Studies in Nonlinear Dynamics & Econometrics

Volume 12, *Issue* 1 2008 *Article* 2

NONLINEAR DYNAMICAL METHODS AND TIME SERIES ANALYSIS

Rank-based Entropy Tests for Serial Independence

Cees Diks* Valentyn Panchenko†

*CeNDEF, University of Amsterdam, C.G.H.Diks@uva.nl

†School of Economics, University of New South Wales, V.Panchenko@unsw.edu.au

Rank-based Entropy Tests for Serial Independence*

Cees Diks and Valentyn Panchenko

Abstract

In nonparametric tests for serial independence the marginal distribution of the data acts as an infinite dimensional nuisance parameter. The decomposition of joint distributions in terms of a copula density and marginal densities shows that in general empirical marginals carry no information on dependence. It follows that the order of ranks is sufficient for inference, which motivates transforming the data to a pre-specified marginal distribution prior to testing. As a test statistic we use an estimator of the marginal redundancy. We numerically study the finite sample properties of the tests obtained when the data are transformed to uniform as well as normal marginals. For comparison purposes we also derive a rank-based test against local ARCH alternatives. The performance of the new tests is compared with a modified version of the BDS test and with the Ljung-Box test.

*The authors wish to thank two anonymous referees for their constructive comments. Research supported by the Netherlands Organization for Scientific Research (NWO) under a MaGW-Vernieuwingsimpuls grant.

1 Introduction

Specifying a parametric family for time series data can be risky when testing for serial independence. Consider, for instance, a family of processes where at least one of the parameters affects both the dependence structure and the marginals (for instance the ARCH parameter in an ARCH(1) model). Because maximum likelihood estimators are typically not invariant under monotonically increasing transformations of the data, misspecification of the marginal distribution may affect estimated parameter values, and hence lead to false conclusions concerning dependence in the data.

To illustrate this, we generated a large number of sequences of independent and identically distributed (i.i.d.) data from a mixture of two normals with zero mean: $X_t = \sigma_t \varepsilon_t$, where $\{\varepsilon_t\}$ is a sequence of independent standard normal variables, while $\sigma_t = 0.5$ with probability 0.9, and $\sigma_t = 4.0$ with probability 0.1. For each of these simulated time series we tested whether the ARCH parameter was significantly different from zero using a standard generalized likelihood ratio test for the restriction $\theta = 0$ in the ARCH model: $X_t = \sqrt{h_t}\varepsilon_t$, $h_t = \gamma + \theta X_{t-1}^2$, $\varepsilon_t \sim N(0,1)$. Repeating this test for $1,000$ sequences at a nominal size of 0.05 for sample size $n = 100$ gave rise to a rejection rate of 0.150, while the rejection rate for i.i.d. normal data of the same sample size was only 0.013. The test being conservative, using a size-corrected critical value gave an even higher rejection rate (0.185). This clearly demonstrates that a parametric test for serial independence can over-reject when the marginals are misspecified.

Outside a parametric framework, marginals are generally uninformative about the dependence structure within a time series. For instance, for a strictly stationary time series process $\{X_t\}$, $t \in \mathbb{Z}$, consider the m-variate distribution of the delay vector $\boldsymbol{X}_t^m := (X_{t-m+1}, \ldots, X_t)'$. Assuming that it has a continuous distribution, its joint probability density function (pdf) can be decomposed into a product of the copula density c_m and the marginal pdf's:

$$f_{X_{t-m+1}}(x_1, \ldots, x_m) = c_m(F_X(x_1), \ldots, F_X(x_m)) f_X(x_1) \times \cdots \times f_X(x_m).$$

The null hypothesis of serial independence, which is often written as the factorization of the joint pdf into a product of marginals, H_0: $f_{X_{t-m+1}}(x_1, \ldots, x_m) = f_X(x_1) \times \cdots \times f_X(x_m)$, for $m = 1, 2, \ldots$, can alternatively be phrased in terms of the copula density only, i.e. without reference to marginals: H_0: $c_m(u_1, \ldots, u_m) = 1$, for $u_i \in [0,1]$, $i = 1, \ldots, m$.

Since the null hypothesis does not impose any structure on the marginals, one might actually ignore marginal information altogether when doing inference. A straightforward way of doing this would be to transform the observed time series $\{X_t\}_{t=1}^n$ to uniform marginals prior to applying a test for serial independence,

e.g. via the empirical cumulative distribution function (CDF) $\widehat{F}_X(X_t) = \#\{s \in \{1,\ldots,n\}|X_s \leq X_t\}/n$, or, as we prefer, a slight modification of that:

$$\widehat{U}_t = \frac{\#\{s \in \{1,\ldots,n\}|X_s \leq X_t\}}{n+1}.$$

The convention to divide by $n+1$ rather than n avoids complications with the largest observation when applying the inverse normal CDF to these uniformized observations. By construction, transforming the data to a canonical marginal distribution, all information in the marginals is ignored. Of course, compared to the parametric case this may lead to inefficiencies, since often information from marginals improves the accuracy of parameter estimators. However, the risks associated with possible model misspecification will be avoided.

In the present context it is natural to study the empirical distribution of m-histories $(\widehat{U}_{t-m+1},\ldots,\widehat{U}_t)$, $t = m,\ldots,n$, which is known as the empirical copula of $(X_{t-m+1},\ldots,X_t)$. Indeed, all statistics that we will consider are functions of the empirical copula of the data at hand. In some cases, such as for the ARCH-copula test discussed below, the values $\widehat{U}_t$ enter the statistics via the inverse normal CDF Φ, i.e. through $\Phi^{-1}(\widehat{U}_t)$.

The idea of transforming to ranks dates back to Spearman's (1904) rank correlation. For an overview of various rank-based statistical methods, see the book by Lehmann and D'Abrera (1998). Also Hallin and co-authors (1985, 1988) have worked on rank-based tests, but with a focus on optimal tests against specific alternatives (such as ARMA). Here we consider a nonparametric test that is perhaps not optimal against a specific alternative, but is consistent against any unspecified fixed alternative. We perform simulations to establish the power. To get an idea of the performance relative to a parametric test designed against a specific alternative, we compare the power with that of a parametric copula-based test against ARCH. The proposed method shares the information theoretical test statistics with Diks and Manzan (2002), but differs in the approach and motivation. Diks and Manzan (2002) aim at determining the order of a time series process, which involves conditional independence rather than independence. Here the aim is to construct nonparametric tests for serial independence with power against a wide range of alternatives.

After introducing the test statistic and the permutation procedure for establishing the statistical significance of the test statistic in section 2, we discuss the bandwidth selection problem in section 3.2 and propose a multiple bandwidth procedure. Section 4 describes the parametric ARCH-copula test. Section 5 presents the simulated size and power properties of the new nonparametric and parametric ARCH-copula test for a number of processes, and compares the results with those obtained with some existing tests. Section 6 concludes.

2 Test statistic

Consider a sample from a strictly stationary time series process: $\{X_t\}_{t=1}^n$. The null hypothesis is:

$$H_0 : \qquad X_t \text{ consists of i.i.d. observations}$$

and we wish to test this against general alternatives with dependence of order $p = m - 1$. Under the null hypothesis the elements of the delay vector $\boldsymbol{X}_t^m := (X_{t-m+1}, \ldots, X_t)$ are independent for each so-called 'embedding dimension' $m > 1$.

As motivated above, the test statistics under consideration are functions of the empirical probability integral transform $\{\widehat{U}_t\}$ of $\{X_t\}$. In general the values $\widehat{U}_s$ may enter the test statistic via the inverse CDF of another distribution with a given marginal CDF, G, say, through $Y_t = G^{-1}(\widehat{U}_t)$. We denote the associated marginal probability density function by $g_1(y)$, and the joint density of the m-dimensional delay vectors $\boldsymbol{Y}_t^m := (Y_{t-m+1}, \ldots, Y_t)$ by $g_m(\boldsymbol{y})$. Pompe (1993) has shown that the following relation holds in the case where $g_1(y)$ is the UNIF$(0, 1)$ density: density:

$$\int_{\mathbb{R}^m} g_m^2(\boldsymbol{y}^m)\,\mathrm{d}\boldsymbol{y}^m \geq \int_{\mathbb{R}^{m-1}} g_{m-1}^2(\boldsymbol{y}^{m-1})\,\mathrm{d}\boldsymbol{y}^{m-1} \int_{\mathbb{R}} g_1^2(y)\,\mathrm{d}y,$$

with equality if and only if $g_m(\boldsymbol{y}) = \prod_{i=1}^m g_1(y_i)$. Pompe stated this inequality in terms of an information theoretical notion called marginal redundancy: let $H_m = \ln \int_{\mathbb{R}^m} g_m^2(\boldsymbol{y}_m)\,\mathrm{d}\boldsymbol{y}_m \geq 0$, then the marginal redundancy of order m is defined as $R_m(Y_{t-m+1}, \ldots, Y_{t-1}; Y_t) = H_m - H_{m-1} - H_1 \geq 0$. It is a measure for the amount of information that $Y_{t-m+1}, \ldots, Y_{t-1}$ contain about Y_t. In terms of the redundancy, the claim is

$$R_m = H_m - H_{m-1} - H_1 \geq 0$$

with equality if and only if $g_m(\boldsymbol{y}) = \prod_{i=1}^m g_1(y_i)$, as long as $\{Y_t\}$ has a uniform marginal distribution. For other marginals this inequality need not hold, and the redundancy may become negative. Although this inequality can be used for constructing a consistent test, this does not guarantee that the test has highest power for particular alternatives. Therefore, in the simulation study we also consider transformations to other than uniform marginals by applying the transformation $Y_t = G^{-1}(\widehat{U}_t)$ prior to estimating the marginal redundancy. In the worst case this leads to a loss of power, but since we implement all tests as permutation tests, there is no risk of an increased type I error. In the simulation study below we estimate the redundancy after transforming to UNIF$(0, 1)$ as well as standard normal marginals, rejecting the null hypothesis if the estimated redundancy is too large. Although there might be exotic examples of joint distributions with normal marginals for which the redundancy would be negative, in practice we find that also in the non-uniform case this one-sided test performs better than a two-sided one.

We use plug-in kernel density estimators for evaluating the quantities H_m. The density estimator evaluated at $\boldsymbol{Y}_i^m = (Y_{i-m+1}, \ldots, Y_i)$ is

$$\begin{aligned} \widehat{g}_m(\boldsymbol{Y}_i^m) &= \frac{1}{n-m+1}\sum_{j=m}^{n} K_h(\boldsymbol{Y}_i - \boldsymbol{Y}_j) \\ &= \frac{1}{n-m+1}\sum_{j=m}^{n}\prod_{s=1}^{m} \kappa_h(Y_{i+1-s} - Y_{j+1-s}), \end{aligned} \tag{1}$$

where the last equality expresses the fact that we assume that the kernel factorizes, i.e. satisfies $K_h(\boldsymbol{z}) = \prod_{i=1}^m \kappa_h(z^i)$, where $\boldsymbol{z}$ is the vector $(z^1, \ldots, z^m)'$, and $\kappa_h(\cdot)$ a one-dimensional pdf with scale parameter h, so that $\kappa_h(s) = h^{-1}\kappa_1(s/h)$. Although the results presented in this paper are based on the Gaussian kernel function $\kappa_h(s) = (2\pi)^{-\frac{1}{2}}h^{-1}\exp\left(-s^2/(2h^2)\right)$, the theory can be developed for general probability kernels, i.e. kernels that are non-negative and normalized such that $\int_{\mathbb{R}} \kappa_h(z)\,\mathrm{d}z = 1$. Note that this normalization differs from that used in the chaos literature, where $\kappa_h(0)$ is usually normalized to 1. This is purely a matter of convention and is done here to emphasize the close connection between correlation integrals and kernel density estimation.

Anticipating the expressions that will arise when the nonparametric density estimators from (1) are plugged into the definition of R_m, we define the m-dimensional correlation integral associated with the kernel K_h as

$$\begin{aligned} C_m(h) &= \int\!\!\int K_h(\boldsymbol{v}-\boldsymbol{w})\, g_m(\boldsymbol{v}) g_m(\boldsymbol{w}) \mathrm{d}\boldsymbol{v}\,\mathrm{d}\boldsymbol{w} \\ &= E[K_h(\boldsymbol{V}-\boldsymbol{W})], \quad \boldsymbol{V} \text{ and } \boldsymbol{W} \sim \boldsymbol{Y}_t^m, \text{ independent.} \end{aligned}$$

In chaos theory the behavior of these correlation integrals with m and h is used to characterize the distribution of the delay vectors $\boldsymbol{Y}_t^m$. Estimators of correlation integrals appear naturally when we consider the nonparametric estimation of $\int_{\mathbb{R}^m} g_m^2(\boldsymbol{y}^m)\,\mathrm{d}\boldsymbol{y}^m = E[g_m(\boldsymbol{Y}_t^m)]$. A straightforward nonparametric estimator of this expectation would just be the sample average of the local density estimates (1) over all possible observed delay vectors $\widehat{Y}_t^m$. This leads to a V-statistic estimator of the correlation integral of $\boldsymbol{Y}_t^m$, based on the kernel function K_h:

$$\begin{aligned} \tilde{C}_m(h) &:= \frac{1}{n-m+1}\sum_{i=m}^{n} \widehat{g}_m(\boldsymbol{Y}_i^m) \\ &= \frac{1}{(n-m+1)^2}\sum_{i=m}^{n}\sum_{j=m}^{n} K_h(\boldsymbol{Y}_i^m - \boldsymbol{Y}_j^m). \end{aligned}$$

The theory developed for U- and V-statistics for weakly dependent processes (see Denker and Keller, 1983) applies, and, among other results, implies consistency of the estimator for bounded kernels under strong mixing: $\tilde{C}_m(h) \xrightarrow{P} C_m(h)$.

For the purpose of inference we may replace the V-statistic estimator by the asymptotically equivalent corresponding U-statistic:

$$\widehat{C}_m(h) = \binom{n-m+1}{2}^{-1} \sum_{i=m}^{n} \sum_{j=m-1}^{n} K_h(\boldsymbol{Y}_i^m - \boldsymbol{Y}_j^m).$$

We thus find the following estimator for the marginal redundancy of order m,

$$\widehat{R}_m = \widehat{H}_m - \widehat{H}_{m-1} - \widehat{H}_1 = \ln \widehat{C}_m(h) - \ln \widehat{C}_{m-1}(h) - \ln \widehat{C}_1(h),$$

where $\widehat{C}_m(h)$ is the estimated correlation integral for embedding dimension m.

3 Monte Carlo tests

We employ a Monte Carlo approach to obtain a p-value for the observed marginal redundancy. Before describing this method in detail, we briefly discuss the possibility of an alternative implementation of the test, using asymptotic theory.

The theory of U-statistics for weakly dependent processes, developed by Denker and Keller (1983, 1986), shows that under strict stationarity and suitable mixing conditions the fixed bandwidth vector of sample correlation integrals $\boldsymbol{\theta}_n := (\widehat{C}_m(h), \widehat{C}_{m-1}(h), \widehat{C}_1(h))'$ is asymptotically multivariate normally distributed with asymptotic mean $\boldsymbol{\theta} = (C_m(h), C_{m-1}(h), C_1(h))'$, and a certain asymptotic covariance $\Sigma_m(h)$, i.e. $\sqrt{n}\,(\boldsymbol{\theta}_n - \boldsymbol{\theta}) \xrightarrow{d} N(0, \Sigma_m(h))$. The functional delta method can then be employed to show that $\widehat{R}_m$ is asymptotically normal with asymptotic mean $\ln C_m(h) - \ln C_{m-1}(h) - \ln C_1(h)$ and asymptotic variance σ_m^2/n, where σ_m^2 can be consistently estimated from the data. Although these asymptotic results are elegant and potentially useful in many contexts, they are limited by the fact that they assume a fixed bandwidth value. For testing purposes, one would ideally like to be able to rely on asymptotic theory which allows for asymptotically optimal bandwidth rates. This would require determining the optimal asymptotic bandwidth rate, as well as a derivation of the limiting distributions of the test statistic for this rate. The fact that the test statistic $\widehat{R}_m$ can be interpreted as a marginal redundancy calculated by replacing the true density by a 'plug-in' kernel density estimate is convenient for showing consistency of $\widehat{R}_m$ under bandwidth rates for which the underlying density estimators are consistent. However, because of the aggregate nature of correlation integrals, the optimal rate need not be the same as that for nonparametric density estimation. Because of the variance-reducing effect of aggregating local density estimates, we expect that undersmoothing relative to optimal density estimation rates is appropriate. Besides the optimal bandwidth rate, applications also require a (possibly data-driven) prescription for choosing the proportionality constant involved.

Although the development of the relevant asymptotics can provide important insights, as well as a test procedure that is computationally less demanding, we here develop an alternative approach based on the concept of Monte Carlo testing, which has the advantage of providing a test with a type I error rate that exactly equals the nominal size for any finite sample size.

The problem of testing for serial independence is highly suitable for the Monte Carlo approach, allowing one to perform simulation-based tests with a type I error rate exactly equal to the nominal size. Consider the time series $\{\widehat{U}_t\}_{t=1}^n$. The elements $\widehat{U}_t$ take each value $\frac{k}{n}$, $k = 1, \ldots, n$ once. That is, the vector consists of a permutation of $\frac{1}{n}, \ldots, \frac{n}{n} = 1$. Under the null hypothesis of serial independence, each permutation of these values is equally likely. Therefore the null distribution of the test statistic $\widehat{R}$ can be simulated exactly by calculating the test statistic for many different random permutations. Since in this particular case the sample marginal of $\{\widehat{U}_i\}_{i=1}^n$ is fixed, critical values could be determined once and for all by a single large simulation, which need not be repeated for each new realization of the time series. However, in practice the null distribution would depend on details such as the time series length n, the bandwidth h, and the particular estimator used (either a V-statistic or a U-statistic). For this reason it is actually more practical to calculate exact p-values 'on the fly' for each specific case. Given today's computational power of PC's this poses no problem even for time series consisting of thousands of observations.

3.1 Single bandwidth permutation test

Although we have implemented the multiple bandwidth permutation test detailed below, it is instructive to consider how exact p-values would be obtained in a Monte Carlo test for a single bandwidth. We denote the test statistic $\widehat{R}_m$ calculated using the original data by $\widehat{R}_m^0$. Next the data are permuted randomly $B-1$ times, and for each permutation a 'bootstrap' version of the test statistic, $\widehat{R}_m^i$, $i = 1, \ldots, B-1$, is calculated. The significance of the originally observed value of the test statistic can be determined by observing that if the original data were generated under the null hypothesis of serial independence, the B values of the test statistics $\widehat{R}_m^0, \ldots, \widehat{R}_m^{B-1}$ are exchangeable. The one-sided p-value of the originally observed value of the test statistic is

$$\widehat{p} = \frac{\sum_{i=0}^{B} I(\widehat{R}_m^i > \widehat{R}_m^0) + L}{B+1}, \tag{2}$$

where $I(\cdot)$ denotes the indicator function taking the value 1 if the condition in brackets is true and 0 otherwise. This p-value is exact in the sense that under the null, it is uniformly distributed on $\frac{1}{B+1}, \frac{2}{B+1}, \ldots, 1$, provided that ties (cases where the test

statistic for permuted data and the original data coincide) are dealt with appropriately.

Let $Z = \sum_{i=0}^{B} I(\widehat{R}_m^i = \widehat{R}_m^0) \geq 1$ denote the number of ties plus one. In case $Z = 1$, $L = 1$, while for $Z > 1$, for L we take a random variable, uniformly distributed on $1, \ldots, Z$. That is, each rank of $\widehat{R}_m^0$ among the $\widehat{R}_m^i$ that happen to be equal to $\widehat{R}_m^0$, is taken to be equally probable. This is equivalent to adding a very small amount of noise to each of the $\widehat{R}_m^i$'s before determining their ranks, thus making the rank of $\widehat{R}_m^0$ among the $\widehat{R}_m^i$ unique. If $0 < \alpha = k/(B+1) < 1$ for some integer k, rejecting whenever $\widehat{p} \leq \alpha$ yields an exact level-α test. Generally, the power of a permutation test decreases if the number of permutations B decreases. Marriott (1979) has shown that that the efficiency loss is small for $B + 1 \geq 5/\alpha$. For instance, for a test with size 5%, it suffices to take $B + 1 = 100$.

Notice that the last term, $\ln \widehat{C}_1(h)$, in the marginal redundancy estimator $\widehat{R}_m$ is in fact a function of the empirical marginal distribution, which is invariant under permutations. This implies that this term is not important for inference concerning the dependence structure, which in turn is reflecting the fact that empirical marginals carry no information on dependence. One might therefore just as well decide to leave this term out of the test statistic when performing the permutation test, as that would have no effect whatsoever on the obtained p-value. It should be noted that the resulting difference $\ln \widehat{C}_m(h) - \ln \widehat{C}_{m-1}(h)$ is closely related to the correlation entropy $K_m(h) := \ln C_{m-1}(h) - \ln C_m(h)$, which was introduced in chaos theory for characterizing predictability of X_t on the basis of $\boldsymbol{X}_{t-1}^{m-1}$. Since larger values of the correlation entropy are associated with smaller predictability, our test for serial independence can be thought of as checking whether the entropy is sufficiently high (and hence the predictability sufficiently small) to be compatible with i.i.d. data with a given marginal distribution. For uniform marginals the maximum possible entropy is achieved exactly under the null hypothesis of serial independence.

3.2 Multiple bandwidth permutation test

The nonparametric test implemented as described above requires a choice for the bandwidth parameter h. Since the null distribution of the test statistic can be obtained easily by simulation, the actual size of the test for any bandwidth can be kept at the nominal value. Therefore, the most important criterion for the bandwidth choice would be the power. In general the bandwidth giving the largest power may depend on the process at hand (the particular alternative under consideration), and on the sample size. In fact one may be able to derive how the optimal bandwidth should scale with the sample size, as done by Diks and Panchenko (2007) for a

nonparametric test for Granger non-causality. However, this would still involve the estimation of the optimal factor of proportionality for the particular process at hand. As an alternative approach, we therefore choose to follow the multiple bandwidth approach developed by Diks and Panchenko (2006) based on Horowitz and Spokoiny (2001).

The idea behind the multiple bandwidth method is that instead of choosing a single bandwidth for testing, one might do better by combining information provided by the test statistics at different bandwidths. The multiple bandwidth method allows one to determine a single overall p-value on the basis of the test statistics at a (possibly large) number of bandwidths specified in advance. The multiple bandwidth method involves estimating a joint vector of p-values, $(\widehat{p}(h_1), \ldots, \widehat{p}(h_d))$, for d different bandwidths $h_1, \ldots, h_d$. We typically choose the bandwidth values h_i equidistant on a logarithmic scale between $h_{\min}$ and $h_{\max}$:

$$h_i \;=\; h_{\max}(h_{\min}/h_{\max})^{\frac{d-i}{d-1}}, \qquad i = 1, \ldots, d. \tag{3}$$

The next step consists of choosing an overall test statistic $\widehat{T}$, which is a single number summarizing evidence for serial dependence from the different bandwidths. Here we take $\widehat{T} = \inf_{h \in H} \widehat{p}(h)$, where $H = \{h_1, \ldots, h_d\}$ is the set of bandwidths under consideration. Again, the empirical distribution could be obtained beforehand by simulation. However, this would then still heavily depend on the bandwidths of choice, so we typically use an on-the-fly simulation method to determine the overall p-value for the observed value of T. Exact p-values can be obtained as follows:

1. Calculate the vector $\widehat{\boldsymbol{R}}_m^0 = (\widehat{R}_m^0(h_1), \ldots, \widehat{R}_m^0(h_d))$ of test statistics for a range of bandwidths: $H = \{h_1, \ldots, h_d\}$. We typically take bandwidth values that are equidistant on a logarithmic scale, as specified in Eq. (3).

2. Randomly permute the data and calculate a bootstrap version of the vector of statistics, $\widehat{\boldsymbol{R}}_m^1 = (\widehat{R}_m^1(h_1), \ldots, \widehat{R}_m^1(h_d))$. Repeat this B times, to obtain a total of B bootstrap vectors $\widehat{\boldsymbol{R}}_m^i$ for $i = 1, \ldots, B$.

3. For each bandwidth h_j, transform $\widehat{R}_m^i(h_j)$ into a (single bandwidth) p-value: $\widehat{p}_i(h_j) = [\sum_{k=0}^{B} I(\widehat{R}_m^k(h_j) > \widehat{R}_m^i(h_j)) + L_j]/(B+1)$, with L_j defined analogously to L in Eq. (2).

4. For each series, select the smallest p-value among all bandwidths and call it $\widehat{T}_i$: $\widehat{T}_i = \inf_{h \in H} \widehat{p}_i(h)$.

5. Calculate an overall p-value on the basis of the rank of $\widehat{T}_0$ among the $\widehat{T}_i$ ($i = 0, \ldots, B$), i.e. $\widehat{p} = [\sum_{i=0}^{B}(\widehat{T}_i < \widehat{T}_0) + L]/(B+1)$ using the ties randomization procedure as in Eq. (2).

In step 3 we pretend each of the permuted series to be the originally observed series and determine the corresponding p-values $\widehat{p}_i(h_j)$ that would have been obtained for series i for each of the different bandwidths h_j. In step 4, for each series the smallest p-value over the different bandwidths is selected (denoted by $\widehat{T}_i$, $i = 0, \ldots, B$). We finally use the exchangeability of the $B+1$ series (the original plus B permuted series) under the null to calculate an overall p-value by establishing the significance of $\widehat{T}_0$ for the actually observed data (step 5). As in the single bandwidth case, if $\alpha = k/(B+1)$ for some k, rejecting the null hypothesis whenever $\widehat{p} \leq \alpha$ yields an exact level-α test.

The power of the multiple bandwidth procedure depends on the alternative under consideration, the range $[h_{\min}, h_{\max}]$, the number d of elements in the bandwidth set H and the number of permutations B. In particular, the range should be wide enough to contain the bandwidths that are most informative concerning the alternative, for a wide range of data generating processes (DGPs).

In order to determine the optimal range $[h_{\min}, h_{\max}]$, we investigate the dependence of the single bandwidth on the sample size n and the data generating process (DGP). All simulations are based on the Gaussian kernel function $\kappa_h(s) = (2\pi)^{-\frac{1}{2}} h^{-1} \exp(-s^2/(2h^2))$. We consider $d = 30$ different bandwidth values h_i ranging from 0.1 to 2.5, equidistant on a logarithmic scale (see Eq. 3). A detailed description of the DGPs used, along with broader simulation results, are presented in section 5. To investigate the dependence of the optimal single bandwidth on n we considered a local ARCH(1) alternative of the form $Y_t \sim N(0, 1 + a_n Y_{t-1}^2)$ with $a_n = n^{-\frac{1}{2}}$. For brevity, here we only present simulations for the test based on the uniform transformation of marginals. Simulations with the normal transformation showed qualitatively similar results.

Figure 1 shows the power as a function of the bandwidth for time series of various lengths n for the ARCH(1) process described above (left panel, $m = 3$) and for various DGPs (right panel, $n = 100$, $m = 3$). The left panel reveals no clear relation between the bandwidth for which maximum power is obtained and the sample size n, although with increasing n one can observe that a near-optimal power is obtained for a wider range of bandwidth values. Based on the results of these simulations we decided to set $h_{\min} = 0.4$ and $h_{\max} = 2$ for the simulations presented below. Also the number of bandwidths d chosen in the range $[h_{\min}, h_{\max}]$ is important for the power. Taking d too small we risk losing informative bandwidths through the grid. Additional simulation results (not shown here) suggest that the empirical power of the multiple bandwidth procedure reduces as the bandwidth range becomes wider.

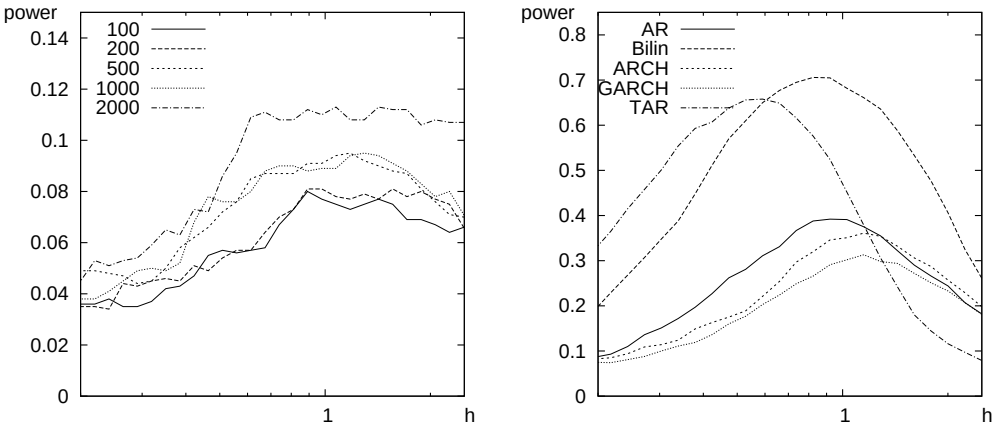

Figure 1: *Observed power as a function of bandwidth* h*. The left panel shows results for various series lengths* n*, for a sequence of local ARCH(1) alternatives converging to the null at rate* $n^{-1/2}$*; the right panel for various DGPs for* $n = 100$*. In all cases: dimension* $m = 3$*, nominal size* $\alpha = 0.05$*, number of permutations* $B + 1 = 100$ *and number of simulations* $1,000$*.*

In practice we suggest taking $d = 5$ different bandwidths, which is the value used in the simulations described below.

4 ARCH-copula test

To evaluate the power of the nonparametric test described above, it is desirable to have a benchmark test which is close to optimal (in the sense of Neyman and Pearson) against the alternatives considered. Since this would require the development of a different benchmark test for each alternative considered, we only consider the ARCH alternative here. As argued in the introduction, one cannot use a standard likelihood ratio test against ARCH dependence. If one wishes to ignore information from the marginals, only tests based on ranks are to be considered. We therefore derive an approximate likelihood ratio test of the uniform copula against the copula of an ARCH(1) process. Since the this copula depends on the ARCH coefficient, we only consider the likelihood ratio of the null likelihood and local alternatives in the direction of ARCH. This approach leads to a simple test with large simulated power against ARCH, which appears to serve as a decent parametric benchmark to compare our test with when applied to ARCH processes. In fact this local test against ARCH also turns out to have considerable power against many other types

of dependence, such as that present in the GARCH model, which is why the tables below also contain results of the ARCH copula test applied to processes other than ARCH.

In short, the resulting ARCH copula test consists of calculating

$$\tilde{T} = \sum_{i=2}^{n} (\Phi^{-1}(u_i))^2 (\Phi^{-1}(u_{i-1}))^2, \tag{4}$$

where Φ denotes the standard normal CDF, and to reject if $\tilde{T}$ is too large. Critical values can be obtained by a single extensive Monte Carlo run, simulating the values of $\tilde{T}$ under the i.i.d. null. The remaining part of this section is devoted to a derivation of this test.

We focus on testing the null of a uniform copula against local alternatives in the direction of the ARCH copula alternative. A standard ARCH(1) process $\{X_t\}$ is considered with normal innovations, i.e.

$$X_t = \sigma_t Z_t$$

where $\{Z_t\}$ is a sequence of independent standard normal random variables, and

$$\sigma_t^2 = \gamma + \theta X_{t-1}^2.$$

Without loss of generality we assume $\gamma = 1$, as it affects the scale of the process but not the distribution of the vector of ranks $S_n := ((n+1)\widehat{U}_1, \ldots, (n+1)\widehat{U}_n)$. Let $\ell(\theta; \boldsymbol{x})$ denote the log-likelihood of the ARCH(1) model with ARCH parameter θ, for a given sequence of observations $\boldsymbol{x} = (x_1, \ldots, x_n)$. The locally most powerful rank test (see e.g. Van der Vaart, 1998) for H_0: $\theta = 0$ against H_a: $\theta > 0$, rejects H_0 for large values of

$$E_0 \left(\dot{\ell}(0; \boldsymbol{x}) | S_n = s \right),$$

where the dot denotes derivation w.r.t. θ, $E_0(\cdot)$ is the expectation operator under the null, and the condition $S_n = s$ indicates that the expectation is conditional on the observed sequence of ranks, s.

The log-likelihood is given by

$$\ell(\theta; \boldsymbol{x}) = \log f_X(x_1) + \sum_{i=2}^{n} \log f_{X_t|X_{t-1}}(x_i|x_{i-1}).$$

The density in the sum is the conditional normal density from the ARCH(1) model. Although the first term can be calculated by determining the marginal density $f_X(x)$

up to order θ. The stationary density $f_X(x;\theta)$ must satisfy

$$\begin{aligned} f_X(x;\theta) &= \int f_X(z;\theta) f_{X_t|X_{t-1}}(x|z)\,\mathrm{d}z \\ &= \int f_X(z;\theta)(2\pi(1+\theta z^2))^{-\frac{1}{2}}\mathrm{e}^{-x^2/(2(1+\theta z^2))}\,\mathrm{d}z. \end{aligned}$$

A Taylor expansion of the transition probability density around $\theta = 0$ gives

$$\begin{aligned} f_X(x;\theta) &= (2\pi)^{-\frac{1}{2}}\mathrm{e}^{-x^2/2}\int f_X(z;\theta)\left[1+\frac{1}{2}\theta\left(x^2-z^2\right)\right]\mathrm{d}z + \mathcal{O}(\theta^2) \\ &= (2\pi)^{-\frac{1}{2}}e^{-z^2/2}\left[1+\frac{1}{2}\theta\left(x^2-1\right)\right] + \mathcal{O}(\theta^2), \end{aligned}$$

where we have used $E[Z^2] = \frac{1}{1-\theta}$, the unconditional variance of the ARCH process under consideration.

Upon taking the derivatives one finds that the locally most powerful rank test rejects for large values of

$$\begin{aligned} &\frac{1}{2}E_0\left[X_1^2 - \sum_{i=2}^{n}\left(X_{i-1}^2 + X_{i-1}^2X_i^2\right)|S_n = s\right] \\ &:= \frac{1}{2}E\left[(\Phi^{-1}(U_1))^2 - \sum_{i=2}^{n}\left((\Phi^{-1}(U_{i-1}))^2 + (\Phi^{-1}(U_{i-1}))^2(\Phi^{-1}(U_i))^2\right)|S_n = s\right], \end{aligned}$$

where the latter expectation is over an i.i.d. sequence of UNIF$(0,1)$ variables U_i, conditioned on having the same rank sequence as the originally observed time series. The calculation of the conditional expectation is cumbersome, but since $E[U_{k:n}|S_n = s] = k/(n+1) = \widehat{U}_{k:n}$, the above expectation is very close to the much more simple statistic

$$\frac{1}{2}(\Phi^{-1}(\widehat{U}_1))^2 - \frac{1}{2}\sum_{i=2}^{n}\left((\Phi^{-1}(\widehat{U}_{i-1}))^2 + (\Phi^{-1}(\widehat{U}_{i-1}))^2(\Phi^{-1}(\widehat{U}_i))^2\right).$$

Note that because the sample values of $\{\widehat{U}_t\}_{t=1}^n$ are fixed, apart from a small edge effect the quantity $(\Phi^{-1}(\widehat{U}_1))^2 - \sum_{i=2}^n(\Phi^{-1}(\widehat{U}_i))^2$ is constant under permutations of the data. Because also $\sum_{i=1}^n(\Phi^{-1}(\widehat{U}_i))^2$ is constant, there is an intuitively more clear formulation of the critical region in terms of the covariance: reject the null if the sample covariance $\mathrm{Cov}((\Phi^{-1}(\widehat{U}_{i-1}))^2, (\Phi^{-1}(\widehat{U}_i))^2)$ is too large. In practice it is more convenient to calculate $\tilde{T}$ as given in Eq. (4) and to reject the null hypothesis if $\tilde{T}$ is too large. Since the test depends on the data only through the sample distribution of $(\widehat{U}_{i-1}, \widehat{U}_i)$, the empirical copula, we refer to the test as the ARCH-copula test. Critical values for the test statistic and/or p-values for observed data can be obtained straightforwardly by simulation.

5 Simulation results

This section investigates the power of the proposed nonparametric tests (further referred to as the R tests) based on the uniform and normal marginal transformations and compares them with that of the parametric ARCH copula test, the BDS test and the Ljung-Box (1978) test. To allow for a better comparison, the BDS test was modified to allow for transformations to uniform or normal marginals prior to testing, and the multiple bandwidth procedure was also implemented for the BDS test.

5.1 Fixed alternatives

We compare the rejection rates of the tests against fixed alternatives for the following stationary DGPs, where $\{\varepsilon_t\}$ is an i.i.d. sequence of $N(0,1)$ random variables:

DGP 0. $Y_t = \varepsilon_t$

DGP 1. $Y_t = \sigma_t \varepsilon_t, \qquad \sigma_t = \begin{cases} 0.5 & \text{w.p. } 0.9, \\ 4.0 & \text{w.p. } 0.1 \end{cases}$

DGP 2. $Y_t = 0.3 Y_{t-1} + \varepsilon_t$

DGP 3. $Y_t = 0.6 \varepsilon_{t-1} Y_{t-2} + \varepsilon_t$

DGP 4. $Y_t = \sqrt{h_t}\varepsilon_t, \qquad h_t = 1 + 0.4 Y_{t-1}^2$

DGP 5. $Y_t = \sqrt{h_t}\varepsilon_t, \qquad h_t = 0.01 + 0.80 h_{t-1} + 0.15 Y_{t-1}^2$

DGP 6. $Y_t = \begin{cases} -0.5 Y_{t-1} + \varepsilon_t, & Y_{t-1} < 1 \\ 0.4 Y_{t-1} + \varepsilon_t, & \text{else} \end{cases}$

The above DGPs or slight modifications of these were previously considered by Diks and Panchenko (2006), Granger *et al.* (2004), Granger and Lin (1994), Hong and White (2005), Brock *et al.* (1996) and others. DGP 0 satisfies the null hypothesis and is included to assess the empirical size of the tests. DGP 1 is the i.i.d. Gaussian mixture model that served as a motivating example in the introduction. DGP 2 is a linear AR(1) process. DGP 3 is a bilinear process introduced by Granger and Andersen (1978). DGPs 4 and 5 are instances of ARCH(1) and GARCH$(1,1)$ processes proposed by Engle (1982) and Bollerslev (1986) respectively. The coefficients of the GARCH$(1,1)$ process are taken close to the corresponding estimates of Bollerslev (1986). DGP 6 is a TAR process proposed by Tong (1978). We used series of length $n = 100$, and the total number of permutations, including the original

DGP	R, multiple h		R at $\widehat{h}^*$		$\widehat{h}^*$	BDS		ARCH	Ljung
	unif	normal	unif	normal		unif	normal	copula	-Box
0. IIDN	0.05	0.05	0.05	0.05	-	0.04	0.05	0.05	0.09
1. MIXN	0.06	0.05	0.06	0.05	-	0.05	0.05	0.05	0.04
2. AR(1)	0.34	0.17	0.39	0.20	1.2	0.42	0.24	0.16	0.39
3. Bilin	0.67	0.67	0.70	0.69	0.8	0.53	0.56	0.47	0.11
4. ARCH	0.29	0.46	0.36	0.53	1.0	0.34	0.55	0.74	0.09
5. GARCH	0.27	0.40	0.31	0.43	1.1	0.25	0.37	0.38	0.12
6. TAR	0.57	0.13	0.66	0.19	0.5	0.69	0.17	0.18	0.09

Table 1: *Simulated rejection rates of the specified tests for various DGPs, nominal size* $\alpha = 0.05$*, series length* $n = 100$*, embedding dimension* $m = 3$ *for the R and BDS tests, number of permutations* $B + 1 = 100$ *and number of simulations* $1,000$*. The multiple bandwidth permutation method was used for both the R test and the BDS test. The columns denoted by '*R *at* $\widehat{h}^*$*' refer to the largest observed single bandwidth rejection rates of the* R *test over an extended grid of* 30 *bandwidths ranging from* 0.1 *to* 2.5*; the corresponding bandwidth* $\widehat{h}^*$ *(if unique) for the transformation to uniform marginals is reported in the column labeled* $\widehat{h}^*$*. The Ljung-Box test was based on the first* 30 *sample autocorrelations.*

series, was set to $B+1 = 100$. The bandwidth set H included $d = 5$ different values in the range $[0.4, 2.0]$ after normalizing the series to unit variance. We considered delay vector dimension $m = 3$. For comparison the BDS test was implemented as a one-sided permutation test, rejecting when the test statistic is too large. The bandwidth range for the BDS test was taken as $[0.5, 2.0]$, which roughly coincides with the typical values for the BDS test found in the literature. As for the R test, $d = 5$ bandwidths were used in this range, and the number of permutations was set to $B + 1 = 100$. All tests were conducted at a nominal size of $\alpha = 0.05$, and the number of simulations was set to $1,000$.

Table 1 reports the observed rejection rates at nominal size $\alpha = 0.05$ for the R test and the BDS test, both for uniform and normal marginal transformations, the parametric ARCH-copula test and the Ljung-Box test based on the first 30 sample autocorrelations. The actual size of all tests is close to the nominal size of 5%, except for the Ljung-Box test applied to i.i.d. normal data, which is slightly over-sized with an observed rejection rate of 9%. All Monte-Carlo methods perform well in terms of size for both i.i.d. processes. This was to be expected, since all these tests are exact, and insensitive to transformations of marginals. It can be observed that the Ljung-Box test only has substantial power for the AR(1) process, which has more linear dependence structure than the others. For all nonlinear

processes each of the nonparametric tests considered outperforms the Ljung-Box test. The nonparametric R test yields powers comparable to those obtained using the multiple bandwidth version of the BDS test. The BDS test performs better for the AR(1), ARCH and TAR processes (DGPs 2, 4 and 6), while the R test performs better for the bilinear process (DGP 3) and GARCH process (DGP 5). As expected the parametric ARCH copula test shows high power against the ARCH(1) alternative (DGP 4), but has relatively low power against the other alternatives. The uniform marginal transformation leads to higher power for AR(1) and TAR processes, DGPs 2 and 6 respectively, while the normal marginal transformation improves power for ARCH(1) and GARCH(1,1) processes, DGPs 4 and 5 respectively. These results hold for both the R test and the BDS test. The columns labeled 'R at $\widehat{h}^*$' report the largest observed single bandwidth rejection rates among 30 bandwidth values ranging from 0.1 to 2.5. The bandwidth value, $\widehat{h}^*$, for which the maximum occurred (if unique), is reported in the next column, only for the transformation to uniform marginals (the values found for the transformation to normal marginals were very similar). It can be observed that the multiple bandwidth procedure loses little in terms of power compared to the optimal single bandwidth. Although it cannot be excluded that data driven bandwidth selection methods may also achieve these optimum single bandwidth powers, or even improve on them, the results appear to suggest that the multiple bandwidth method leaves little room for improvement.

Based on the somewhat mixed simulation results presented here, one might conclude that there is no reason to prefer the R test over the BDS test. The sizes for both tests are correct and their powers are comparable, although one test may have more power against certain alternatives, and the other against other alternatives. Firstly, we would like to emphasize that, to make the comparison more reasonable, the results presented here under the label BDS test are for a modified version of the BDS test, which also employs the multiple bandwidth procedure. Secondly, we are not aware of a consistency result for the BDS test against any class of fixed alternatives. It follows from Pompe's inequality stated in the section 2, and the consistency of U-statistics, that the R test with uniform marginals is consistent against all fixed deviations from independence. Thirdly, the marginal redundancy is suited to picking up the conditional density of X_t given the $m-1$ previous observations $X_{t-m+1}, \ldots, X_{t-1}$. In doing so, it uses likelihood-based scores, and in that sense closely mimics locally optimal rank tests, but in a nonparametric fashion. This might be a possible explanation for the fact that the R test outperforms the BDS test in particular for DGPs 3 and 5, which exhibit dependence of order higher than 1.

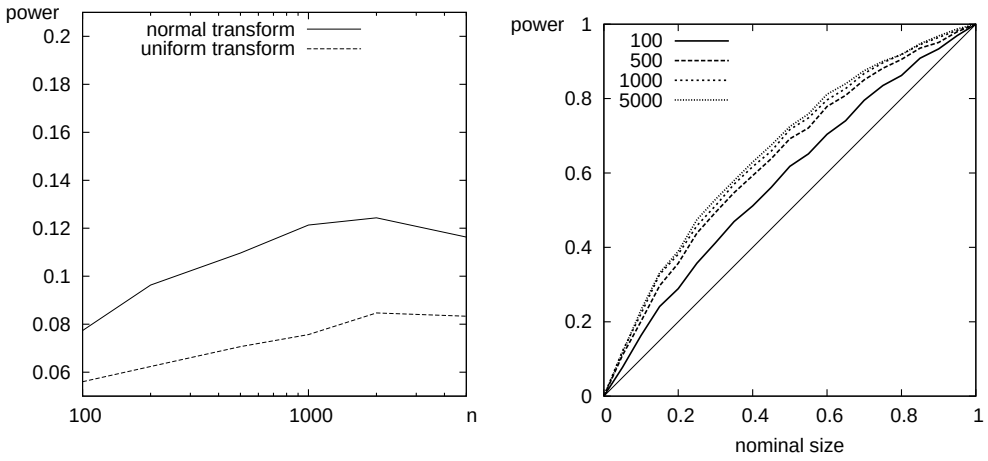

Figure 2: *Observed power against local ARCH(1) alternatives converging to the null at rate* $n^{-1/2}$*, as a function of sample size* $n = 100, \ldots, 5,000$ *at nominal size* $\alpha = 0.05$ *(left panel); as a function of nominal size for the test based on normal transformation (right panel). Embedding dimension* $m = 3$*, number of permutations* $B + 1 = 100$*, number of simulations* $3,000$*.*

5.2 Local alternatives

We next consider the power against local alternatives. The components of the test statistic for the R test are estimated using U-statistics, which in the non-degenerate case converge at the parametric rate $n^{-1/2}$. Therefore, we expect the test to have nontrivial asymptotic power at the rate $n^{-1/2}$ and illustrate this via simulations. Following Hong and White (2005) we consider a sequence of processes, the lag j dependence of which is described by the following joint pdf:

$$f_{jn}(y_t, y_{t+j}) = f(y_t)f(y_{t+j})[1 - a_n q_j(y_t, y_{t+j}) + r_{jn}(y_t, y_{t+j})], \tag{3}$$

where $q_j(y_t, y_{t+j})$ is a function characterizing the deviation from the null hypothesis, a_n governs the rate of convergence to the null as $n \to \infty$, and $r_{jn}(y_t, y_{t+j})$ is a higher order term obtained from the Taylor series expansion of $f_{jn}(y_t, y_{t+j})$ around the point $a_n = 0$. See Hong and White (2005) for assumptions on $q_j(\cdot, \cdot)$ and $r_{jn}(\cdot, \cdot)$ which ensure that $f_{jn}(\cdot, \cdot)$ is a proper density function.

The simulations are based on an ARCH(1) process $Y_t \sim N(0, 1 + a_n Y_{t-1}^2)$ with $a_n = n^{-1/2}$. The joint density of (Y_t, Y_{t+1}) can be represented in the form (3) with $q_j(y_t, y_{t+j}) = y_t y_{t+j}$. Figure 2 (left panel) shows the rejection rates (powers) of the considered test against a sequence of local alternatives which converges to the

null at the usual parametric rate $a_n = n^{-1/2}$, where $n = 100, \ldots, 5,000$. A horizontal line in the graph would indicate the parametric rate. After an initial transient period for small n, the curves level out, suggesting that both tests asymptotically approach the parametric rate. The observed small deviations from the horizontal line are due to estimation uncertainty, but are within the 5% error bounds. The nontrivial asymptotic power for the R test against this sequence of local alternatives can also be observed for other values of the nominal size, as illustrated by the power-size plots for the transformation to normal marginals for increasing sample sizes n shown in the right panel of Figure 2.

6 Summary

Model misspecification may lead to increased rejection rates for parametric tests for serial independence. This was shown in a simple ARCH example where the process was actually i.i.d. but with misspecified marginals. On the basis of this observation we explored the idea of full-heartedly deleting all information in marginals prior to testing by transforming to a pre-specified marginal distribution. In the case of a uniform marginal distribution the marginal redundancy was shown to be a promising measure of dependence, on which consistent tests against any alternative can be based. The bandwidth selection problem was addressed and it was argued why a multiple bandwidth procedure, using information from a number of different bandwidths simultaneously, was implemented. The resulting tests were compared to a modified version of the BDS test (implemented as a permutation test with multiple bandwidth procedure), a parametric ARCH copula test and the Ljung-Box test. Our simulations, which were carried out for relatively small sample sizes, showed that none of the rank-based tests was uniformly most powerful against all alternatives. Transforming to uniform marginals was found to be considerably better than to normal marginals for the nonparametric detection (either with the BDS test or with the R test) of deviations from independence against the AR(1) and the TAR processes, which are characterized by a strong dependence of the conditional mean on past realized values. For the processes considered, the BDS test performed better for first order processes, while the R test had more power against alternatives with higher order dependence.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31,** 307–327.

Brock, W., Dechert, W., Scheinkman, J. and LeBaron, B (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, **15,** 197–236.

Denker, M. and Keller, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **64,** 505–522.

Denker, M. and Keller, G. (1986). Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics*, **44,** 67–93.

Diks, C. and Manzan, S. (2002). Tests for serial independence and linearity based on correlation integrals. *Studies in Nonlinear Dynamics in Econometrics*, **6** (2)**,** article 2, 1–22.

Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, **30,** 1647–1669.

Diks, C. and Panchenko, V. (2007). Nonparametric tests for serial independence based on quadratic forms. *Statistica Sinica*, **17,** 81–97.

Engle, Robert (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50,** 987–1008.

Granger, C. W. and Andersen, A. P. (1978). *An introduction to bilinear time series models*. Göttingen: Vandenhoek and Ruprecht.

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

Granger, C. W., Maasoumi, E. and Racine, J. (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis*, **25,** 649–669.

Hallin, M., Ingenbeek, J.-F. and Puri, M. L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *Annals of Statistics*, **13,** 1156–1181.

Hallin, M. and Mélard, G. (1988). Rank-based tests for randomness against first-order serial dependence. *Journal of the American Statistical Association*, **83,** 1117–1128.

Hong, Y. and White, H. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica*, **73,** 837–901.

Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69,** 599–631.

Lehmann, E. L. and D'Abrera, H. J. M. (1998). *Nonparametrics: Statistical Methods Based on Ranks, rev. ed.* Prentice-Hall: Englewood Cliffs, NJ.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65,** 297–202.

Marriott, F. (1979). Barnard's Monte Carlo tests: How many simulations? *Applied Statistics*, **28,** 75–77.

Pompe, B. (1993). Measuring statistical dependences in time series. *Journal of Statistical Physics*, **73,** 587–610.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, **15,** 72–101.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*, *Amsterdam* (ed. C. H. Chen), pp. 101–141. Sijhoff and Noordhoff.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press.