TI 2001-085/1
Tinbergen Institute Discussion Paper

Tests for Serial Independence and Linearity based on Correlation Integrals

Cees Diks
Sebastiano Manzan

Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), Faculty of Economics and Econometrics, University of Amsterdam, and Tinbergen Institute

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Tests for serial independence and linearity based on correlation integrals

Cees Diks & Sebastiano Manzan

Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), Faculty of Economics and Econometrics, and Tinbergen Institute, University of Amsterdam
Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

E-mail: manzan@fee.uva.nl

August 3, 2001

Abstract

We propose information theoretic tests for serial independence and linearity in time series. The test statistics are based on the conditional mutual information, a general measure of dependence between lagged variables. In case of rejecting the null hypothesis, this readily provides insights into the lags through which the dependence arises. The conditional mutual information is estimated using the correlation integral from chaos theory. The significance of the test statistics is determined with a permutation procedure and a parametric bootstrap in the tests for serial independence and linearity, respectively. The size and power properties of the tests are examined numerically and illustrated with applications to some benchmark time series.

Keywords: serial independence, linearity, bootstrap, permutation test, nonparametric estimation, nonlinear time series analysis, correlation integral

1 Introduction

It is well known that processes with zero autocorrelation can still exhibit higher order dependence or nonlinear dependence. This has motivated the development of tests for serial independence with power against alternatives exhibiting general types of dependence. The nonparametric approach avoids restrictive assumptions on the marginal distribution of the process. Nonparametric measures of divergence between two distributions are of relevance, since they can be used to develop one-sided tests for serial independence. This has led several authors to consider nonparametric divergence measures in a time series context. For example, Robinson (1991) considers the Kullback-Leibler information, while Skaug and Tjøstheim (1993a) study the Blum, Kiefer & Rosenblatt (1961) statistic. Some divergence measures based on the probability density functions are compared in Skaug and Tjøstheim (1993b). Other recently proposed nonparametric tests for serial independence in time series can be found, for example, in Chan and Tran (1992), Delgado (1996), Aparicio and Escribano (1998) and Hong (1998). For a review on nonparametric testing for serial independence we refer to Tjøstheim (1996).

Although evidence against the null hypothesis of independence for a particular time series suggests the presence of structure in the time series, it usually provides little insight into the nature of this structure. Since linear models are a benchmark class of models with known properties, testing the hypothesis of linearity is a natural next step in practice. One way of testing for linearity is by applying a test for independence to the residuals of an estimated linear model. A rejection of the null hypothesis provides evidence suggesting that some structure is left in the residuals upon removing linear dependence, and hence that a linear model is not appropriate. Brock *et al.* (1996) have shown that this approach is aymptotically consistent for their BDS test, provided that the model parameters are estimated root-N consistently. An alternative approach to testing for linearity is to compare linear and nonparametric statistics, such as estimators of the conditional mean and variance, as proposed by Hjellvik and Tjøstheim (1995) and Hjellvik *et al.* (1998). This avoids pre-whitening of the time series (which typically leads to a reduction in power) and also preserves the order of dependence in the time series.

Our test statistics are closely related to the δ statistic introduced by Wu *et al.* (1993) for measuring conditional dependence. The δ statistic is defined in terms of ratios of correlation integrals. Correlation integrals originate from the study of chaotic systems, where they are important means of characterizing the dynamics of deterministic processes. Their estimation is relatively straightforward. The connection between generalized correlation integrals and information theoretic quantities, established by Prichard and Theiler (1995), shows that our statistics correspond to the second order conditional mutual information. The information theoretic quantities used by Granger and Lin (1994) are also related to ours. However, their test statistic is a generalisation of the autocorrelation function, while ours generalises the partial autocorrelation function. This renders our statistics more suitable for investigating the lag dependence, which may serve as a first step for model selection. Since the number of parameters in a parametric nonlinear time series model (such as a TAR model) typically increases fast with the number of lags selected, lag selection criteria are important for constructing parsimonious time series models. For some recent approaches to lag selection see: Auestad and Tjøstheim (1990), Cheng and Tong (1991), Tschernig and Yang (2000).

In this paper we employ bootstrap methods for determining the significance of the test statistics. In the case of testing for independence this leads to an exact level-α test. The main advantages, besides the bootstrap approach, are the following. i) The test statistics are based on information theoretic quantities. Since these are nonlinear functionals of the density function they can capture dependence in higher moments of the distribution, thus not limiting the analysis to linear dependence. ii) The conditional mutual information provides insights into the lag dependence in the time series. iii) In the linearity test we compare nonparametric and linear parametric information-theoretic quantities for the original time series. The advantage over comparing estimates of those for the residuals is that the lag dependence in the time series is preserved. iv) The connections between information theory and correlation integrals are used for efficient nonparametric estimation.

In section 2 we briefly review the information theoretic quantities, while section 3 describes the estimation methods based on correlation integrals. Sections 4 and 5 discuss the test of independence and linearity, respectively. In section 6 the size and power properties of the tests are investigated numerically for a number of linear and nonlinear models. Section 7 illustrates our approach with applications to some empirical time series.

2 Information theory

Information theory was introduced by Shannon and Wiener and its statistical application pioneered by Kullback (1959). Since our approach is closely connected with information theory we will give a brief overview here.

Let X be a continuous, possibly vector-valued, random variable with probability density function $f_X(x)$. The *Shannon entropy* is defined as

$$H(X) = -\int \ln f_X(x) f_X(x)\,\mathrm{d}x, \tag{1}$$

which is just the expected value of $-\ln f_X$, $-E(\ln f_X)$. Note that the Shannon entropy is scale dependent. For example, for the transformed variable aX, where a is a real constant, the entropy becomes $H(aX) = H(X) + m \ln a$ where m is the dimension of x.

For a pair of random variables X, Y with joint probability density function $f_{X,Y}(x, y)$, the joint entropy reads

$$H(X, Y) = -\int\int \ln f_{X,Y}(x, y) f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y. \tag{2}$$

The *conditional entropy* of X given Y is the mean entropy of X, conditional on Y:

$$H(X|Y) = -\int \ln f_{X|Y}(x \mid y) f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y, \tag{3}$$

where $f_{X|Y}(x \mid y)$ denotes the conditional probability density function of X, given $Y = y$. It can be easily verified that $H(X|Y) = H(X, Y) - H(Y)$. Note that $H(X|Y)$ is not invariant under changing its arguments. However, the *mutual information*, defined as

$$I(X, Y) = \int\int \ln\left(\frac{f_{X,Y}(x, y)}{f_X(x) f_Y(y)}\right) f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y, \tag{4}$$

is a symmetric measure of dependence between X and Y. The mutual information measures the average information contained in one of the random variables about the other. The symmetry follows directly from the definition and also becomes obvious after expressing it in terms of entropies: $I(X, Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$. The mutual information is invariant not only under scale transformations of X and Y, but more generally, under all continuous one-to-one transformations of X and Y.

The mutual information is non-negative, $I(X, Y) \geq 0$, with equality if and only if $f_{X,Y} = f_X f_Y$. This property makes it a useful quantity for testing independence hypotheses. The mutual information is a special case of the Kullback-Leibler information, defined for two pdf's $f(x)$ and $g(x)$ as:

$$I_{\mathrm{KL}} = \int \ln\left(\frac{f(x)}{g(x)}\right) f(x)\,\mathrm{d}x. \tag{5}$$

Robinson (1991) proposed a test for independence based on the Kullback-Leibler information, which in that case reduces to the mutual information.

For testing conditional independence of X and Y, given a third random variable Z, it is useful to consider the *conditional mutual information*, defined by

$$I(X, Y|Z) = \int\int\int \ln\left(\frac{f_{X|Y,Z}(x|y, z)}{f_{X|Z}(x|z)}\right) f_{X,Y,Z}(x, y, z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \tag{6}$$

The conditional mutual information quantifies the average amount of additional information in Y about X, given the information about X already contained in Z. This can be seen by expressing it as $I(X,Y|Z) = H(X|Z) - H(X|Y,Z) = -H(X,Y,Z) + H(X,Z) + H(Y,Z) - H(Z)$. Because the entropy does not increase upon conditioning on additional information, $I(X,Y|Z) \geq 0$, with equality if and only if X and Y are conditionally independent, given Z.

3 Correlation integrals

The generalized order-q correlation integral of X is defined as

$$C_q(X;\epsilon) = \left[\int \left(\int I_{(\|x-x'\|\leq\epsilon)} f_X(x')\mathrm{d}x'\right)^{q-1} f_X(x)\mathrm{d}x\right]^{\frac{1}{q-1}}, \tag{7}$$

where $I_{(\cdot)}$ denotes the indicator function taking values 0 and 1, and $\|\cdot\|$ denotes the supremum norm

$$\|\mathbf{x}\| = \sup_{i=1,\ldots,\dim\mathbf{x}} \mid x_i \mid . \tag{8}$$

The parameter ϵ plays the role of a bandwidth. Correlation integral estimates are being used frequently in chaos theory to study fractal structures and to characterize deterministic time series. Correlation integrals are also useful for testing for serial independence, because the generalized correlation integral factorises when the elements of X are i.i.d. (independent and identically distributed). The factorisation for $q = 2$ was used in the BDS test for independence, based on $C_2(X;\epsilon)$.

To describe the relation between information theoretic quantities and correlation integrals it is convenient to notice that the Shannon entropy is a special case of a generalised entropy, the *Renyi entropy*, defined by

$$H_q(X) = -\frac{1}{q-1}\ln\int (f_X(x))^{q-1} f_X(x)\,\mathrm{d}x, \tag{9}$$

where q denotes the order of the Renyi entropy. Indeed, by taking the limit $q \to 1$, one obtains using l'Hopital's rule,

$$\lim_{q\to1} H_q(X) == -\left.\frac{\mathrm{d}}{\mathrm{d}q}\right|_{q=1}\int f_X(x)^{(q-1)} f_X(x)\,\mathrm{d}x = -\int \ln f_X(x) f_X(x)\,\mathrm{d}x, \tag{10}$$

which is the Shannon entropy.

Upon taking logarithms in Eqn. (7) we obtain

$$-\ln C(X;\epsilon) = -\frac{1}{q-1}\ln\left[\int \left(\int I_{(\|x-x'\|\leq\epsilon)} f_X(x')\mathrm{d}x'\right)^{q-1} f_X(x)\mathrm{d}x\right], \tag{11}$$

which is very similar to the generalized Renyi entropy, given in Eqn. (9), the only difference being the replacement of the term $f_X(x)$ within brackets by an integral of $f_X(x')$ over an ϵ-ball around x. The inner integral in Eqn. (7) behaves as $\epsilon^m f_X(x)$ for small ϵ, where m is the dimension of X. Thus, up to an ϵ dependent scale factor, the correlation integral will correspond to the integral in Eqn. (9). The relationship between $H_q(X)$ and $C_q(X;\epsilon)$ for ϵ small is

$$H_q(X) \simeq -\ln C_q(X;\epsilon) + m\ln\epsilon. \tag{12}$$

This shows that estimated correlation integrals provide nonparametric estimates of $H_q(X)$, and vice versa. To give an example of how this leads to estimates of information theoretic quantities, let us consider $I_q(X, Y)$ the q-th order mutual information between X and Y, given by

$$I_q(X, Y) = H_q(X) + H_q(Y) - H_q(X, Y). \tag{13}$$

Given estimated correlation integrals $\widehat{C}_q(X; \epsilon)$, $\widehat{C}_q(Y; \epsilon)$ and $\widehat{C}_q(X, Y; \epsilon)$, an estimator for $I_q(X, Y)$ is given by

$$\widehat{I}_q(X, Y) = \ln \widehat{C}_q(X, Y; \epsilon) - \ln \widehat{C}_q(X; \epsilon) - \ln \widehat{C}_q(Y; \epsilon). \tag{14}$$

Note that the terms proportional to $m \ln \epsilon$ cancel because the dimension of (X, Y) is the sum of those of X and Y. A similar cancellation occurs in the conditional mutual information, for which we obtain analogously:

$$\widehat{I}_q(X, Y|Z) = \ln \widehat{C}_q(X, Y, Z; \epsilon) - \ln \widehat{C}_q(X, Z; \epsilon) - \ln \widehat{C}_q(Y, Z; \epsilon) + \ln \widehat{C}_q(Z; \epsilon). \tag{15}$$

Further details on the connection between correlation integrals and information theory can be found in Prichard and Theiler (1995).

The choice $q = 2$ is by far the most popular in chaos analysis, since it allows for efficient estimation algorithms. The conditional mutual information $I_q(X, Y \mid Z)$ strictly speaking is not positive definite for $q \neq 1$. This means that it is possible to construct examples of variables X and Y, which are conditionally dependent given Z, and for which $I_2(X, Y \mid Z)$ is zero or negative. If $I_2(X, Y \mid Z)$ is zero, the test based on I_2 asymptotically does not have unit power against this alternative. This situation appears to be very exceptional, and usually $I_2(X, Y \mid Z)$ is either positive or negative. This suggests that a one-sided test, rejecting for $I_2(X, Y; \epsilon)$ large, is not always optimal. In practice, however, I_2 behaves much like I_1 in that we usually observe larger power for one-sided tests (rejecting for large I_2) than for two-sided tests. This led us to choose $q = 2$, together with a one-sided implementation of the test.

Prichard and Theiler (1995) use the standard indicator function kernel but mention that it might not be optimal from a statistical point of view. To their defence we might add that in chaos theory this has never been an issue, since there the definition of the correlation integral is usually taken as a starting point. Estimation using U-statistics then gives the indicator function kernel in a natural way. Here a similar point of view can be taken. In the present testing context we have no a priori reasons to assume that redefining the correlation integral using another kernel function will lead to more powerful tests. This might seem counterintuitive in a nonparametric setting, since it is known that some kernels are better for nonparametric function estimation than others. However, there are important differences between nonparametric function estimation problems and nonparametric tests. For example, in a nonparametric function estimation problem the bandwidth should tend to zero at the proper rate asymptotically to obtain consistency. For testing this need not be the case as is clearly illustrated by the fact that the asymptotic theory for the BDS test holds for any fixed value of ϵ. In any case, it should be kept in mind that we wish to estimate a functional of the pdf, and not the pdf itself.

It is beyond the scope of this paper to examine which kernels are optimal for testing. However, one essential property should not be left unnoticed here: the choice of the kernel function cannot be made independently from that of the norm used to calculate distances. For example, factorisation of the correlation integral for i.i.d. data occurs for the indicator function only with the L_∞ norm, and for the gaussian kernel only with the L_2 norm.

4 Testing for Serial Independence

In the present paper we propose to investigate and test for independence in time series using the conditional mutual information defined above. The advantage of using this as a test statistic is that it captures dependences while the conditioning on intermediate values of the time series will also give insights on the order of the underlying process.

First let us consider mutual information in a time series setting. For a stationary time series $\{X_t\}_{t=1}^T$ we define the delay vectors as

$$\mathbf{X}_t^m = (X_t, \ldots, X_{t-m+1})', \tag{16}$$

where the prime denotes the transposed. The dimension m is referred to as the embedding dimension. The total number of vectors, N, obtained in this way is $N = T - m + 1$.

The conditional mutual information between X_t and X_{t-m} given the intermediate observations, $\mathbf{X}_{t-1}^{m-1}$ is given by

$$I(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1}) = -H(\mathbf{X}_{t-1}^{m+1}) + 2H(\mathbf{X}_t^m) - H(\mathbf{X}_{t-1}^{m-1}). \tag{17}$$

The conditional mutual information has a particular interpretation in a time series setting: if X_t is a Markov process of order k, the conditional probability density depends only in the last k lagged values of the time series and further lags contain no additional information. The conditional mutual information between X_t and X_{t-m} will become zero for $m > k$ and positive for $m \leq k$. In this sense the conditional mutual information can be interpreted as an order identifier.

Another useful interpretation is the following. The average amount of information about X_t in $\mathbf{X}_{t-1}^m$ is given by $I(X_t, \mathbf{X}_{t-1}^m)$, while the average amount of information about X_t in $\mathbf{X}_{t-1}^{m-1}$ only is given by $I(X_t, \mathbf{X}_{t-1}^{m-1})$. If these two information measures are subtracted, one arrives at

$$I(X_t, \mathbf{X}_{t-1}^m) - I(X_t, \mathbf{X}_{t-1}^{m-1}) = I(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1}), \tag{18}$$

the conditional mutual information. This clearly demonstrates that the conditional mutual information quantifies the average amount of extra information that $\mathbf{X}_{t-1}^m$ contains about X_t, in addition to the information already in $\mathbf{X}_{t-1}^{m-1}$. If X_{t-m} contains no extra information about values of X_t in addition to that in X_{t-m-1}, $I(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1}) = 0$. If, on the other hand, X_{t-m} does contain extra information on X_t, we expect $I(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1}) > 0$. We thus wish to perform a one-sided test based on $I(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1})$, which can be estimated from correlation integrals.

Upon introducing $C_m(\epsilon)$ and $\widehat{C}_m(\epsilon)$ as shorthand notation for $C_2(\mathbf{X}_t^m; \epsilon)$ and its estimator $\widehat{C}_2(\mathbf{X}_t^m; \epsilon)$, respectively, we may write

$$\widehat{I}(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1}) = -2 \ln \widehat{C}_m(\epsilon) + \ln \widehat{C}_{m+1}(\epsilon) + \ln \widehat{C}_{m-1}(\epsilon). \tag{19}$$

The second order ($q = 2$) correlation integral for the m-dimensional delay vectors $\mathbf{X}_t^m$ is

$$C_m(\epsilon) = \int \int I_{(\|s-t\| \leq \epsilon)} f_{\mathbf{X}_m}(s) f_{\mathbf{X}_m}(t) \, \mathrm{d}s \, \mathrm{d}t. \tag{20}$$

Because this is just the expectation of the kernel function, $E(I_{(\|\mathbf{X}_m^1 - \mathbf{X}_m^2\| \leq \epsilon)})$, it can be estimated straightforwardly in a U-statistics framework, by

$$\widehat{C}_m(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} I_{(\|\mathbf{X}_i^m - \mathbf{X}_j^m\| \leq \epsilon)}. \tag{21}$$

Note that the conditional mutual information is an unbounded measure of conditional dependence. In our implementation we use a transformed version of the mutual information,

$$\widehat{\delta}_m(\epsilon) = 1 - \exp(-\widehat{I}(X_t, X_{t-m}|(X_{t-1}, ..., X_{t-m+1})) = 1 - \frac{[\widehat{C}_m(\epsilon)]^2}{\widehat{C}_{m-1}(\epsilon)\widehat{C}_{m+1}(\epsilon)}, \quad (22)$$

which takes values between 0 and 1. The use of $\widehat{\delta}_m(\epsilon)$ was first proposed by Savit and Green (1991) to determine the dimension of a chaotic attractor. Wu *et al.* (1993) derived the asymptotic distribution under the null hypothesis of an i.i.d. process. The asymptotic distribution is

$$T^{\frac{1}{2}}\widehat{\delta}_m(\epsilon) \xrightarrow{d} N(0, V_{\delta_m}) \quad (23)$$

where the asymptotic variance is given by

$$V_{\delta_m} = 4\left\{\left(\frac{K_1(\epsilon)}{C_1(\epsilon)}\right)^{m-1}\left[\left(\frac{K_1(\epsilon)}{C_1(\epsilon)}\right)^2 - 1\right]\right\}^2. \quad (24)$$

For the correlation integral at embedding dimension 1, the estimator $\widehat{C}^1(\epsilon)$ is used in the BDS test, while $K^1(\epsilon)$ is estimated by

$$\widehat{K_1}(\epsilon) = \frac{2}{N(N-1)(N-2)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sum_{k=j+1}^{N} I_{(|x_i-x_j|\leq\epsilon)}I_{(|x_j-x_k|\leq\epsilon)}. \quad (25)$$

It is known that the normal approximation based on the asymptotic distribution does not always perform well for moderate sample sizes. In the simulation study we will show that problems also arise for $\widehat{\delta}_m(\epsilon)$. This is the main motivation for using a bootstrap approach for determining the null distribution of the test statistic.

Recently, Diks (1999) suggested to test for independence using a permutation test for the BDS statistic. The advantages are an exact level test and a simplification of the test procedure: $\widehat{C}_1(\epsilon)$ and $\widehat{K_1}(\epsilon)$ are functions of the order statistic and hence are invariant under permutation of the time series. In this way p-values are not affected by normalisation of location and scale of the statistics. In addition permutation tests are easily implemented, and avoid the cumbersome computations needed to obtain the asymptotic variance. The test procedure is thus composed of the following steps:

1. Calculate $\widehat{\delta}_m(\epsilon)$ for the time series $\{X_t\}_{t=1}^T$.
2. Randomly permute the time series and obtain $\{\widetilde{X}_t\}_{t=1}^T$.
3. Calculate the test statistic on $\{\widetilde{X}_t\}_{t=1}^T$, denoted by $\widetilde{\delta}_m(\epsilon)$.
4. Repeat steps 2-3 B times. In the simulations we set B to 199.
5. Calculate the one-sided bootstrap p-value as

$$\widehat{p} = \frac{1 + \#\left[\widetilde{\delta}_m(\epsilon) \geq \widehat{\delta}_m(\epsilon)\right]}{1+B}$$

6. Reject the null hypothesis of independence if $\widehat{p} \leq \alpha$, where α denotes the chosen significance level.

5 Testing for Linearity

We test for linearity by comparing a nonparametric estimate of the conditional mutual information with a parametric counterpart. This amounts to comparing the extra amount of information contained in X_{t-m} about X_t with the expected amount of extra information under the null of linearity.

For linear gaussian processes, we have $\mathbf{X}_t^m \sim \mathcal{N}(\mu_m, \Sigma_m)$ where Σ_m denotes the variance-covariance matrix of the m-dimensional vector of lagged values of the process $\{X_t\}_{t=1}^T$. The gaussian Renyi entropy for $\mathbf{X}_t^m$ then becomes

$$\mathcal{H}(\mathbf{X}_t^m) = -\int \ln\left\{|2\pi\Sigma_m|^{-\frac{1}{2}}\right\}\left[-\frac{1}{2}(x-\mu_m)'\Sigma_m^{-1}(x-\mu_m)\right] f_{X_m}(x)\,\mathrm{d}x = \frac{m}{2}\ln(2\pi)+\frac{1}{2}\ln|\Sigma_m|, \tag{26}$$

which is independent of the order q.

The gaussian mutual information and conditional mutual information become

$$\mathcal{I}(X_t, X_{t-1}^m) = \frac{1}{2}\ln\left(\frac{|\Sigma_1||\Sigma_m|}{|\Sigma_{m+1}|}\right), \tag{27}$$

and

$$\mathcal{I}(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1}) = \frac{1}{2}\ln\left(\frac{|\Sigma_m|^2}{|\Sigma_{m+1}||\Sigma_{m-1}|}\right) \tag{28}$$

respectively. It follows that the linearized version of δ_m, δ_m^{lin}, is given by

$$\delta_m^{\mathrm{lin}} = 1-\exp(-I(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1})) = 1-\sqrt{\frac{|\Sigma_{m-1}||\Sigma_{m+1}|}{|\Sigma_m|^2}}. \tag{29}$$

Because Σ_m is a symmetric positive definite matrix we can factorize it as $\Sigma_m = L_m' L_m$ where L_m is a lower triangular matrix. It is then immediate that $|\Sigma_m| = |L_m|^2 = \prod_{j=1}^m l_j^2$ where l_j is the j-th diagonal element of L. We can now express δ_m^{lin} as

$$\delta_m^{\mathrm{lin}} = 1-\frac{l_{m+1}}{l_m}. \tag{30}$$

The test statistic is an estimate of $\mu_m(\epsilon) = \delta_m(\epsilon) - \delta_m^{\mathrm{lin}}(\epsilon)$, which quantifies the difference between the general and the linearized $\delta_m(\epsilon)$. Upon subtracting the estimators for $\delta_m(\epsilon)$ and $\delta_m^{\mathrm{lin}}(\epsilon)$, one obtains

$$\widehat{\mu}_m(\epsilon) = \widehat{\delta}_m(\epsilon) - \widehat{\delta}_m^{\mathrm{lin}}(\epsilon) = \frac{\widehat{l}_{m+1}}{\widehat{l}_m} - \frac{\left[\widehat{C}_m(\epsilon)\right]^2}{\widehat{C}_{m-1}(\epsilon)\widehat{C}_{m+1}(\epsilon)}, \tag{31}$$

where for $\widehat{l}_m$ a consistent estimator is used, based on triangularization of the sample variance-covariance matrix.

Again, we set up a bootstrap procedure to approximate the null distribution of the test statistic. The test is composed of the following steps:

1. Calculate $\widehat{\mu}_m(\epsilon)$ for the time series $\{X_t\}_{t=1}^T$.

2. Estimate an AR(d) model for $d = 1, ..., d^{max}$ and choose the optimal order $\widehat{d}$ according to a selection criterium. In the simulation and the empirical applications we used AIC selection criterium.

3. Generate data using the estimated parameters and gaussian innovations; the bootstrap time series is given by

$$\widetilde{x}_t = \sum_{i=1}^{\widehat{d}} \widehat{\beta}\widetilde{x}_{t-1} + \epsilon_t \tag{32}$$

with $\widehat{\beta}$ the estimated parameters and ϵ_t drawn from the standard normal distribution.

4. Calculate $\widetilde{\mu}_m(\epsilon)$ for the bootstrap time series.

5. Repeat steps 3–4 B times. We use B equal to 199.

6. Calculate the one-sided bootstrap p-value as

$$\widehat{p} = \frac{1 + \# \left[\widetilde{\mu}_m(\epsilon) \geq \widehat{\mu}_m(\epsilon)\right]}{1 + B}$$

7. Reject the null hypothesis of linearity if $\widehat{p} \leq \alpha$, where α denotes the significance level.

6 Simulations

6.1 Test for Serial Independence

Before examining the power of our test for various models, we first examine the size of the asymptotic test for independence for the δ statistic. Note that checking the size of the permutation test for independence is not necessary, since the permutation test by construction has exact level. Table 1 shows the size of the asymptotic test for sample sizes of 100, 200 and 500 and for ϵ equal to 0.5, 1.0, 1.5 and 2,based on 1000 simulations. In all cases the asymptotic test has a tendency of over-rejecting. As expected, increasing the time series length T improves the size of the asymptotic test. For all time series lengths, the size improves upon taking larger bandwidths, but increasing the embedding dimension m has an adverse effect. These results clearly demonstrate the overall poor performance of the asymptotic test for moderate sample sizes.

Table 1 near here

We next investigate the finite sample performance of our test for independence for the models given in Table 2. Throughout we use 1000 simulations for each case, keeping the number of bootstrap replications fixed at $B = 199$. The results are shown in Table 3. Apart from the AR(1) model and the Asymmetric Tent Map (ATM), they are all nonlinear stochastic models with zero autocorrelation at all lags. For these models the application of autocorrelation based tests would fail to detect any dependence. The ATM is included because it is an example of a chaotic model that has the autocorrelation structure of an AR(1) process. We analyzed the conditional dependence in the first four lags for sample sizes $T = 100$ and $T = 200$. We considered three values of the bandwidth equal to 0.5, 1.0 and 1.5.

Table 2 near here

Table 3 near here

For the AR(1) model the permutation test has power close to unity for the first lag for all sample sizes. For higher lags the rejection rate is close to the nominal level, confirming the ability of the test to detect conditional dependence, which only occurs through the first lag. Note that the power for the lags larger than 1 are even smaller than the size. This possibly results from the fact that there is conditional independence in this process for higher lags, but no unconditional independence (our null hypothesis, under which the bootstrap is performed).

For the chaotic ATM, the test has unit power at lag one for all our choices for the time series length and the bandwidth. The obtained rejection rates for this model were zero for all higher lags, which have no conditional dependence.

The BILINEAR model exhibits conditional dependence through the first two lags. For this model larger sample sizes clearly improve the power of the test. As expected the test has power against this alternative only for the first and second lag.

The results for the NLMA model show power only at the third lag: for $\epsilon = 0.5$ the power goes from 0.25 to 0.54 while for $\epsilon = 1.0$ it ranges from 0.70 to 0.94. This clearly demonstrates that the performance of the test depends on the choice of the bandwidth.

The test has also power against the (first order) TAR model: for $\epsilon = 1.0$ the rejection rate is 0.62 for samples size 100 and 0.89 for $T = 200$. In this case $\epsilon = 1.5$ shows less power than the smaller bandwidths for the first lag ($m = 1$). Here it can also be observed that there is some power in the second lag. We conjecture that this "leakage" of power is the result of taking a bandwidth too large compared to the length scale on which the conditional distribution of X_t given past observations changes.

For the ARCH(1) model the test has remarkably high power already at sample size $T = 100$: for $\epsilon = 1.0$ it goes from 0.86 to 0.99 for time series lengths of $T = 200$. Some marginal power is also detected in the second lag and no evidence of deviations from the null occur in the third and fourth lag.

The test also has power against the GARCH alternative. For the GARCH(1,1) model the test has power for all four lags analyzed. In this case the interpretation in terms of order is not possible, as the model for X_t is of infinite Markov order.

Although the optimal bandwidth is expected to depend on the alternative at hand, a bandwidth of 1.0 appears to be reasonable for the processes examined here.

There are of course many tests for independence with which we can compare ours. However, it appears unreasonable to compare an omnibus test with a test which has power against specific alternatives. It can be expected that tests which are designed to pick up specific types of dependence, such as changes in conditional mean or variance, have larger power for specific alternatives than omnibus tests such as the BDS test and ours. Therefore we decided to compare our test only with the BDS test. The latter can also easily be implemented as a permutation test, so that the size is exact and power comparisons are meaningful. Even taking this into account it can hardly be expected that our test or the BDS test is uniformly more powerful than the other, which makes direct power comparisons for specific models not very interesting. However, we can compare our results qualitatively to the BDS test, focusing on the behaviour of the power function with changing lag m. Since the BDS test is sensitive to dependence, and not only conditional dependence, we expect it to have a tendency of rejecting beyond lags for which the first evidence for dependence is found.

Table 4 shows the results for some of the models obtained with the permutation version of the BDS test with $\epsilon = 1.0$ and $T = 100$. In a comparison with Table 3, it can be observed that the BDS test has a tendency of rejecting for embedding dimension $m > k$ when there is conditional dependence only up to $m = k$. These results illustrate our earlier point that the δ test is more suitable for obtaining insights into the lag dependence structure than the BDS test. In most cases with conditional dependence on the m'th lag, the power of the BDS test is larger than that of the δ test, but an exception is found for the NLMA model ($m = 3$).

Table 4 near here

6.2 Test for Linearity

We first show the size properties of the test in Table 5. For ϵ smaller than 1.5, the test is correctly sized at all the four lags taken into account. For higher bandwidth values the first lag has the tendency to underreject the null hypothesis while larger lags seem to be relatively unaffected by the choice of ϵ. These size considerations seem to suggest to take the bandwidth in the interval 0.5–1.0. Table 5 shows results for the AR(1) parameter equal to 0.6. Similar results were found upon changing the value of the AR(1) coefficient.

Table 5 near here

Table 6 near here

Table 6 shows the power of the test for linearity for time series lengths $T = 100$ and $T = 200$ and for different bandwidth values. For $\epsilon = 1.0$ and $T = 200$ the test has power (at 5% significance level) 0.63 and 0.85 in the first and second lag respectively against the BILINEAR alternative, 0.93 on the third lag for the NLMA model, 0.85 on the first lag for TAR model and 0.99 for the ARCH(1) model. The test has also power on various lags for the GARCH(1,1) model.

Power considerations suggest that $\epsilon = 0.5$ performs poorly in comparison with higher bandwidths. A reasonable trade-off between size and power seems to suggest a choice of $\epsilon \approx 1.0$. For some models the power of the linearity test is slightly smaller than for the independence test: this lack of power can be due to the fact that we are testing for a more general null hypothesis with the linearity test. The models examined here do not exhibit linear structure. The fact that the differences in power are small suggests that little is lost in terms of size and power when the more general null hypothesis of linearity is tested for.

7 Empirical Applications

We applied our tests for independence and linearity to three benchmark time series in the statistical literature: the lynx data ($T = 114$), the sunspot data ($T = 288$) and the blowfly data ($T = 361$). See Tong (1990) for further details related to these time series. The results are summarized in Tables 7–9.

Table 7 near here

Table 8 near here

Table 9 near here

For the relatively short lynx time series there were only very few neighbouring points in high embedding dimensions (an effect known as the curse of dimensionality). Therefore we tested up to a maximum lag m of 5 for the lynx data and 10 for the other two time series. We applied the tests using different bandwidths: $\epsilon = 0.5, 1.0$ and 1.5. For the smallest bandwidth, we found evidence suggesting dependence for the lynx data at lags $m = 1$ and $m = 2$ while evidence for nonlinearity was found only at $m = 1$. Even though evidence for dependence is present for all banwidths, the information concerning the relevant lags is mixed. While for small ϵ the test rejects, it does not necessarily do so for larger ϵ. This is a typical feature of TAR models as shown by the results in Tables 3 and 6 where maximal power is reached for the smaller bandwidth values.

For the sunspot data the results are similar: independence is rejected at lags 1 and 3 but for the second lag only for the smallest bandwidth. Linearity is also rejected for the smallest ϵ-value at the first two lags.

The blowfly data show signs of dependence for the first lag but not for any higher lags. No evidence for non-linearity is found for the blowfly data.

8 Conclusion

In this paper we propose information theoretic bootstrap tests for independence and linearity. The results of the simulation study show that the test for independence has good power properties at moderate sample sizes, when compared to the BDS test, and in addition provides insights into the lag dependence in the data generating process. The power of both tests typically increases when larger bandwidth values ϵ are taken. However, care should be taken to avoid "leakage" of power to other lags as a result of taking the bandwidth ϵ too large. The choice $\epsilon = 1$ appears to be a reasonable trade-off between these effects for the models examined. The size of the independence test by construction is equal to the nominal size. For the model examined, the size of the linearity test turned out to be also close to the nominal level. Moreover, for models without linear structure, the power of the linearity test was found to be close to that of the independence test. This suggests that little is lost in terms of size and power when testing the more general null hypothesis of linearity instead of independence.

Acknowledgements

The authors wish to thank the Netherlands' Organisation for Scientific Research (NWO) for their financial support, and to acknowledge Dag Tjøstheim for his comments on an earlier draft of this manuscript.

References

Aparicio, F. M. and Escribano, A. (1998). Information-theoretic analysis of serial dependence and cointegration. *Studies in Nonlinear Dynamics and Econometrics*, **3,** number 3, 119–140.

Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika*, **77,** number 4, 669–687.

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, **32,** 485–98.

Brock, W. A., Dechert, W. D., Scheinkman, J. A. and B, LeBaron (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, **15,** number 3, 197–235.

Chan, N. H. and Tran, L. T. (1992). Nonparametric tests for serial dependence. *Journal of Time Series Analysis*, **13,** number 1, 19–28.

Cheng, B. and Tong, H. (1991). On consistent nonparametric order determination and chaos. *Journal of the Royal Statistical Society* B, **54,** number 2, 427–449.

Delgado, M. A. (1996). Testing serial independence using the sample distribution function. *Journal of Time Series Analysis*, **17,** number 3, 271–285.

Diks, C. (1999). Consistent testing for serial independence. Technical Report. CeNDEF working paper 99-02, University of Amsterdam.

Granger, C. W. and Lin, J. L. (1994). Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis*, **15,** number 4, 371–384.

Hjellvik, V. and Tjøstheim, D. (1995). Nonparametric tests of linearity for time series. *Biometrika*, **82,** number 2, 351–368.

Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using polynomial approximation. *Journal of Statistical Planning and Inference*, **68,** number 2, 295–321.

Hong, Y. (1998). Testing for pairwise independence via the empirical distribution function. *Journal of Royal Statistical Society* B, **60,** 429–453.

Kullback, S. (1959). *Information Theory and Statistics*. New York: Wiley & Sons.

Prichard, D. and Theiler, J. (1995). Generalized redundancies for time series analysis. *Physica* D, **84,** 476–493.

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

Savit, R. and Green, M. (1991). Time series and dependent variables. *Physica* D, **50,** number 50, i95–116.

Skaug, H. J. and Tjøstheim, D. (1993a). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika*, **80,** number 3, 591–602.

Skaug, H. J. and Tjøstheim, D. (1993b). Nonparametric tests of serial independence. In *Developments in Time Series Analysis: In Honour of Maurice B. Priestley* (ed. T. Subba Rao), pp. 207–230. Chapman & Hall, London.

Tjøstheim, D. (1996). Measures of dependence and tests of independence. *Statistics*, **28,** 249–284.

Tong, H. (1990). *Nonlinear time series analysis: A dynamical system approach.* Oxford University Press, Oxford.

Tschernig, R. and Yang, L. (2000). Nonparametric lag selection for time series. *Journal of Time Series Analysis*, **21,** number 4, 457–487.

Wu, K., Savit, R. and Brock, W. (1993). Statistical tests for deterministic effects in broad band time series. *Physica* D, **69,** 172–188.

	$T = 100$				$T = 200$				$T = 500$			
ϵ	$m = 1$	2	3	4	1	2	3	4	1	2	3	4
0.5	0.21	0.29	0.40	0.48	0.14	0.23	0.32	0.42	0.09	0.14	0.25	0.34
1.0	0.13	0.15	0.19	0.23	0.08	0.09	0.13	0.17	0.07	0.06	0.07	0.10
1.5	0.11	0.12	0.13	0.16	0.08	0.09	0.09	0.10	0.06	0.06	0.06	0.08
2.0	0.15	0.16	0.17	0.16	0.08	0.10	0.10	0.10	0.06	0.07	0.07	0.07

Table 1

Name	Model
AR(1)	$y_t = 0.6y_{t-1} + u_t$
ATM	$y_t = 1.25y_{t-1}I(0 \leq y_{t-1} \leq 0.8) + 5(1 - y_{t-1})I(0.8 < y_{t-1} \leq 1)$
BILINEAR	$y_t = 0.6u_{t-1}y_{t-2} + u_t$
NLMA	$y_t = 0.6u_{t-3}^2 + u_t$
TAR	$y_t = -0.5y_{t-1}I(y_{t-1} \leq 1) + 0.6y_{t-1}I(y_{t-1} > 1) + u_t$
ARCH(1)	$y_t = \sqrt{h_t}u_t,\ h_t = 1 + 0.6y_{t-1}^2$
GARCH(1,1)	$y_t = \sqrt{h_t}u_t,\ h_t = 1 + 0.3y_{t-1}^2 + 0.6h_{t-1}$

Table 2

		$T = 100$				$T = 200$			
Model	ϵ	$m = 1$	2	3	4	1	2	3	4
AR(1)	0.5	0.91	0.03	0.02	0.01	1.00	0.02	0.02	0.01
	1.0	0.97	0.03	0.02	0.02	1.00	0.02	0.01	0.02
	1.5	0.98	0.04	0.02	0.02	1.00	0.04	0.01	0.02
ATM	0.5	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	1.0	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	1.5	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
BILINEAR	0.5	0.26	0.26	0.02	0.02	0.50	0.58	0.04	0.01
	1.0	0.37	0.54	0.04	0.03	0.63	0.86	0.07	0.03
	1.5	0.37	0.58	0.05	0.02	0.64	0.88	0.05	0.02
NLMA	0.5	0.03	0.05	0.25	0.03	0.06	0.06	0.54	0.02
	1.0	0.05	0.07	0.66	0.03	0.06	0.06	0.95	0.02
	1.5	0.05	0.06	0.70	0.03	0.06	0.05	0.96	0.03
TAR	0.5	0.61	0.04	0.02	0.02	0.91	0.05	0.04	0.01
	1.0	0.62	0.06	0.03	0.03	0.89	0.07	0.03	0.03
	1.5	0.45	0.08	0.03	0.04	0.73	0.11	0.04	0.03
ARCH	0.5	0.73	0.05	0.02	0.02	0.98	0.06	0.02	0.02
	1.0	0.86	0.06	0.02	0.03	0.99	0.08	0.03	0.02
	1.5	0.86	0.07	0.02	0.01	0.99	0.10	0.02	0.02
GARCH	0.5	0.47	0.21	0.06	0.02	0.61	0.36	0.15	0.07
	1.0	0.64	0.40	0.17	0.06	0.90	0.68	0.37	0.13
	1.5	0.63	0.38	0.18	0.07	0.89	0.72	0.41	0.16

Table 3

Model	$m=1$	2	3	4
AR(1)	0.98	0.96	0.94	0.92
ATM	1.00	1.00	1.00	1.00
BILINEAR	0.39	0.60	0.62	0.61
NLMA	0.07	0.08	0.18	0.25
TAR	0.64	0.60	0.53	0.46
ARCH	0.87	0.82	0.77	0.71
GARCH	0.63	0.70	0.73	0.75

Table 4

	$T=100$				$T=200$			
ϵ	$m=1$	2	3	4	1	2	3	4
0.5	0.06	0.05	0.04	0.04	0.04	0.05	0.04	0.05
1.0	0.06	0.04	0.05	0.05	0.06	0.04	0.05	0.05
1.5	0.03	0.05	0.04	0.06	0.01	0.05	0.03	0.04
2.0	0.00	0.05	0.05	0.05	0.00	0.04	0.05	0.05

Table 5

		$T = 100$				$T = 200$			
Model	ϵ	$m = 1$	2	3	4	1	2	3	4
	0.5	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
ATM	1.0	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	1.5	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	0.5	0.24	0.21	0.04	0.03	0.53	0.55	0.04	0.02
BILINEAR	1.0	0.36	0.53	0.05	0.02	0.63	0.85	0.05	0.02
	1.5	0.40	0.59	0.04	0.03	0.65	0.89	0.05	0.03
	0.5	0.05	0.04	0.20	0.01	0.04	0.06	0.48	0.01
NLMA	1.0	0.05	0.06	0.64	0.02	0.06	0.07	0.93	0.02
	1.5	0.05	0.06	0.72	0.03	0.05	0.07	0.95	0.02
	0.5	0.57	0.03	0.02	0.02	0.89	0.04	0.03	0.02
TAR	1.0	0.58	0.04	0.03	0.04	0.85	0.08	0.03	0.03
	1.5	0.40	0.06	0.04	0.05	0.64	0.11	0.05	0.04
	0.5	0.73	0.04	0.02	0.01	0.96	0.06	0.01	0.01
ARCH	1.0	0.85	0.06	0.03	0.02	0.99	0.11	0.02	0.02
	1.5	0.86	0.07	0.03	0.03	0.99	0.13	0.03	0.03
	0.5	0.50	0.19	0.05	0.02	0.83	0.42	0.10	0.02
GARCH	1.0	0.61	0.41	0.18	0.07	0.90	0.71	0.38	0.13
	1.5	0.63	0.43	0.26	0.12	0.91	0.75	0.47	0.22

Table 6

ϵ	$m=1$	2	3	4	5
	Independence Test				
0.5	**0.01**	**0.01**	0.15	0.45	0.53
1.0	**0.01**	**0.01**	**0.01**	0.18	0.68
1.5	**0.01**	0.87	**0.01**	**0.02**	0.50
	Linearity Test				
0.5	**0.02**	0.68	0.07	0.52	0.72
1.0	0.10	0.43	0.40	0.39	0.95
1.5	0.28	0.61	0.36	0.33	0.71

Table 7

ϵ	$m=1$	2	3	4	5	6	7	8	9	10
	Independence Test									
0.5	**0.00**	**0.00**	**0.04**	0.39	0.50	0.50	0.55	0.48	0.10	**0.03**
1.0	**0.00**	0.26	**0.01**	**0.02**	0.35	0.53	0.36	0.34	0.32	0.38
1.5	**0.00**	0.99	**0.01**	**0.00**	0.07	0.46	0.31	0.05	0.04	0.16
	Linearity Test									
0.5	**0.00**	**0.00**	0.07	0.29	0.79	0.87	0.55	0.71	0.62	0.34
1.0	0.12	0.42	**0.03**	0.06	0.23	0.76	0.44	0.31	0.40	0.15
1.5	0.37	0.55	0.43	0.13	**0.04**	0.56	0.53	0.20	0.26	0.15

Table 8

ϵ	$m=1$	2	3	4	5	6	7	8	9	10
	Independence Test									
0.5	**0.00**	0.57	0.32	0.48	0.49	0.49	0.44	0.46	0.42	0.44
1.0	**0.00**	0.64	0.49	0.40	0.41	0.50	0.47	0.52	0.52	0.44
1.5	**0.00**	0.47	0.63	0.45	0.44	0.37	0.42	0.50	0.51	0.46
	Linearity Test									
0.5	0.17	0.71	0.71	0.82	0.73	0.76	0.72	0.70	0.27	0.33
1.0	0.89	0.40	0.70	0.83	0.77	0.91	0.86	0.94	0.95	0.47
1.5	0.75	0.53	0.46	0.59	0.79	0.64	0.70	0.87	0.83	0.67

Table 9