has $n = 1063$ observations and has been differenced and log-transformed. Such a series has white noise type correlogram, while the partial

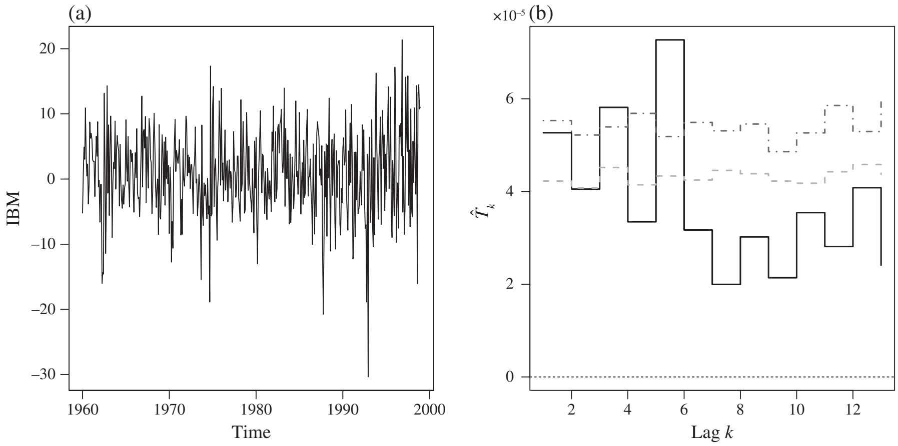

Fig. 1. (a) Monthly log returns in percentages of IBM stock from 1960 to 1999. (b) Plot of $\hat{T}_k$ for the IBM series at lags 1 to 12 with rejection bands at 95% (dashed) and 99% (dash-dot).

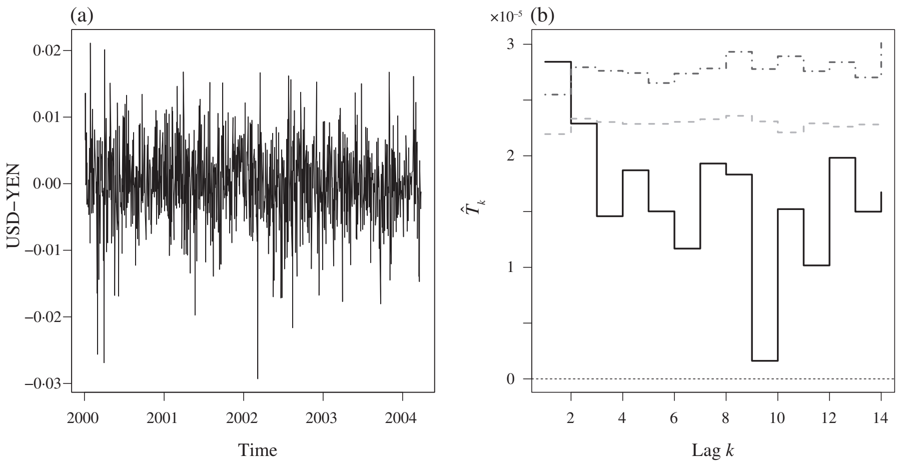

Fig. 2. (a) Differenced and log-transformed daily exchange rates between the U.S. dollar and Japanese yen from 3 January 2000 to 26 March 2004. (b) Plot of $\hat{T}_k$ for the exchange rate series at lags 1 to 14 with rejection bands at 95% (dashed) and 99% (dash-dot).

correlogram is significant at lag 1. The data are displayed in Fig. 2(a), and the plot of $\hat{T}_k$ at lags 1 to 14 is shown in Fig. 2(b).

The evidence against linearity is clear in both series, as the two tests give the same outcome in each case. In particular, for the IBM data there are possible nonlinear effects at lags 3 and 5; see Fig. 1(b). For the daily USD-YEN exchange rate, the tests suggest a significant effect at lag 1. If we compare 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 infeasible to perform a model identification solely on the basis of such plots, the information conveyed by our test can help considerably. In this instance, the results point to a complex dependence upon past shocks that is consistent with findings reported in the literature.

8. CONCLUSIONS

Our tests, being based on pairwise comparisons, can be applied in situations where other tests may fail. For instance, high-frequency time series may show periodicities at distant lags due to the sampling rate. In such cases it would be infeasible to apply tests that require the specification of a nonlinear model or a Volterra series expansion involving many lags. Moreover, S_ρ is a measure of dependence that involves the whole bivariate distribution function, and this offers a potential advantage over tests based on specific moments or aspects of the distributions. Our procedures have a high computational burden, so we have created an R package that implements a parallel version of them. The package can be found at www2.stat.unibo.it/giannerini/software.html and is forthcoming on CRAN.

ACKNOWLEDGEMENT

The authors thank the editor and the referees for valuable suggestions, and Qiwei Yao and Francesco Battaglia for helpful comments on an earlier version of the manuscript. This work was supported by the Italian Ministry of University and Research. Simone Giannerini acknowledges the support of the Institute for Mathematical Sciences of the National University of Singapore.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online contains the proof of Proposition 1, further results from the simulation study, and further discussions.

APPENDIX

Proof of Proposition 2

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

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 $g : [-1, 1] \to [0, 1]$, $g(x) = 1 - 2(1-x^2)^{1/4}(4-x^2)^{-1/2}$ is a continuous bounded function, from Theorem 5.1.4 of Fuller (1996) it follows that $\hat{S}_k^{\mathrm{p}} = g(\hat{\rho}_k) \to g(\rho_k) = S_k$ in probability. Furthermore, since $0 < \hat{S}_k < 1$ almost surely for all k, from Theorem 6.2.4 of Sen et al. (2009) it follows that $\hat{S}_k^{\mathrm{p}} \to S_k$ in L^2.

Proof of Proposition 4

(i) Conditions 1–5 enable us to apply the framework of Tjøstheim (1996). The quantity to be estimated can also be written as

$$S_k = 1 - \iint B\left\{u(x_1, x_2)\right\} w(x_1, x_2)\, \mathrm{d}F(x_1, x_2)$$

where

$$u(x_1, x_2) = \left\{ f_{X_t}, (x_1) f_{X_{t+k}}(x_2),\ f_{X_t, X_{t+k}}(x_1, x_2) \right\},$$

$$B\left\{u(x_1, x_2)\right\} = \left\{ f_{X_t}(x_1) f_{X_{t+k}}(x_2) \right\}^{1/2} \left\{ f_{X_t, X_{t+k}}(x_1, x_2) \right\}^{-1/2},$$

so that

$$\hat{S}_k^{\mathrm{u}} = 1 - \iint B\left\{\hat{u}(x_1, x_2)\right\} w(x_1, x_2)\, \mathrm{d}\hat{F}(x_1, x_2).$$

Now we have

$$\hat{S}_k^{\mathrm{u}} - S_k = \iint B\left\{u(x_1, x_2)\right\} w(x_1, x_2) \left\{\mathrm{d}F(x_1, x_2) - \mathrm{d}\hat{F}(x_1, x_2)\right\} \tag{A1}$$

$$\iint \left[B\left\{u(x_1, x_2)\right\} - B\left\{\hat{u}(x_1, x_2)\right\}\right] w(x_1, x_2)\, \mathrm{d}\hat{F}(x_1, x_2). \tag{A2}$$

By the ergodic theorem, (A1) $\to 0$ in L^2 as $n \to \infty$. To prove that (A2) $\to 0$ in L^2, note that there exists an integer N such that for $n \geqslant N$ we have $K_n = \mathrm{pr}\{u(x_1, x_2) \in A\} = 1$, 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

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

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

(ii) The regularity assumptions, Conditions 1–5, allow us to apply the theoretical framework outlined in Tjøstheim (1996). The proof then follows from that of Tjøstheim (1996) by taking $u(x_1, x_2) = \{f_{X_t}, (x_1) f_{X_{t+k}}(x_2), f_{X_t, X_{t+k}}(x_1, x_2)\}$ and $B\{u(x_1, x_2)\} = \{f_{X_t}(x_1) f_{X_{t+k}}(x_2)\}^{1/2} \times \{f_{X_t, X_{t+k}}(x_1, x_2)\}^{-1/2}$.

Proof of Theorem 1

(i) The results follow directly from Propositions 3 and 4 and from the algebra of convergence in L^2.

(ii) From Propositions 3 and 4 and the algebra of convergence in distribution it follows that $n^{1/2}(S_k^{\mathrm{u}} - S_k) - n^{1/2}(S_k^{\mathrm{p}} - S_k) = n^{1/2}(S_k^{\mathrm{u}} - S_k^{\mathrm{p}}) \to N(0, \sigma_{\mathrm{a}}^2)$ in distribution, where $\sigma_{\mathrm{a}}^2 = \sigma_{\mathrm{p}}^2 + \sigma_{\mathrm{u}}^2$. Hence, in distribution,

$$\frac{n^{1/2}(S_k^{\mathrm{u}} - S_k^{\mathrm{p}})}{\sigma_{\mathrm{a}}} = \frac{(n\hat{T}_k)^{1/2}}{\sigma_{\mathrm{a}}} \to N(0, 1), \qquad \frac{n\hat{T}_k}{\sigma_{\mathrm{a}}^2} \to \chi_1^2.$$

Proof of Proposition 5

Since the sample periodogram $I(\mathrm{x}, \omega) = (2\pi)^{-1} \sum_{k=-(n-1)}^{n-1} \hat{\gamma}_k \exp(-\mathrm{i}k\omega)$ and the sample autocovariance function $\hat{\gamma}_k$ of x at lag k are related through an invertible function, the preservation of the sample autocorrelation in the surrogate series is equivalent to preservation of the sample periodogram. In fact, $V = \{\bar{x}, \hat{\gamma}_k, k = 1, 2, \ldots\}$ is a joint sufficient statistic for a linear Gaussian process. Moreover, it is easy to show that the test statistics $\hat{S}_k^{\mathrm{u}}$ and $\hat{T}_k$ are asymptotically independent of any finite set of X_t for which $t \in \mathbb{N}$. To this end, consider the statistic $\hat{T}_k = (\hat{S}_k^{\mathrm{u}} - \hat{S}_k^{\mathrm{p}})^2$ and let $\mathcal{I} = i_1, \ldots, i_N$ be a finite subset of indices in $\mathbb{N}$. We can write

$$\hat{S}_k^{\mathrm{u}} = \frac{1}{2} \iint \left[\left\{(n-k)^{-1} \sum_{i \in \mathcal{I}} h_1^{-1} h_2^{-1} K_{12}\right\}^{1/2} - \left(n^{-1} \sum_{i \in \mathcal{I}} h_1^{-1} K_1 \times n^{-1} \sum_{i \in \mathcal{I}} h_2{}^{-1} K_2\right)^{1/2}\right]^2$$

$$+ \left[\left\{(n-k)^{-1} \sum_{i \in \{\mathbb{N} - \mathcal{I}\}} h_1^{-1} h_2^{-1} K_{12}\right\}^{1/2} - \left(n^{-1} \sum_{i \in \{\mathbb{N} - \mathcal{I}\}} h_1^{-1} K_1 \times n^{-1} \sum_{i \in \{\mathbb{N} - \mathcal{I}\}} h_2^{-1} K_2\right)^{1/2}\right]^2$$

$$\times\, w(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2,$$

where $K_1 = K\{(x_1 - X_i)/h_1\}$, $K_2 = K\{(x_2 - X_i)/h_2\}$ and $K_{12} = K\{(x_1 - X_i)/h_1, (x_2 - X_i)/h_2\}$. Now, since the first term of the integrand vanishes as $n \to \infty$ and the estimator $\hat{S}_k^{\mathrm{u}}$ is asymptotically Gaussian with limiting variance that does not depend on any finite subset of observations, the result follows immediately. The same argument holds for $\hat{S}_k^{\mathrm{p}}$. In fact, let $\hat{\rho}_k$ be the sample autocorrelation function of $\{X_t\}$ at lag k and let $\hat{S}_k^{\mathrm{p}} = 1 - 2(1 - \hat{\rho}_k^2)^{1/4}/(4 - \hat{\rho}_k^2)^{1/2}$. Then we have that $\hat{\rho}_k = n^{-1}\sum_{i\in\{\mathcal{I}\}} X_i X_{i+k} + n^{-1}\sum_{i\in\{\mathbb{N}-\mathcal{I}\}} X_i X_{i+k}$. Again, since the first of the two terms in the sum vanishes as the sample size diverges and since $\hat{\rho}_k$ is asymptotically Gaussian with limiting variance $\zeta = \sum_{i=1}^{\infty}\{\rho_{(i+k)}\rho_{(i-k)} - 2\rho_i\rho_k\}^2$, the asymptotic independence holds. In turn, since $\hat{S}_k^{\mathrm{p}}$ is a piecewise monotone function of $\hat{\rho}_k$, the result follows.

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'. The statistic $\hat{T}_k = (\hat{S}_k^{\mathrm{u}} - \hat{S}_k^{\mathrm{p}})^2$ has two components. The parametric component can be written as $\hat{S}_k^{\mathrm{p}} = g_1\{(n-k)^{-1}\sum_{t=k+1}^{n} h(X_t, X_{t-k})\}$, i.e., 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 in (3). The second component $\hat{S}_k^{\mathrm{u}}$, 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

$$\hat{S}_k^{\mathrm{u}} = 1 - \iint \left\{ f_1(x_1) \times f_2(x_2) \times f_{12}(x_1, x_2) \right\}^{1/2} \mathrm{d}x_1\,\mathrm{d}x_2 = 1 - \frac{\text{const}}{(n-k)^{1/2}} \sum_{\mathrm{x}\in\mathbb{R}^2} g_2(X_t, X_{t+k}),$$

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 above, 'const' is a real constant that depends on n, k, h_1 and h_2, and $g_2(x_1, x_2) = \{f_1(x_1) f_2(x_2) f_{12}(x_1, x_2)\}^{1/2}$. From Conditions 2 and 3, 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 in Bickel & Bühlmann (1999); so the result follows directly from the consistency of the smoothed sieve bootstrap process. The parametric estimator $\hat{S}_k^{\mathrm{p}}$ also satisfies the conditions of Theorem 3.3 in Bühlmann (1997).

References

Bickel, P. J. & Bühlmann, P. (1997). Closure of linear processes. *J. Theor. Prob.* **10**, 445–79.

Bickel, P. J. & Bühlmann, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5**, 413–46.

Brockwell, P. J. & Davis, R. A. (2009). *Time Series: Theory and Methods*. New York: Springer.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3**, 123–48.

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

Chan, K.-S. (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 Institute Communications*. Providence, Rhode Island: American Mathematical Society, pp. 77–97.

Chan, K.-S. & Tong, H. (2001). *Chaos: A Statistical Perspective*. New York: Springer.

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

Fuller, W. A. (1996). *Introduction to Statistical Time Series*. New York: John Wiley & Sons, 2nd ed.

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*. Amsterdam: Elsevier, pp. 43–63.

Granger, C. W. J. & Lin, J. L. (1994). Using the mutual information coefficient to identify lags in nonlinear models. *J. Time Ser. Anal.* **15**, 371–84.

Granger, C. W. J., Maasoumi, E. & Racine, J. (2004). A dependence metric for possibly nonlinear processes. *J. Time Ser. Anal.* **25**, 649–69.

Graves, S. (2014). *FinTS: Companion to Tsay (2005) Analysis of Financial Time Series*. R package version 0.4-5, available at http://cran.r-project.org/web/packages/FinTS/index.html.

Hjellvik, V. & Tjøstheim, D. (1995). Nonparametric tests of linearity for time series. *Biometrika* **82**, 351–68.

Hjellvik, V., Yao, Q. & Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plan. Infer.* **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. *J. Am. Statist. Assoc.* **84**, 157–64.

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

MAASOUMI, E. (1993). A compendium to information theory in economics and econometrics. *Economet. Rev.* **12**, 137–81.

MAASOUMI, E. & RACINE, J. S. (2009). A robust entropy-based test of asymmetry for discrete and continuous processes. *Economet. Rev.* **28**, 246–61.

R DEVELOPMENT CORE TEAM (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10**, 441–51.

ROBINSON, P. (1991). Consistent nonparametric entropy-based testing. *Rev. Econ. Studies* **58**, 437–53.

RUSTICELLI, E., ASHLEY, R. A., BEE DAGUM, E. & PATTERSON, D. M. (2009). A new bispectral test for nonlinear serial dependence. *Economet Rev.* **28**, 279–93.

SCHREIBER, T. (1998). Constrained randomization of time series data. *Phys. Rev. Lett.* **90**, 2105–8.

SEN, P. K., SINGER, J. M. & PEDROSO DE LIMA, A. C. (2009). *From Finite Sample to Asymptotic Methods in Statistics*. Cambridge: Cambridge University Press.

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

SKAUG, H. J. & 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 in Statistics*. New York: Springer, pp. 363–77.

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.

TJØSTHEIM, D. (1996). Measures of dependence and tests of independence. *Statistics* **28**, 249–84.

TONG, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

TSAY, R. S. (2000). Nonlinear time series models: Testing and applications. In *A Course in Time Series Analysis*, D. Peña, G. C. Tiao & R. S. Tsay, eds., chap. 10. New York: John Wiley & Sons, pp. 267–85.

TSAY, R. S. (2010). *Analysis of Financial Time Series*. Hoboken: Wiley, 3rd ed.