

# *Studies in Nonlinear Dynamics & Econometrics*

---

*Volume 6, Issue 2*

2002

*Article 2*

---

## Tests for Serial Independence and Linearity Based on Correlation Integrals

Cees Diks\*

Sebastiano Manzan†

\*CeNDEF – University of Amsterdam

†CeNDEF – University of Amsterdam

*Studies in Nonlinear Dynamics & Econometrics* is produced by The Berkeley Electronic Press (bepress). <http://www.bepress.com/snde>

Copyright ©2002 by the authors.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher bepress.

# Tests for Serial Independence and Linearity Based on Correlation Integrals

## **Abstract**

We propose information theoretic tests for serial independence and linearity in time series against nonlinear dependence on lagged variables, based on the conditional mutual information. The conditional mutual information, which is a general measure for dependence, is estimated using the correlation integral from chaos theory. The significance of the test statistics is determined by means of bootstrap methods. The size and power properties of the tests are examined by simulation and illustrated with applications to real US GNP data.

## 1. Introduction

It is well known that processes with zero autocorrelation can exhibit higher order dependence or nonlinear dependence. This has motivated the development of tests for serial independence with power against general types of dependence. The recent literature shows an increasing interest in nonparametric approaches since they avoid making restrictive assumptions on the marginal distribution of the process. Usually a nonparametric measure of divergence between two distributions is taken to be the basis of the test. 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) and Aparicio and Escribano (1998).

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. For example, the structure could be either linear or nonlinear. The properties of linear models are well-known, and linear modelling is relatively straightforward compared to nonlinear modelling. Therefore, before moving to nonlinear models for an observed time series, one should at least have some evidence for the presence of nonlinear dependence. Therefore, when dependence is found, 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 the BDS test for independence provides a consistent specification test when applied to residuals, provided that the model parameters are estimated root- $N$  consistently. An alternative approach to testing for linearity is that of comparing 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.

The statistics used in this paper are closely related to the  $\delta$  statistic introduced by Wu *et al.* (1993) for measuring conditional dependence. The  $\delta$  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).

The proposed tests are characterized by the following three properties. Firstly, 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. Secondly, the conditional mutual information is used, rather than mutual information. This provides insights into the lag dependence in the time series. Thirdly, in the linearity test we compare nonparametric and linear parametric information-theoretic quantities for the original time series. The advantage over testing for dependence in residuals is that the lag dependence in the time series is preserved.

In section 2. we briefly review some 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 a macroeconomic time series.

## 2. Information theory

Information theory was introduced by Shannon (1948) and Wiener (1948) 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

$$H(X) = - \int \ln f_X(x) f_X(x) dx, \quad (2.1)$$

which is just the expected value of  $-\ln f_X(X)$ ,  $-E(\ln f_X(X))$ . Similarly, for a pair of random variables  $X, Y$  with joint probability density function  $f_{X,Y}(x, y)$ , the joint entropy reads

$$H(X, Y) = - \int \int \ln f_{X,Y}(x, y) f_{X,Y}(x, y) dx dy. \quad (2.2)$$

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

$$H(X|Y) = - \int \int \ln f_{X|Y}(x | y) f_{X,Y}(x, y) dx dy, \quad (2.3)$$

where  $f_{X|Y}(x | y)$  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

$$I(X, Y) = \int \int \ln \left( \frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)} \right) f_{X,Y}(x, y) \, dx \, dy, \quad (2.4)$$

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  $Y$ . It is also non-negative,  $I(X, Y) \geq 0$ , with equality holding if and only if  $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ . This property makes it a useful quantity for testing independence hypotheses.

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

$$I(X, Y|Z) = \int \int \int \ln \left( \frac{f_{X|Y,Z}(x|y, z)}{f_{X|Z}(x|z)} \right) f_{X,Y,Z}(x, y, z) \, dx \, dy \, dz. \quad (2.5)$$

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  $I(X, Y|Z) = H(X|Z) - H(X|Y, Z) = -H(X, Y, Z) + H(X, Z) + H(Y, Z) - H(Z) = I(X|Y, Z) - I(X|Z)$ . We have  $I(X, Y|Z) \geq 0$ , with equality if and only if  $X$  and  $Y$  are conditionally independent, given  $Z$ .

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

$$H_q(X) = -\frac{1}{q-1} \ln \int (f_X(x))^{q-1} f_X(x) \, dx, \quad (2.6)$$

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

$$\lim_{q \rightarrow 1} H_q(X) = -\int \ln f_X(x) f_X(x) \, dx, \quad (2.7)$$

which is just the Shannon entropy  $H(X)$  defined in Eq. (2.1).

### 3. Correlation integrals

Next let us describe the connection with correlation integrals, and the way correlation integrals can be used as estimators for information theoretic quantities.

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

$$C_q(X; \epsilon) = \left[ \int \left( \int I_{(\|x-x'\| \leq \epsilon)} f_X(x') dx' \right)^{q-1} f_X(x) dx \right]^{\frac{1}{q-1}}, \quad (3.1)$$

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

$$\|\mathbf{x}\| = \sup_{i=1, \dots, \dim \mathbf{x}} |x_i|. \quad (3.2)$$

The parameter  $\epsilon$  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  $C_2(X; \epsilon)$ .

Upon taking logarithms in Eqn. (3.1) we obtain

$$-\ln C_q(X; \epsilon) = -\frac{1}{q-1} \ln \left[ \int \left( \int I_{(\|x-x'\| \leq \epsilon)} f_X(x') dx' \right)^{q-1} f_X(x) dx \right], \quad (3.3)$$

which differs from the generalized Renyi entropy, given in Eqn. (2.6), only in that it has the term  $f_X(x)$  within brackets replaced by an integral of  $f_X(x')$  over an  $\epsilon$ -ball around  $x$ . The inner integral in Eqn. (3.1) behaves as  $\epsilon^m f_X(x)$  for small  $\epsilon$ , where  $m$  is the dimension of  $X$ . Thus, up to an  $\epsilon$  dependent scale factor, the correlation integral will correspond to the integral in Eqn. (2.6). The relationship between  $H_q(X)$  and  $C_q(X; \epsilon)$  for  $\epsilon$  small is

$$H_q(X) \simeq -\ln C_q(X; \epsilon) + m \ln \epsilon. \quad (3.4)$$

This shows that estimated correlation integrals provide nonparametric estimates of  $H_q(X)$ . 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

$$I_q(X, Y) = H_q(X) + H_q(Y) - H_q(X, Y). \quad (3.5)$$

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

$$\widehat{I}_q(X, Y) = \ln \widehat{C}_q(X, Y; \epsilon) - \ln \widehat{C}_q(X; \epsilon) - \ln \widehat{C}_q(Y; \epsilon). \quad (3.6)$$

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:

$$\widehat{I}_q(X, Y|Z) = \ln \widehat{C}_q(X, Y, Z; \epsilon) - \ln \widehat{C}_q(X, Z; \epsilon) - \ln \widehat{C}_q(Y, Z; \epsilon) + \ln \widehat{C}_q(Z; \epsilon). \quad (3.7)$$

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  $I_q(X, Y|Z)$  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  $I_2(X, Y|Z)$  is zero or negative. If  $I_2(X, Y|Z)$  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  $I_2(X, Y|Z)$  is either positive or negative. This suggests that a one-sided test, rejecting for  $I_2(X, Y|Z; \epsilon)$  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.

## 4. Testing for Serial Independence

In this section we describe our approach to testing for serial independence in a time series setting. Let  $\{X_t\}_{t=1}^T$  be an observed time series of length  $T$ . We test the following null hypothesis:

$$H_0 : \quad X_t \text{ is i.i.d.}$$

our test statistic is based on the conditional mutual information defined above. By doing so the test is designed to have power against alternatives with conditional dependence, which has the advantage that the p-values obtained at different orders provide information about the lag structure of the time series.

We define delay vectors as

$$\mathbf{X}_t^m = (X_t, \dots, X_{t-m+1})', \quad (4.1)$$

where the prime denotes the transposed. The number of elements  $m$  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

$$I(X_t, X_{t-m} | \mathbf{X}_{t-1}^{m-1}) = -H(\mathbf{X}_{t-1}^{m+1}) + 2H(\mathbf{X}_t^m) - H(\mathbf{X}_{t-1}^{m-1}). \quad (4.2)$$

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  $X_{t-\ell}$  will become zero for  $\ell > k$  and positive for  $\ell \leq k$ . 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

$$I(X_t, \mathbf{X}_{t-1}^m) - I(X_t, \mathbf{X}_{t-1}^{m-1}) = I(X_t, X_{t-m} | \mathbf{X}_{t-1}^{m-1}), \quad (4.3)$$

the conditional mutual information. This demonstrates that the conditional mutual information quantifies the average amount of extra information that  $X_{t-m}$  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  $\mathbf{X}_{t-1}^{m-1}$ ,  $I(X_t, X_{t-m} | \mathbf{X}_{t-1}^{m-1}) = 0$ . If, on the other hand,  $X_{t-m}$  does contain extra information on  $X_t$ , we expect  $I(X_t, X_{t-m} | \mathbf{X}_{t-1}^{m-1}) > 0$ . We thus propose to perform a one-sided test based on  $I(X_t, X_{t-m} | \mathbf{X}_{t-1}^{m-1})$ , 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

$$\widehat{I}(X_t, X_{t-m} | \mathbf{X}_{t-1}^{m-1}) = -2 \ln \widehat{C}_m(\epsilon) + \ln \widehat{C}_{m+1}(\epsilon) + \ln \widehat{C}_{m-1}(\epsilon). \quad (4.4)$$

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

$$C_m(\epsilon) = \int \int I_{(\|s-t\| \leq \epsilon)} f_{\mathbf{X}_m}(s) f_{\mathbf{X}_m}(t) ds dt. \quad (4.5)$$

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

$$\widehat{C}_m(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N I_{(\|\mathbf{x}_i^m - \mathbf{x}_j^m\| \leq \epsilon)}. \quad (4.6)$$

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,

$$\widehat{\delta}_m(\epsilon) = 1 - \exp(-\widehat{I}(X_t, X_{t-m} | (X_{t-1}, \dots, X_{t-m+1}))) = 1 - \frac{[\widehat{C}_m(\epsilon)]^2}{\widehat{C}_{m-1}(\epsilon) \widehat{C}_{m+1}(\epsilon)}, \quad (4.7)$$

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

$$T^{\frac{1}{2}} \widehat{\delta}_m(\epsilon) \xrightarrow{d} N(0, V_{\delta_m}) \quad (4.8)$$



where the asymptotic variance is given by

$$V_{\delta_m} = 4 \left\{ \left( \frac{K_1(\epsilon)}{C_1(\epsilon)} \right)^{m-1} \left[ \left( \frac{K_1(\epsilon)}{C_1(\epsilon)} \right)^2 - 1 \right] \right\}^2, \quad (4.9)$$

and  $\widehat{C}_1(\epsilon)$  indicates the correlation integral at embedding dimension 1 while  $K_1(\epsilon)$  is estimated by

$$\widehat{K}_1(\epsilon) = \frac{2}{N(N-1)(N-2)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{k=j+1}^N I_{(|x_i-x_j|\leq\epsilon)} I_{(|x_j-x_k|\leq\epsilon)}. \quad (4.10)$$

It is known that the normal approximation based on the asymptotic distribution does not always perform well for small 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.

We will implement the test for independence as a Monte Carlo test. The Monte Carlo approach was first suggested by Barnard (1963) in the context of testing a simple null hypothesis. The idea is to construct the null distribution of the test statistic by calculating the test statistic for a large number of independent realisations generated by the null model. One can derive the significance of an empirically observed value of the test statistic using the fact that, under the null hypothesis, the test statistic for the original data and the artificial data are independent draws from the null distribution. In cases where one wishes to test a composite null hypothesis this procedure cannot be applied directly, since the true null process still depends on unknown model parameters. It was shown by Besag and Diggle (1977) and Engen and Lillegård (1997) that one can still obtain an exact level Monte Carlo test for composite hypotheses by conditioning on a minimal set of sufficient statistics under the null hypothesis. The approach is very flexible in that it allows for testing null hypotheses which are completely unspecified apart from some properties, as the symmetry tests proposed by Diks and Tong (1999) show. Under the IID null the order statistics provide a minimal and sufficient statistic. Under the null, and conditionally on the order statistic, each permutation of the observed data is equally probable, so that conditioning on the order statistic leads to a permutation test. Because a permutation test can be implemented easily, it offers a convenient way for obtaining an exact test.

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

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

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

In the bootstrap literature it is often emphasized that one should consider test statistics which are, at least asymptotically, pivotal under the null hypothesis, that is, their distribution should not depend on any unknown parameters under the null, see e.g. Beran (1988). The Monte Carlo approach, by conditioning on a minimal sufficient statistic, automatically satisfies this requirement, even for finite sample sizes. The reason is that after conditioning on a minimal and sufficient statistic under the null, the null distribution of any statistic, by construction, does not depend on any unknown parameters.

## 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 compare 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  $\Sigma_m$  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

$$\mathcal{H}_q(\mathbf{X}_t^m) = \frac{m}{2} \ln(2\pi) + \frac{1}{2} \ln |\Sigma_m| + \frac{m}{2} \frac{\log(q)}{(q+1)}, \quad (5.1)$$

which depends from the order  $q$ . The mutual information and conditional mutual information for linear Gaussian processes become

$$\mathcal{I}_q(X_t, X_{t-1}^m) = \frac{1}{2} \ln \left( \frac{|\Sigma_1| |\Sigma_m|}{|\Sigma_{m+1}|} \right), \quad (5.2)$$

and

$$\mathcal{I}_q(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1}) = \frac{1}{2} \ln \left( \frac{|\Sigma_m|^2}{|\Sigma_{m+1}| |\Sigma_{m-1}|} \right) \quad (5.3)$$

respectively, which are both independent of the Renyi order  $q$ . It follows that  $\delta_m$  for a linear Gaussian random process behaves as  $\delta_m^{\text{lin}}$  given by

$$\delta_m^{\text{lin}} = 1 - \exp(-I(X_t, X_{t-m} \mid \mathbf{X}_{t-1}^{m-1})) = 1 - \sqrt{\frac{|\Sigma_{m-1}| |\Sigma_{m+1}|}{|\Sigma_m|^2}}. \quad (5.4)$$

Because  $\Sigma_m$  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  $\delta_m^{\text{lin}}$  as

$$\delta_m^{\text{lin}} = 1 - \frac{l_{m+1}}{l_m}. \quad (5.5)$$

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

$$\hat{\mu}_m(\epsilon) = \hat{\delta}_m(\epsilon) - \hat{\delta}_m^{\text{lin}}(\epsilon) = \frac{\hat{l}_{m+1}}{\hat{l}_m} - \frac{[\hat{C}_m(\epsilon)]^2}{\hat{C}_{m-1}(\epsilon)\hat{C}_{m+1}(\epsilon)}, \quad (5.6)$$

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

In the case of testing for linearity it is less straightforward to set up a Monte Carlo test than for the independence test, and instead we set up a parametric bootstrap procedure to approximate the null distribution of the test statistic.

The test is composed of the following steps:

1. Calculate  $\hat{\mu}_m(\epsilon)$  for the time series  $\{X_t\}_{t=1}^T$ .
2. Estimate an AR( $d$ ) model for  $d = 1, \dots, d^{\text{max}}$  and choose the optimal order  $\hat{d}$  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

$$\tilde{x}_t = \sum_{i=1}^{\hat{d}} \hat{\beta}_i \tilde{x}_{t-i} + \epsilon_t \quad (5.7)$$

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

4. Calculate  $\tilde{\mu}_m(\epsilon)$  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

$$\hat{p} = \frac{1 + \# [\tilde{\mu}_m(\epsilon) \geq \hat{\mu}_m(\epsilon)]}{1 + B}$$

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

## 6. Simulations

This section describes the results obtained for the tests in simulations. Because the time series are scaled to unit sample variance prior to analysis, the values quoted for the bandwidth parameter  $\epsilon$  can be thought of as being expressed in terms of the number of standard deviations of the time series.

### 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  $\delta$  statistic. Recall 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  $\epsilon$  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. Also it can be observed that for small  $\epsilon$  the asymptotic approximation is poor. In that case, the correlation integrals are determined by a small number of distances so that the assumption of normality is no longer realistic. For similar reasons, increasing  $m$  also leads to a poor approximation of the asymptotic theory. These results clearly demonstrate the overall poor performance of the asymptotic test for small sample sizes.

$\epsilon$	$T = 100$				$T = 200$				$T = 500$			
	$m = 1$	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:** Size of the asymptotic test for  $\hat{\delta}_m(\epsilon)$  at the 5% nominal level for i.i.d. Gaussian noise, for various values of the bandwidth parameter  $\epsilon$ , time series length  $T$ , and embedding dimension (order)  $m$ .

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 AR(1) process and the Asymmetric Tent Map (ATM) have the same autocorrelation structure but the second process is a chaotic map. For the Nonlinear AR (NLAR) model we will consider dependence in different lags and investigate the behavior of the statistic in identifying the order. The remaining processes are nonlinear stochastic models with zero autocorrelation at all lags. For these models the application of autocorrelation based tests would fail to detect any dependence. Because of the curse of dimensionality, 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.

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	$y_t = 0.6u_{t-1}y_{t-2} + u_t$
NLAR(k)	$y_t =  y_{t-k} ^{0.8} + u_t$
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:** Time series models used to investigate the power of the tests. The innovations  $u_t$  are drawn independently from the standard normal distribution.

The results are shown in Table 3. 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. Notice 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 result for the NLAR model confirm the ability of the test to have very high power to detect dependences. When the true order is 1 the test has power higher than 0.90 on the first lag for the range of  $\epsilon$  considered. Instead when the true order is 3 the bandwidth becomes relevant for the power of the test: for  $\epsilon = 0.5$  it is 0.38 and increases to 0.89 for  $\epsilon = 1$ . For sample size of 200 the power increases significantly. The lower power of NLAR(3) in comparison with NLAR(1) for  $T = 100$  can be explained as the result of the curse of dimensionality: estimating multivariate densities for small bandwidth affects the statistical precision of the test statistics.

The test has also power against the (first order) TAR model: for  $\epsilon = 1.0$  the rejection rate is 0.62 for sample size 100 and 0.89 for 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 behavior 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  $\delta$  test is more suitable for obtaining insights into the lag dependence structure than the BDS test.

## 6..2 Test for Linearity

We first show the size properties of the test in Table 5. For  $\epsilon$  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  $\epsilon$ . 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 6 shows the power of the test for linearity for time series lengths  $T = 100$  and  $T = 200$  and for different bandwidth values. We also show the power of the  $V23$  test for linearity proposed by Teräsvirta *et al.* (1993). This test

has power against a wide range of alternative specifications of the conditional mean of the process. The order used in the  $V23$  test is the true order of the simulated model.

The test has unit power against the alternative of ATM. Even though it has the same linear structure of an AR model, the underlying chaotic dynamics is well addressed by the test. Similar results are obtained by the  $V23$  test. 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. Comparing the power with the parametric test, it turns out that for smaller sample size the curse of dimensionality affects the performance of the test whereas for bigger sample size the tests have similar power. For the NLAR process the power is higher for  $\epsilon = 1$  and increases with the sample size. The order is correctly detected for the 2 models considered. For  $\epsilon = 1$  and  $N = 100$  the power for the first order model is 0.29 and for the third order is 0.18. Also for the linearity test occurs the decay in power for the model with dependence in the third lag. In comparison with the independence test the power has significantly decreased. This is because the test for linearity considers only nonlinear dependence whereas the independence test was rejecting also for the presence of linear structure. The comparison with the  $V23$  test shows that it generally performs slightly worse than ours, particularly for the NLAR(1). The power is 0.57 on the first lag for TAR and increases to 0.89 for the bigger sample size. Also for this model our test has higher power than the  $V23$  test in the smaller sample and similar for  $T = 200$ . Very high power is also present for the ARCH(1) model. The test also has power on various lags for the GARCH(1,1) model. For the ARCH and GARCH models no results are quoted for the  $V23$  test, since this test is not consistent against dependence in higher moments.

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  $\epsilon \approx 1.0$ . In addition, the comparison with the  $V23$  test suggests that our test in most cases performs at least as well. The additional advantage of the nonparametric linearity test is that it can also capture dependence occurring in higher moments.

## 7. Empirical Applications

There have been many investigations concerning the presence of nonlinear dynamics in real US GNP data. In a TAR framework, Tiao and Tsay (1994) proposed a 4 regimes model involving the first 2 lags and Potter (1995) a 2 regimes model using lags 1,2 and 5. We apply the tests proposed here to the time series of log-differences of quarterly seasonally adjusted real GNP from the first quarter of 1947 to the last quarter of 2000. The use of seasonally adjusted data is common practice, and might introduce nonlinearities in the data. Investigating the effect of seasonal adjustment on the tests proposed here are beyond

the purpose of this article. In total we have 216 observations and we calculate the  $\delta$  statistic up to lag 5 to avoid the curse of dimensionality.

The results are summarized in Table 7. For  $\epsilon = 0.5$  the null of independence is rejected for the first 2 lags. The test for linearity instead rejects only for the second lag and this can be interpreted as an indication that the first order dependence was mainly of linear nature. Higher bandwidths confirm the rejection of the null hypothesis for the first 2 lags but also the third and fourth lags show rejections for both tests. These conclusions are partly in accordance with the results of Tiao and Tsay and Potter. However, there is also evidence that taking only the first 2 lags may lead to some structure left in the time series. The reason can be an unexplained dynamics in the conditional mean or in higher moments. As shown in Table 3 and 6 rejection of the null can arise also because of heteroskedastic structure such as for GARCH models where the power is spread in the first 4 lags.

## 8. Conclusion

In this paper we propose information theoretic bootstrap tests for independence and linearity. The results of the simulation study show that the tests have good power properties at moderate sample sizes, when compared to the BDS test and the V23 test. In addition they provide insights into the lag dependence in the data generating process. The power of both nonparametric tests typically increases when larger bandwidth values  $\epsilon$  are taken. However, care should be taken to avoid “leakage” of power to other lags as a result of taking the bandwidth  $\epsilon$  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. We thank an anonymous referee for suggestions, which substantially improved the manuscript.



## References

- Aparicio, F. M., and A. Escribano Aparicio and Escribano (1998): "Information-theoretic analysis of serial dependence and cointegration," *Studies in Nonlinear Dynamics and Econometrics*, **3**, 119–140.
- Auestad, B., and D. Tjøstheim Auestad and Tjøstheim (1990): "Identification of nonlinear time series: First order characterization and order determination," *Biometrika*, **77**, 669–687.
- Barnard, G. A. Barnard (1963): "Discussion of a paper by M. S. Bartlett," *Journal of the Royal Statistical Society B*, **25**, 294.
- Beran, R. Beran (1988): "Prepivoting test statistics: a bootstrap view of asymptotic refinements," *Journal of the American Statistical Association*, **83**, 687–697.
- Besag, J., and P. J. Diggle Besag and Diggle (1977): "Simple Monte Carlo tests for spatial pattern," *Applied Statistics*, **26**, 327–33.
- Blum, J. R., J. Kiefer, and M. Rosenblatt Blum *et al.* (1961): "Distribution free tests of independence based on the sample distribution function," *Annals of Mathematical Statistics*, **32**, 485–98.
- Brock, W. A., W. D. Dechert, J. A. Scheinkman, and B. LeBaron Brock *et al.* (1996): "A test for independence based on the correlation dimension," *Econometric Reviews*, **15**, 197–235.
- Chan, N. H., and L. T. Tran Chan and Tran (1992): "Nonparametric tests for serial dependence," *Journal of Time Series Analysis*, **13**, 19–28.
- Cheng, B., and H. Tong Cheng and Tong (1991): "On consistent nonparametric order determination and chaos," *Journal of the Royal Statistical Society B*, **54**, 427–449.
- Delgado, M. A. Delgado (1996): "Testing serial independence using the sample distribution function," *Journal of Time Series Analysis*, **17**, 271–285.
- Diks, C. and H. Tong Diks and Tong (1999): "A test for symmetries of multivariate probability distributions," *Biometrika*, **86**, 605–614.
- Engen, S. and M. Lillegård Engen and Lillegård (1997): "Stochastic simulations conditioned on sufficient statistics," *Biometrika*, **84**, 235–40.
- Granger, C. W. and J. L. Lin Granger and Lin (1994): "Using the mutual information coefficient to identify lags in nonlinear models," *Journal of Time Series Analysis*, **15**, 371–384.
- Hjellvik, V. and D. Tjøstheim Hjellvik and Tjøstheim (1995): "Nonparametric tests of linearity for time series," *Biometrika*, **82**, 351–368.

- Hjellvik, V., Q. Yao, and D. Tjøstheim Hjellvik *et al.* (1998): "Linearity testing using polynomial approximation," *Journal of Statistical Planning and Inference*, **68**, 295–321.
- Kullback, S. Kullback (1959): *Information Theory and Statistics*. New York: Wiley & Sons.
- Potter, S. Potter (1995): "A nonlinear approach to US GNP," *Journal of Applied Econometrics*, **10**, 109.
- Prichard, D. and J. Theiler, Prichard and Theiler (1995): "Generalized redundancies for time series analysis," *Physica D*, **84**, 476–493.
- Robinson, P. M. Robinson (1991): "Consistent nonparametric entropy-based testing," *Review of Economic Studies*, **58**, 437–453.
- Savit, R. and M. Green Savit and Green (1991): "Time series and dependent variables" *Physica D*, **50**, 95–116.
- Shannon, C. E. Shannon (1948): "A mathematical theory of communication" *Bell Systems Technical Journal*, **27**, 379–423, 623–656.
- Skaug, H. J. and D. Tjøstheim Skaug and Tjøstheim (1993a): "A nonparametric test of serial independence based on the empirical distribution function," *Biometrika*, **80**, 591–602.
- Skaug, H. J. and D. Tjøstheim Skaug and Tjøstheim (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.
- Teräsvirta, T., c-F. Lin, and C.W.J. Granger Teräsvirta *et al.* (1993): "Power of the neural network linearity test," *Journal of Time Series Analysis*, **14**, 209–220.
- Tiao, G.C. and R.S. Tsay, Tiao and Tsay (1994): "Some advances in non-linear and adaptive modelling in time-series," *Journal of Forecasting*, **13**, 109.
- Tschernig, R. and L. Yang, Tschernig and Yang (2000): "Nonparametric lag selection for time series" *Journal of Time Series Analysis*, **21**, 457–487.
- Wiener, N. Wiener (1948): *Cybernetics*. New York: John Wiley & Sons.
- Wu, K., R. Savit, and W. Brock, Wu *et al.* (1993): "Statistical tests for deterministic effects in broad band time series," *Physica D*, **69**, 172–188.

Model	$\epsilon$	$T = 100$				$T = 200$			
		$m = 1$	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
NLAR(1)	0.5	0.92	0.02	0.02	0.01	1.00	0.03	0.03	0.01
	1.0	0.97	0.03	0.02	0.02	1.00	0.03	0.03	0.02
	1.5	0.96	0.04	0.02	0.02	1.00	0.07	0.03	0.02
NLAR(3)	0.5	0.07	0.09	0.38	0.01	0.06	0.09	0.81	0.01
	1.0	0.07	0.10	0.89	0.02	0.06	0.10	1.00	0.02
	1.5	0.06	0.10	0.93	0.02	0.06	0.08	1.00	0.02
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:** Power of the independence test at the 5% nominal level for models in Table 2.

Model	$m = 1$	2	3	4
AR(1)	0.97	0.96	0.94	0.92
ATM	1.00	1.00	1.00	1.00
BILINEAR	0.37	0.60	0.62	0.61
NLAR(1)	0.97	0.94	0.91	0.87
NLAR(3)	0.07	0.10	0.41	0.55
TAR	0.62	0.60	0.53	0.46
ARCH	0.86	0.82	0.77	0.71
GARCH	0.64	0.70	0.73	0.75

**Table 4:** Power of the BDS test at the 5% nominal level for models in Table 2 ( $T = 100$ ,  $\epsilon = 1.0$ ).

$\epsilon$	$T = 100$				$T = 200$			
	$m = 1$	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:** Size of the linearity test at the 5% nominal level for an AR(1) with parameter equal to 0.6.

Model	$\epsilon$	$T = 100$				$T = 200$			
		$m = 1$	2	3	4	1	2	3	4
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
	V23	1.00				1.00			
BILINEAR	0.5	0.24	0.21	0.04	0.03	0.53	0.55	0.04	0.02
	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
	V23	0.73				0.89			
NLAR(1)	0.5	0.26	0.03	0.04	0.03	0.47	0.03	0.05	0.05
	1.0	0.29	0.05	0.03	0.04	0.48	0.04	0.05	0.05
	1.5	0.11	0.06	0.04	0.04	0.17	0.06	0.06	0.05
	V23	0.13				0.32			
NLAR(3)	0.5	0.04	0.05	0.11	0.05	0.03	0.06	0.19	0.04
	1.0	0.05	0.05	0.18	0.06	0.03	0.05	0.31	0.05
	1.5	0.06	0.04	0.11	0.06	0.04	0.04	0.12	0.05
	V23	0.15				0.26			
TAR	0.5	0.57	0.03	0.02	0.02	0.89	0.04	0.03	0.02
	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
	V23	0.49				0.80			
ARCH	0.5	0.73	0.04	0.02	0.01	0.96	0.06	0.01	0.01
	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
GARCH	0.5	0.50	0.19	0.05	0.02	0.83	0.42	0.10	0.02
	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:** Power of the linearity test at the 5% nominal level for models in Table 2.

$\epsilon$	$m = 1$	2	3	4	5
	Independence Test				
0.5	<b>0.02</b>	<b>0.02</b>	0.19	0.15	0.49
1.0	<b>0.01</b>	<b>0.02</b>	<b>0.04</b>	<b>0.02</b>	0.11
1.5	<b>0.01</b>	<b>0.05</b>	0.16	<b>0.02</b>	0.15
	Linearity Test				
0.5	0.18	<b>0.02</b>	0.23	0.22	0.42
1.0	<b>0.02</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	0.13
1.5	<b>0.03</b>	<b>0.04</b>	0.12	<b>0.01</b>	0.13

**Table 7:** *p*-values of the independence and linearity test for the quarterly growth rate of real US GNP ( $T = 216$ ).