

## Non parametric portmanteau tests for detecting non linearities in high dimensions

Jan G. De Gooijer & Ao Yuan

To cite this article: Jan G. De Gooijer & Ao Yuan (2016) Non parametric portmanteau tests for detecting non linearities in high dimensions, Communications in Statistics - Theory and Methods, 45:2, 385-399, DOI: [10.1080/03610926.2013.815209](https://doi.org/10.1080/03610926.2013.815209)

To link to this article: <https://doi.org/10.1080/03610926.2013.815209>



© 2016 The Author(s). Published with license by Taylor & Francis© Jan G. De Gooijer and Ao Yuan.



Published online: 13 Jan 2016.



Submit your article to this journal [↗](#)



Article views: 406



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



# Non parametric portmanteau tests for detecting non linearities in high dimensions

Jan G. De Gooijer<sup>a</sup> and Ao Yuan<sup>b</sup>

<sup>a</sup>Department of Quantitative Economics, University of Amsterdam, Amsterdam, The Netherlands; <sup>b</sup>Department of Biostatistics, Bioinformatics and Biomathematics Georgetown University, Washington DC, USA

## ABSTRACT

We propose two non parametric portmanteau test statistics for serial dependence in high dimensions using the correlation integral. One test depends on a cutoff threshold value, while the other test is freed of this dependence. Although these tests may each be viewed as variants of the classical Brock, Dechert, and Scheinkman (BDS) test statistic, they avoid some of the major weaknesses of this test. We establish consistency and asymptotic normality of both portmanteau tests. Using Monte Carlo simulations, we investigate the small sample properties of the tests for a variety of data generating processes with normally and uniformly distributed innovations. We show that asymptotic theory provides accurate inference in finite samples and for relatively high dimensions. This is followed by a power comparison with the BDS test, and with several rank-based extensions of the BDS tests that have recently been proposed in the literature. Two real data examples are provided to illustrate the use of the test procedure.

## ARTICLE HISTORY

Received 24 December 2012  
Accepted 7 June 2013

## KEYWORDS

Correlation integral; Non linear time series; Rank-based BDS test statistics; Second-order Rényi entropy.

## MATHEMATICS SUBJECT CLASSIFICATION

62G10; 62G20; 62M10

## 1. Introduction

Testing for serial independence has been extensively studied in the literature. Within the context of (non) linear time series analysis, a wide variety of tests have been proposed. Among them, are many procedures for testing first-order serial dependence ( $m = 2$  dimensions) based on functionals measuring differences between two estimated densities. Often this is coupled with a desire to avoid specific assumptions about the marginal distributions under the null hypothesis of serial independence. Not surprisingly, therefore, non parametric test statistics are predominant in the literature. Using kernel density estimators, research papers in this area are by Robinson (1991), Skaug and Tjøstheim (1993, 1996), Delgado (1996), and Hong and White (2005), among others. Also various tests of serial independence based on rank statistics have been proposed, including Hallin and Puri (1992), Ferguson et al. (2000), and Genest, Ghoudi, and Rémillard (2007) (henceforth GGR, 2007).

Since it is now generally believed that many empirical time series, while non linear, are generated by high-dimensional processes, it is natural to consider non parametric serial independence portmanteau tests for dimensions  $m > 2$ . Indeed, both Skaug and Tjøstheim (1993), and Hong and White (2005) proposed tests for multidimensional data using a linear combination of their original tests when  $m = 2$ . However, Delgado (1996) remarked that the

**CONTACT** Jan G. De Gooijer [j.g.degooijer@uva.nl](mailto:j.g.degooijer@uva.nl) Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands.

© 2016 Jan G. De Gooijer and Ao Yuan. Published with license by Taylor & Francis.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

asymptotic null distribution of these test statistics can be different from tests based on the joint density of  $m > 2$  consecutive time series observations.

Motivated by high-dimensional non linear deterministic and stochastic phenomena in economics, and using the correlation integral of Grassberger and Procaccia (1984) (see, e.g., Sec. 2), Brock et al. (1996) proposed a non parametric test for higher variate serial independence, often referred to as BDS test. The BDS test is free of nuisance parameters, but it suffers from some problems. For instance, the power of the BDS test depends on the choice of value of the dimensional distance parameter. Another problem is that the BDS test, though asymptotically normal under the null, has high rates of Type I error, especially for non Gaussian data. Moreover, the BDS test is inconsistent. In an attempt to mitigate some or all of these problems, GGR (2007) advocated a circular version of the BDS test and seven rank-based extensions of the BDS test statistic (see Sec. 4).

In this article, we address the above issues in a direct way. Rather than using the theory of rank statistics, we propose two non parametric portmanteau test statistics based on the correlation integral. The first test statistic depends on a cutoff threshold value. And the second test statistic is free of this dependence. Both tests can be viewed as variants of the traditional BDS test statistic. In contrast, our test statistics are consistent, and asymptotically normally distributed. Moreover, the finite-sample rejection probabilities of the tests do not differ too much from their asymptotic level, for both Gaussian and uniformly distributed data. Thus, the tests avoid some of the major weaknesses of the BDS test statistic. Further, they can be relatively easily computed using a simple adjustment of available computer code.

The article is organized as follows: Sec. 2 introduces the two non parametric portmanteau test statistics. The asymptotic properties of the tests are discussed in Sec. 3. In Sec. 4, we report the results of a Monte Carlo study, comparing the finite-sample performance of our tests with the BDS test, and with the eight rank-based BDS tests of GGR (2007). In Sec. 5, two empirical examples are given to illustrate the use of the test procedure, and in Sec. 6 we offer some concluding remarks. We relegate all technical arguments to an Appendix.

## 2. Test statistics

Let  $\{Y_t : t = 1, \dots, n\}$  be a sample from a real-valued, strictly stationary time series process with values in  $\mathbb{R}$ . Rather than focusing on a single time series in  $\mathbb{R}$ , we embed  $\{Y_t\}$  in an  $m$ -dimensional space  $\mathbb{R}^m$ , where  $m \in \mathbb{N}^+ \equiv \{2, 3, \dots\}$ . It can be constructed from  $\{Y_t\}$  as the set of random vectors,  $\{Z_t : t = 1, \dots, n - m + 1\}$  defined by  $Z_t = (Z_{t,1}, \dots, Z_{t,m}) = (Y_t, \dots, Y_{t+m-1})$ . Assume that  $\{Z_t\}$  admits a common continuous joint density function  $f_m(z_1, \dots, z_m)$  for  $Z = (z_1, \dots, z_m)$ . Denote the marginal density by  $f(z)$ . The problem under consideration is that of testing the null hypothesis  $\mathbb{H}_0$ :  $\{Y_t\}$  is independent and identically distributed (i.i.d.) versus the alternative hypothesis of serial  $m - 1$  dependence. Now, if  $\{Y_t\}$  is i.i.d. the joint density will be equal to the product of the individual marginals, and the hypothesis of interest is

$$\mathbb{H}_0 : f_m(z_1, \dots, z_m) = f(z_1) \cdots f(z_m), \text{ for all } (z_1, \dots, z_m) \in \mathbb{R}^m.$$

Any deviation from this equality is evidence of serial dependence.

Various functionals measuring the divergence from independence between densities have been proposed in the literature. It is natural to require that a measure of functional dependence has, at least, the following basic properties: non negativity, maximal information, and invariance under continuous monotone increasing transformations. A divergence measure

satisfying these three properties is the mutual information, defined by

$$I_m \equiv \int \ln \left( \frac{f_m(z_1, \dots, z_m)}{f(z_1) \cdots f(z_m)} \right) f_m(z_1, \dots, z_m) dz_1 \cdots dz_m, \tag{2.1}$$

where the integral is taken over the support of  $\{Z_t\}$ , and with the convention that  $0 \ln(0/0) \equiv 0$ . It is easy to see that (2.1) can be expressed in terms of the Shannon entropy, or the second-order Rényi entropy,  $H(Z) \equiv - \int f_m(z_1, \dots, z_m) \ln f_m(z_1, \dots, z_m) dz_1 \cdots dz_m$  as

$$R_m = \sum_{i=1}^m H(Z_i) - H(Z), \tag{2.2}$$

where  $H(Z_i) = - \int \ln f(z_i) \ln f(z_i) dz_i$  is the  $i$ th marginal version of  $H(Z)$ .

A serial dependence measure can be obtained by relating  $R_m$  to the second-order correlation integral, which in the i.i.d. case is defined as  $C(m, \delta) = \int \int \mathbf{1}(\|s - t\| \leq \delta) f_m(s) f_m(t) ds dt$  where  $\|\cdot\|$  denotes the maximum norm on  $\mathbb{R}^m$ ,  $\mathbf{1}(\cdot)$  is the indicator function, and  $\delta \equiv \delta_n > 0$  a cutoff threshold value or bandwidth. Pompe (1993) and Prichard and Theiler (1995) showed that, for  $\delta$  small,  $H(Z) \approx - \ln C(m, \delta)$ . In fact, putting  $R_m$  in terms of the correlation integral, Prichard and Theiler (1995) find:

$$R_{m,\delta} = \ln C(m, \delta) - \ln\{C(1, \delta)\}^m. \tag{2.3}$$

We see that (2.3) has a similar structure as the divergence measure  $C(m, \delta) - \{C(1, \delta)\}^m$  that forms the basis of the traditional BDS test statistic.

Note that  $C(m, \delta)$  is just the expectation of the kernel function, that is,  $E(\mathbf{1}(\|Z_i - Z_j\| \leq \delta))$  with  $Z_i \sim f_m$ , where ‘ $\sim$ ’ denotes equivalence in distribution. Hence, it can be estimated straightforwardly in a  $U$ -statistic framework by

$$C_{n,\delta}(m) = \binom{n - m + 1}{2}^{-1} \sum_{1 \leq i < j \leq n - m + 1} \mathbf{1}(\|Z_i - Z_j\| \leq \delta). \tag{2.4}$$

Given (2.3) and (2.4) the first non parametric portmanteau test statistic takes the form

$$\mathcal{I}_{n,\delta}(m) = \ln(C_{n,\delta}(m)/C_{n,\delta}^m(1)). \tag{2.5}$$

In principle, the naive kernel in (2.4) may be replaced by more sophisticated kernels. But fully kernel-based non parametric estimation of  $C(m, \delta)$  is usually unattractive because estimation accuracy decreases rapidly, for fixed sample sizes  $n$ , as  $m$  increases. Besides, issues linked with kernel estimation in high dimensions like bandwidth selection, will complicate a direct comparison between (2.5) and the test statistics proposed by GGR (2007). Also note that the test statistic  $\mathcal{I}_{n,\delta}(m)$  has an invariance property in the following sense. For any monotone increasing transformation  $Z'_i = g(Z_i)$  of the data,  $\mathbf{1}(\|Z_i - Z_j\| \leq \delta) = \mathbf{1}(\|Z'_i - Z'_j\| \leq \delta')$  with  $\delta' = g^{-1}(\delta)$ . So the null hypothesis of the original data corresponds to the null hypothesis after the transformation. Thus, after the transformation, by Theorem 1 in Sec. 3, the asymptotic null distribution of  $\mathcal{I}_{n,\delta}(m)$  is still normal only with parameter  $\delta$  changed to  $\delta'$ .

Note that the test statistic (2.5) depends on the choice of  $\delta$ . This may be viewed as a serious limitation in practice. A common way to get around this problem is to integrate out  $\delta$  with respect to some specified probability density function  $h(\cdot)$  on  $[a, b]$  ( $0 < a < b < \infty$ ). Thus, the second test statistic can be defined as follows

$$\tilde{\mathcal{I}}_n(m) = \int_a^b \mathcal{I}_{n,\delta}(m) h(\delta) d\delta, \tag{2.6}$$

where, without loss of generality, we assume  $\delta \in [a, b]$ . A specific choice of  $\delta$  is made later.

### 3. Asymptotic properties

Let  $C \equiv C(\delta) = P_{\mathbb{H}_0}(|Y_1 - Y_2| \leq \delta) = \int \int \mathbf{1}(|u - v| \leq \delta) dF(u) dF(v)$ ,  $\eta(y) = E_{\mathbb{H}_0}(\mathbf{1}(|y - Y_2| \leq \delta))$ ,  $\tau^2 = \text{Var}_{\mathbb{H}_0}(\eta(Y_1)) = K - C^2$ , with  $K \equiv K(\delta) = E\eta^2(Y_1) = \int [F(u + \delta) - F(u - \delta)]^2 dF(u)$ , and  $\gamma \equiv \gamma(\delta) = P_{\mathbb{H}_0}(|Y_1 - Y_2| \leq \delta, |Y_1 - Y_3| \leq \delta)$ . In addition,  $\xrightarrow{D}$  stands for convergence in distribution.

**Theorem 1.** Assume  $K > C^2$ , i.e.  $\tau^2 > 0$ . Then,

(i) For fixed  $\delta > 0$ , under  $\mathbb{H}_0$ , as  $n \rightarrow \infty$  we have

$$\sqrt{n}\mathcal{I}_{n,\delta}(m) \xrightarrow{D} \mathcal{N}(0, \sigma^2(m, \delta)),$$

where  $\sigma^2(m, \delta) = 4(\frac{\tau_m^2}{C^{2m}} - m^2 \frac{K-C^2}{C^2})$ ,  $\tau_m^2 = K^m - C^{2m} + 2 \sum_{i=1}^{m-1} (K^{m-i} C^{2i} - C^{2m})$ , and  $K \equiv \gamma$

(ii) Under  $\mathbb{H}_0$ , as  $n \rightarrow \infty$  first and then let  $\delta \rightarrow 0$ , we have

$$\sqrt{n}\mathcal{I}_{n,\delta}(m) \xrightarrow{D} \mathcal{N}(0, \sigma^2(m)),$$

where  $\sigma^2(m) = \lim_{\delta \rightarrow 0} \sigma^2(m, \delta)$ , assume exist.

For fixed  $\delta$ ,  $\sigma^2(m)$  can be consistently estimated by  $\sigma_n^2(m) \equiv \sigma_n^2(m, \delta)$ , which is  $\sigma^2(m)$  with  $C$  and  $K$  replaced by the following estimates

$$C_n := \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \mathbf{1}(|Y_i - Y_j| \leq \delta) \xrightarrow{a.s.} C,$$

$$K_n := \binom{n}{3}^{-1} \sum_{1 \leq i < j < k \leq n} \mathbf{1}(|Y_i - Y_j| \leq \delta, |Y_i - Y_k| \leq \delta) \xrightarrow{a.s.} K,$$

where the dependence on  $m$  has been suppressed for notational clarity. Similarly, using  $K_n$  and  $C_n$ , a consistent estimator for  $\tau_m^2$  can be obtained.

For fixed  $m$ , assume  $\delta$  varies in  $[a, b]$ . Let  $W(\cdot) = W(m, \cdot)$  denote a Gaussian process on  $[a, b]$ , that is, for each fixed  $\delta \in [a, b]$ ,  $W(\delta)$  is a Gaussian random variable with  $E(W(\delta)) = 0$ , and for each fixed  $\delta_1, \delta_2 \in [a, b]$ ,  $\text{Cov}(W(\delta_1), W(\delta_2)) = \sigma^2(m, \delta_1 \wedge \delta_2)$ , where  $\delta_1 \wedge \delta_2 = \min(\delta_1, \delta_2)$ , for all  $\delta_1, \delta_2 \in [a, b]$ . The following theorem states the asymptotic null distribution of  $\tilde{\mathcal{I}}_n(m)$ . It is also a uniform weak convergence result for  $\mathcal{I}_n(m, \cdot) := \mathcal{I}_{m,\cdot}(m)$  on the space  $C[a, b]$ , the space of all bounded continuous functions on  $[a, b]$ , equipped with the Skorohod metric.

**Theorem 2.** Under  $\mathbb{H}_0$ , completed with conditions of Theorem 1, we have

$$\sqrt{n}\tilde{\mathcal{I}}_n(m) \xrightarrow{D} \int_a^b W(\delta)h(\delta)d\delta \sim \mathcal{N}(0, V_h^2(m)),$$

where  $V_h^2(m) = \int_a^b \int_a^b \sigma^2(m, s \wedge t)h(s)h(t)dsdt$ .

### 4. Finite-sample performance

The purpose of this section is to investigate the finite-sample performance of (2.5) and (2.6) vis-à-vis the standard BDS test statistic and eight high-dimensional serial correlation test statistics proposed by GGR (2007). For ease of reference, these latter tests can be described as follows. The first two test statistics, denoted by  $\tilde{S}_{n,\epsilon}^*$  and  $\tilde{S}_{n,\epsilon}$ , are a circular version and a

rank-based BDS (RBDS) statistic, respectively. Both statistics depend on the distance parameter  $\epsilon$ , fixed at 0.3 by GGR (2007) and in the computations shown below.

To eliminate the finite-sample behavior of  $\tilde{S}_{n,\epsilon}^*$  and  $\tilde{S}_{n,\epsilon}$  from  $\epsilon$ , GGR (2007) introduced six additional RBDS test statistics. These statistics use three functionals, that is, direct integration, the Kolmogorov–Smirnov functional, and the Cramér-von Mises functional, and two different  $m$ -dimensional empirical processes; see GGR (2007) for a definition of these processes. Tests derived from the first empirical process are denoted by  $\tilde{I}_n^*$ ,  $\tilde{M}_n^*$ , and  $\tilde{T}_n^*$ , respectively. Tests based on the second empirical process are denoted by  $\tilde{I}_n$ ,  $\tilde{M}_n$ , and  $\tilde{T}_n$ , respectively. Under  $\mathbb{H}_0$ , the six RBDS test statistics converge in distribution to centered Gaussian variables.

For (2.5) and (2.6) all test results are based on the empirical distribution of 100 bootstrapped  $p$ -values, computed over 1,000 independent runs. Following GGR (2007), Monte Carlo results for the BDS and RBDS test statistics are based on 10,000 time series generated for each statistic. Simulation results showed that, in finite-samples, bootstrap approximations to the distribution of our two test statistics are much better than sample approximations in the right and left tail.

Although the asymptotic variance of  $\mathcal{I}_{n,\delta}(m)$  depends on  $\delta$ , minimizing the variance over  $\delta$  is intractable. Therefore, we simply set  $\delta_n = 2\hat{\sigma}_Y n^{-1/5}$ , where  $\hat{\sigma}_Y$  is an estimate of the standard deviation of  $\{Y_t\}$ , and we use Theorem 1 (ii) to evaluate the asymptotic result. For the evaluation of  $\tilde{\mathcal{I}}_n(m)$  we need to specify a probability density function  $h(\delta)$ . Since the range of optimal values of  $\delta$  can be wide, we set  $\delta_j = u\hat{\sigma}_Y$  ( $j = 1, \dots, M$ ) with  $u$  sampled from a  $U[0.5, 2.5]$  distribution. In addition, we approximate the integration in (2.6) by  $(1/M) \sum_{j=1}^M W(\delta_j)$ . All power results presented below are for  $M = 100$ . Similar results were obtained for  $M = 200$ .

#### 4.1. Fixed alternatives: Raw data

In this section, we consider the size and rejection rates of the two proposed test statistics against fixed alternatives for stationary data generating processes (DGPs) with known parameters. The DGPs are listed in Table 1. They are identical to the ones used by Hong and White (2005) and GGR (2007).

Table 2 reports bootstrapped empirical rejection rates (in percentages) of the tests  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  under DGP 0 for both i.i.d.  $\mathcal{N}(0, 1)$  and i.i.d.  $U(0, 1)$  distributions. The tests are reasonably well sized at all three (10%, 5%, and 1%) significance levels with rejection frequencies close to the nominal levels and for both sample sizes, and both distributions.

Note that both test statistics  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  behave well for both distributions. By contrast, a limitation of the BDS test statistic is that its rate of converge depends on the choice

**Table 1.** Data generating processes (DGPs) used for the Monte Carlo simulation study.

DGP	Name	Model specification
0	i.i.d.	$Y_t = \epsilon_t$
1	AR(1)	$Y_t = 0.3Y_{t-1} + \epsilon_t$
2	ARCH(1)	$Y_t = h_t^{1/2}\epsilon_t, h_t = 1 + 0.8Y_{t-1}^2$
3	Threshold GARCH(1,1)	$Y_t = h_t^{1/2}\epsilon_t$ , with $h_t = 0.25 + 0.6h_{t-1} + 0.5Y_{t-1}^2\mathbf{1}(\epsilon_t < 0) + 0.2Y_{t-1}^2\mathbf{1}(\epsilon_t \geq 0)$
4	Bilinear AR(1)	$Y_t = 0.8Y_{t-1}\epsilon_{t-1} + \epsilon_t$
5	Non linear MA(1)	$Y_t = 0.8\epsilon_{t-1}^2 + \epsilon_t$
6	Threshold AR(1)	$Y_t = 0.4Y_{t-1}\mathbf{1}(Y_{t-1} > 1) - 0.5Y_{t-1}\mathbf{1}(Y_{t-1} \leq 1) + \epsilon_t$
7	Fractional AR(1)	$Y_t = 0.8 Y_{t-1} ^{0.5} + \epsilon_t$
8	Sign AR(1)	$Y_t = \text{sign}(Y_{t-1}) + 0.43\epsilon_t$

**Table 2.** Bootstrap levels of the test statistics  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  under i.i.d.  $\mathcal{N}(0, 1)$  and under i.i.d.  $U(0, 1)$  for DGP 0.

m	n = 100						n = 200					
	$\mathcal{I}_{n,\delta}(m)$			$\tilde{\mathcal{I}}_n(m)$			$\mathcal{I}_{n,\delta}(m)$			$\tilde{\mathcal{I}}_n(m)$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
$\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$												
2	10.1	4.7	1.1	12.1	5.6	0.4	8.0	4.4	0.8	11.4	5.4	1.7
4	9.5	4.6	0.8	9.6	5.1	1.1	8.6	4.5	1.1	14.0	6.6	0.8
6	10.0	5.5	1.3	10.1	5.2	0.8	10.3	5.1	1.0	12.4	6.7	1.3
$\varepsilon_t \sim i.i.d. U(0, 1)$												
2	8.8	4.8	0.9	14.8	6.7	0.9	11.4	5.4	0.7	9.7	4.4	0.8
4	9.5	5.1	0.6	13.7	6.3	1.0	11.1	5.7	0.8	9.8	5.6	1.0
6	10.1	5.1	1.0	14.1	6.3	0.8	10.8	6.0	1.2	10.8	6.3	0.9

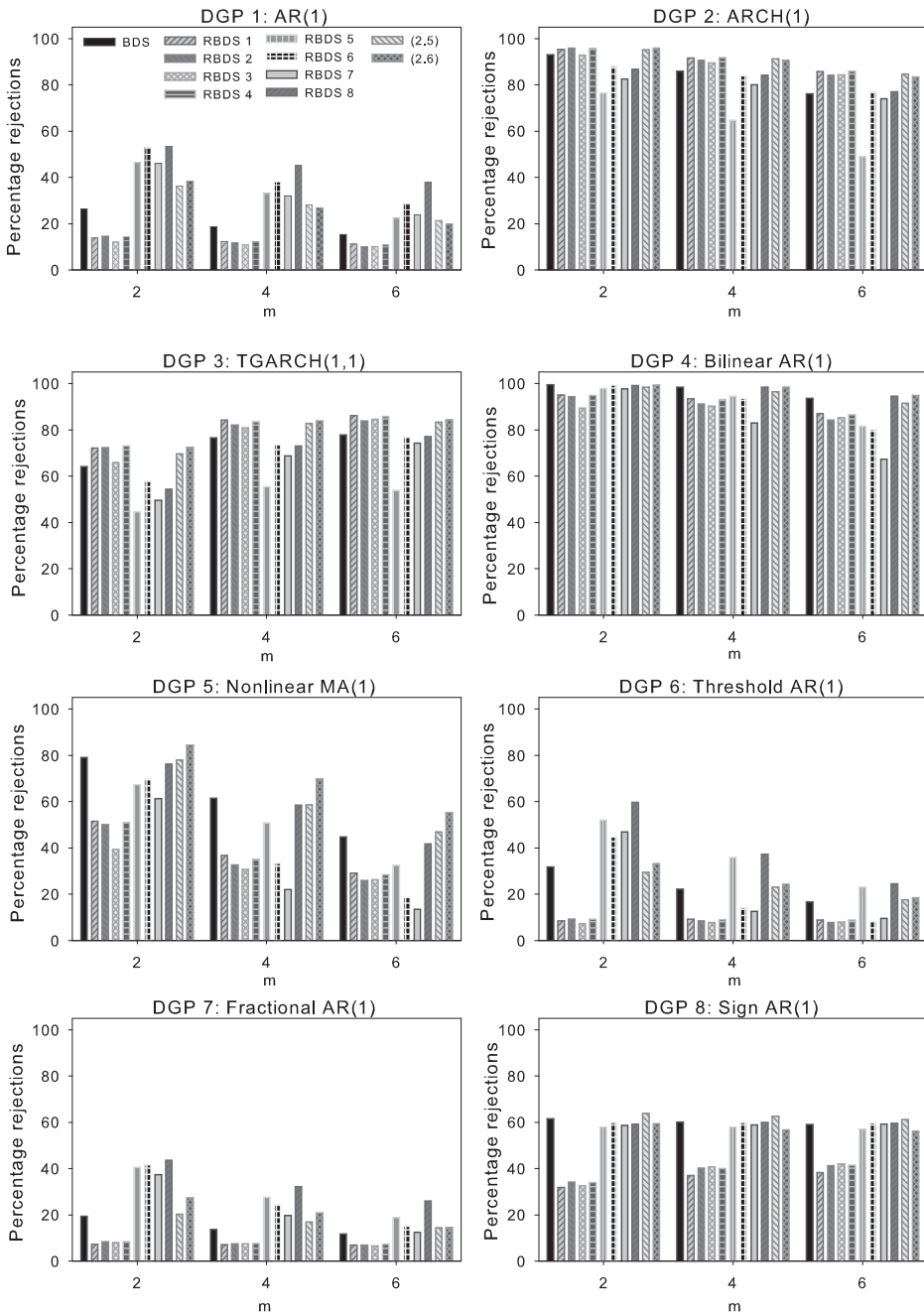
of  $\epsilon$ , the dimension  $m$ , and the shape of the distribution function. With the i.i.d.  $U(0, 1)$  distribution, for example, a nominal 5% test rejects between 8% and 44% of the time when  $n = 100$  and  $m = 2$ ; see, for example, Brock et al. (1991, Table A.5). No serious size distortions are reported by GGR (2007) for the RBDS test statistics.

Figure 1 is partly based on GRR (2007, Table 4). It shows percentages of the empirical rejection rates of eleven statistics under DGPs 1–8 with i.i.d.  $\mathcal{N}(0, 1)$  innovations, at the 5% nominal level, and for  $n = 100$  and  $m = 2, 4$ , and 6. Several points are noteworthy. First,  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  are equally powerful as the best performing RBDS test under DGPs 2, 3, 4, and 8, and more powerful under DGP 5 for all values of  $m$ . Second,  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  are less powerful than the best performing RBDS test under DGPs 1, 6, and 7. Third, across all test statistics the power remains fairly constant for all values of  $m$  under DGPs 2, 3, and 8. The test statistics  $\tilde{S}_{n,0.3}^*$ ,  $\tilde{I}_n^*$ ,  $\tilde{M}_n^*$ , and  $\tilde{T}_n^*$  (RBDS 1–4) also have constant, but very low power, for all values of  $m$  under DGPs 1, 6, and 7. For all other DGPs rejections decrease as  $m$  increases. Fourth, the test statistics  $\tilde{S}_{n,0.3}^*$ ,  $\tilde{I}_n^*$ ,  $\tilde{M}_n^*$ , and  $\tilde{T}_n^*$  (RBDS 1–4) are roughly equally powerful. In addition, we see a clear difference between the performance of the first set of four RBDS test statistics as opposed to the second set of four RBDS test statistics. Not surprisingly, under all DGPs,  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  are far more powerful than the Hong–White test statistic and the Skaug–Tjøstheim test statistic. Hence, these results have been omitted from the comparison.

Table 3 reports the empirical rejection rates under DGPs 1–8 with i.i.d.  $U(0, 1)$  innovations for  $n = 100$ . The simulations permit several observations. First, under DGP 8, almost all test statistics have rejection rates quite close to the nominal size of 5%. This is not surprising, since in this case the sign AR(1) process (DGP 8) reduces to a uniformly distributed white noise process. A clear exception is the BDS statistic with far too high rejection rates; see Brock et al. (1991) for similar Monte Carlo results.

Second, under DGPs 2 and 3,  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  are more powerful than almost all other tests, except for BDS. The power of our test statistics is considerably higher than the powers of  $\tilde{S}_{n,0.3}^*$ ,  $\tilde{I}_n^*$ ,  $\tilde{M}_n^*$ , and  $\tilde{T}_n^*$  under DGPs 1–5, and 7. Under DGP 4, BDS,  $\mathcal{I}_{n,\delta}(m)$ ,  $\tilde{\mathcal{I}}_n(m)$ ,  $\tilde{S}_{n,0.3}$ ,  $\tilde{I}_n$ ,  $\tilde{M}_n$ , and  $\tilde{T}_n$ , are equally powerful. Under DGPs 5 and 6,  $\tilde{S}_{n,0.3}$ ,  $\tilde{I}_n$ ,  $\tilde{M}_n$ , and  $\tilde{T}_n$ , slightly outperform  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$ . Third, the power of  $\tilde{\mathcal{I}}_n(m)$  almost always is higher than the power of  $\mathcal{I}_{n,\delta}(m)$ .

In summary, the suggested test statistics  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  remarkably outperform many of the (R)BDS test statistics under various DGPs when the innovations are i.i.d.  $\mathcal{N}(0, 1)$  or i.i.d.  $U(0, 1)$ , and when the sample size is relatively small, and for all values of  $m$ .



**Figure 1.** Percentage rejections of the BDS statistic, the RBDS test statistics,  $\mathcal{I}_{n,\delta}(m)$  (2.5) and  $\tilde{\mathcal{I}}_n(m)$  (2.6);  $n = 100$  and 5% nominal level with  $\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$ .

#### 4.2. Fixed alternatives: Residuals

In this section, we compare and contrast the size and rejection rates of eleven portmanteau test statistics against AR(2) processes as fixed alternatives, using residuals obtained from a fitted AR(1) process. Specifically, when testing for serial dependence, replacing the true



**Table 3.** Percentages rejections of eleven test statistics at the 5% nominal level under i.i.d.  $U(0, 1); n = 100$ .

DGP	m	BDS	Rank-based BDS test statistics									
			$\tilde{S}_{n,0.3}^*$	$\tilde{T}_n^*$	$\tilde{M}_n^*$	$\tilde{T}_n^*$	$\tilde{S}_{n,0.3}$	$\tilde{T}_n$	$\tilde{M}_n$	$\tilde{T}_n$	$\mathcal{I}_{n,\delta}(m)$	$\tilde{\mathcal{I}}_n(m)$
1	2	44.0	8.8	9.2	7.2	9.5	30.9	39.0	32.2	45.5	35.8	41.1
	4	36.4	7.7	7.0	6.5	7.7	20.5	25.9	21.7	34.9	21.5	27.9
	6	34.6	7.1	6.4	6.5	7.4	14.7	19.9	17.2	29.0	12.9	18.6
2	2	45.6	10.1	10.4	8.5	10.8	24.0	27.4	23.8	31.1	36.6	39.9
	4	36.6	10.2	9.5	8.9	10.4	15.8	16.7	14.0	22.9	21.7	28.6
	6	34.0	9.9	8.9	9.1	10.1	11.1	12.7	10.9	18.6	12.7	20.0
3	2	23.7	6.4	6.8	5.9	6.9	6.3	6.1	5.7	6.2	69.5	71.5
	4	28.8	7.0	6.5	6.2	7.1	5.8	6.3	5.6	6.5	77.7	83.0
	6	34.3	7.1	6.8	6.9	7.5	5.3	6.0	5.6	6.8	74.8	83.5
4	2	99.7	89.4	87.8	79.9	88.8	100.0	100.0	100.0	100.0	99.7	99.9
	4	99.6	72.7	67.7	65.6	71.1	100.0	100.0	99.9	100.0	99.2	99.7
	6	98.9	59.2	54.3	55.5	58.1	99.5	99.9	99.7	100.0	94.8	85.2
5	2	75.9	17.8	18.4	14.4	19.0	92.1	95.3	92.3	96.6	80.3	82.5
	4	78.4	15.0	13.5	12.6	14.5	92.6	94.0	87.0	98.3	82.0	84.9
	6	70.0	13.1	11.7	11.8	13.0	81.3	84.6	72.6	96.5	72.4	69.9
6	2	89.1	91.5	89.9	83.1	91.3	99.8	99.9	99.9	99.9	87.9	91.7
	4	80.8	75.6	70.6	68.6	74.5	98.7	99.2	97.9	99.7	78.6	84.1
	6	70.5	60.6	55.2	57.2	59.9	94.5	97.3	94.7	99.3	68.2	69.6
7	2	51.1	10.8	11.0	8.2	11.5	40.5	49.8	41.8	56.8	41.1	51.2
	4	41.7	9.2	8.5	7.9	9.4	27.9	35.0	29.1	45.3	25.6	38.7
	6	37.6	8.5	7.5	7.7	8.6	19.4	26.9	23.8	38.3	15.2	24.7
8	2	23.5	4.8	5.2	4.6	5.2	5.4	4.9	5.5	5.3	4.6	6.7
	4	27.7	5.1	5.1	5.0	5.1	5.0	5.1	5.0	5.2	3.4	6.3
	6	34.9	5.1	5.0	4.9	5.5	5.1	5.0	4.7	5.3	4.3	6.3

parameter value by an estimator introduces a non negligible random term in the empirical distribution function, and this can affect the finite-sample behavior of the test statistic under study. To study this issue, we assume that the true DGP is the AR(2) process  $Y_t = (\theta + 0.3)Y_{t-1} - 0.3\theta Y_{t-2} + \varepsilon_t$ , where  $\varepsilon_t \sim i.i.d. \mathcal{N}(0, 1)$ . However, we fit an AR(1) model to the generated time series and, using least squares residuals, test for serial dependence. Table 4 shows the percentage rejections for parameter  $\theta \in \{0, 0.4, 0.8\}$ .

For  $\theta = 0$  the above AR(2) process reduces to an AR(1) process. All rejection rates are fairly close to the nominal level of 5%, which is not surprising since the residuals will be close to Gaussian white noise. For  $\theta = 0.4$ , the residuals are non i.i.d. and hence for the test statistics  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  the results of Theorems 1 and 2 do not apply. Nonetheless, the empirical rejection rates of these tests and the (R)BDS tests do not deviate too much from the nominal level of 5%. Finally, as the value of  $\theta$  increases and becomes equal to 0.8, all tests indicate residual dependence.

**Table 4.** Percentages of empirical rejection rates of eleven test statistics using least squares residuals obtained from fitting an AR(1) model to an AR(2) process; 5% nominal level,  $n = 100$ .

$\theta$	m	BDS	Rank-based BDS test statistics									
			$\tilde{S}_{n,0.3}^*$	$\tilde{T}_n^*$	$\tilde{M}_n^*$	$\tilde{T}_n^*$	$\tilde{S}_{n,0.3}$	$\tilde{T}_n$	$\tilde{M}_n$	$\tilde{T}_n$	$\mathcal{I}_{n,\delta}(m)$	$\tilde{\mathcal{I}}_n(m)$
0	2	4.8	4.1	4.4	4.4	4.7	3.7	3.3	4.2	4.4	3.3	3.2
	4	4.7	4.0	3.7	4.3	4.5	3.9	5.2	4.8	4.3	4.5	3.5
	6	4.9	4.7	4.2	4.5	4.9	3.9	5.3	5.1	3.7	4.7	3.8
0.4	2	4.9	4.5	5.4	4.9	5.0	4.8	3.8	4.8	4.6	4.5	4.6
	4	5.0	5.3	5.5	5.1	5.7	5.4	5.6	5.2	6.5	5.3	5.6
	6	5.5	5.4	4.8	5.2	5.9	6.1	5.3	5.2	6.3	4.8	5.2
0.8	2	18.1	7.8	10.8	7.9	9.4	25.3	28.5	23.7	29.7	13.0	13.2
	4	12.9	8.0	9.5	7.7	8.4	15.9	18.2	14.5	24.3	12.9	12.2
	6	10.9	7.8	8.5	7.5	8.2	11.4	13.3	11.2	19.5	11.7	10.5

An investigation of parameter estimation uncertainty on the properties of the two proposed test statistics in an analytic way is rather difficult. However, we expect that parameter uncertainty has no impact on the limiting distribution of  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$ , because parametric model parameter estimators typically converge to the true parameter values at a much faster rate than our kernel-based non parametric estimator of  $C(m, \delta)$ .

## 5. Illustrative examples

### 5.1. S&P 500 index

As a first illustration, we consider the daily S&P 500 stock price index from January 1, 1992 to December 31, 2003. The complete series was analyzed by Hong and White (2005). Here we consider two subperiods. Period 1 (11/2000–02/2003;  $n = 608$ ) corresponds to the worst decline in the S&P 500 Index since 1931, with the end of the “dot-com bubble” around November 2000. Period 2 (03/2003–12/2003;  $n = 218$ ) corresponds to an upward trend with moderate volatility, indicating the start of a new bull market in the first quarter of 2003. We test the geometric random walk hypothesis, which is equivalent to testing the log-returns for serial independence.

Table 5 reports the bootstrapped  $p$ -values of the (R)BDS, the Skaug–Tjøstheim test statistic  $\mathcal{J}_n(m)$ , and the Hong–White test statistic  $\mathcal{T}_n(m)$ . With abuse of notation, these latter two test statistics are the sum of  $m$  single-lag two-dimensional test statistics; see Hong and White (2005) for the definitions of  $\mathcal{J}_n(m)$  and  $\mathcal{T}_n(m)$ . For the first (downward) downward period, the results of almost all test statistics suggest that the log-returns are not i.i.d. with much stronger evidence from  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$  for higher dimensions than from the other test statistics. On the other hand, the test statistics BDS,  $\tilde{\mathcal{I}}_n$ ,  $\tilde{M}_n$ , and  $\mathcal{T}_n(m)$  do not reject  $\mathbb{H}_0$ . The second (upward) period shows a complete different picture. There, except for the test statistics  $\tilde{\mathcal{I}}_n$  and  $\mathcal{J}_n(m)$ , all test results suggest that the log-returns are i.i.d., that is, the S&P 500 daily stock price index follows a random walk. Given these results, and considering the volatility of the series, a related null hypothesis worth investigating is the martingale difference hypothesis.

### 5.2. Canadian lynx data

As a second illustration, we consider the classic log (base 10) lynx data ( $n = 114$ ) that have been extensively analyzed. It is generally believed that this series is non linear, but there is no

**Table 5.** Bootstrap  $p$ -values of tests for daily S&P 500 stock returns; period 1 (11/2000–02/2003;  $n = 608$ ), period 2 (03/2003–12/2003;  $n = 218$ ). Boldface numbers indicate rejection of  $\mathbb{H}_0$  at a nominal 5% level.

Period	$m$	BDS	Rank-based BDS test statistics											
			$\tilde{S}_{n,0.3}^*$	$\tilde{T}_n^*$	$\tilde{M}_n^*$	$\tilde{T}_n^*$	$\tilde{S}_{n,0.3}$	$\tilde{T}_n$	$\tilde{M}_n$	$\tilde{T}_n$	$\mathcal{I}_{n,\delta}(m)$	$\tilde{\mathcal{I}}_n(m)$	$\mathcal{T}_n(m)$	$\mathcal{J}_n(m)$
1	2	0.21	0.07	0.07	0.14	0.08	0.22	0.57	0.53	0.91	0.07	<b>0.01</b>	0.73	<b>0.04</b>
	4	0.29	<b>0.00</b>	<b>0.00</b>	<b>0.02</b>	<b>0.00</b>	<b>0.00</b>	0.30	0.59	0.09	<b>0.00</b>	<b>0.00</b>	0.73	<b>0.00</b>
	6	0.36	<b>0.00</b>	<b>0.00</b>	<b>0.02</b>	<b>0.00</b>	<b>0.00</b>	0.30	0.58	<b>0.01</b>	<b>0.00</b>	<b>0.00</b>	0.74	<b>0.00</b>
	8	0.43	<b>0.00</b>	<b>0.00</b>	<b>0.02</b>	<b>0.00</b>	<b>0.00</b>	0.29	0.76	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.74	<b>0.00</b>
2	2	0.22	0.93	0.91	0.31	0.89	0.25	0.33	0.22	<b>0.00</b>	0.63	0.84	0.83	0.12
	4	0.30	0.97	0.91	0.49	0.80	0.54	0.10	0.85	<b>0.00</b>	0.43	0.39	0.81	<b>0.04</b>
	6	0.36	0.49	0.41	0.34	0.48	0.09	0.12	0.88	<b>0.00</b>	0.32	0.26	0.81	<b>0.01</b>
	8	0.46	0.15	0.13	0.15	0.15	0.00	0.31	0.75	<b>0.00</b>	0.29	0.07	0.80	<b>0.02</b>

**Table 6.** Bootstrap  $p$ -values of eleven test statistics for serial independence applied to the residuals of four time series models fitted to the log lynx series ( $n = 114$ ). Boldface numbers indicate rejection of  $\mathbb{H}_0$  at a nominal 5% level.

Model	$m$	BDS	Rank-based BDS test statistics									
			$\tilde{S}_m^*$	$\tilde{T}_m^*$	$\tilde{M}_m^*$	$\tilde{T}_m^*$	$\tilde{S}_m$	$\tilde{T}_m$	$\tilde{M}_m$	$\tilde{T}_m$	$\mathcal{I}_{n,\delta}(m)$	$\tilde{\mathcal{I}}_n(m)$
AR(2)	2	0.25	0.08	0.07	0.55	<b>0.04</b>	<b>0.04</b>	0.67	0.54	<b>0.01</b>	<b>0.02</b>	<b>0.04</b>
	4	0.31	<b>0.02</b>	<b>0.01</b>	0.38	<b>0.01</b>	0.08	0.40	0.12	<b>0.01</b>	<b>0.02</b>	<b>0.02</b>
	6	0.43	<b>0.01</b>	<b>0.01</b>	0.62	<b>0.01</b>	0.51	0.56	0.04	<b>0.02</b>	0.21	0.06
SETAR(2;7,2)	2	0.26	0.39	0.33	0.64	0.34	0.51	0.59	0.81	0.21	0.24	0.29
	4	0.34	0.15	0.15	0.67	0.13	0.40	0.94	0.28	0.09	0.19	0.34
	6	0.44	0.23	0.25	0.58	0.21	0.96	0.63	0.15	0.08	0.32	0.34
SETAR(3;1,7,2)	2	0.25	0.57	0.66	0.38	0.63	0.60	0.98	0.56	0.13	0.38	0.42
	4	0.32	0.26	0.40	0.27	0.32	0.48	0.92	0.27	0.15	0.20	0.25
	6	0.41	0.41	0.44	0.17	0.41	0.64	0.62	0.15	0.14	0.16	0.25
ExpAR(2)	2	0.25	<b>0.02</b>	0.12	0.32	<b>0.01</b>	<b>0.02</b>	0.12	0.15	<b>0.01</b>	<b>0.00</b>	<b>0.00</b>
	4	0.33	<b>0.02</b>	0.14	0.39	<b>0.01</b>	0.23	0.14	0.68	<b>0.02</b>	<b>0.01</b>	<b>0.01</b>
	6	0.43	<b>0.01</b>	0.38	0.55	<b>0.00</b>	0.12	0.38	0.32	<b>0.04</b>	<b>0.03</b>	<b>0.01</b>

agreement on which non linear model is most appropriate for the data. Tong (1990, Chapter 7) reproduced four time series models proposed in the literature. For the original series, the set of models consists of an AR(2) model, a SETAR(2;7,2) model, and a SETAR(3;1,7,2) model. For the mean-deleted log lynx series, we consider an EXPAR(2) model. Using three parametric portmanteau tests statistics, Tong (1990, p. 388) detected “no obvious lack of fit” of the specified SETAR(2;7,2) model. In addition, he remarked that the DGP underlying the series can be well represented by the fitted SETAR(3;1,7,2) model. By contrast, the adequacy of the fitted AR(2) and the fitted EXPAR(2) model was seriously questioned.

Table 6 shows  $p$ -values, based on 1,000 bootstrap replicates, of the (R)RBDS statistics and our test statistics (2.5) and (2.6) applied to the residuals of the abovementioned models. For the AR(2) model we see that BDS,  $\tilde{S}_m^*$ ,  $\tilde{T}_m^*$ , and  $\tilde{M}_m^*$  fail to reject the null hypothesis of residual serial independence at a 5% nominal level, and for almost all values of  $m$ . For the fitted SETAR(2;7,2) and SETAR(3;1,7,2) models, the  $p$ -values suggest that both these models adequately capture the non linear phenomena in the data. Finally, for the fitted ExpAR(2) model, we observe very pronounced evidence of residual dependence from the reported  $p$ -values of the RBDS test statistics  $\tilde{S}_{n,0.3}^*$ ,  $\tilde{T}_n^*$ ,  $\tilde{T}_n$ , and from our test statistics  $\mathcal{I}_{n,\delta}(m)$  and  $\tilde{\mathcal{I}}_n(m)$ . Unfortunately, the BDS test statistic fails to notice any model inadequacy in the residuals obtained from the AR(2) and ExpAR(2) specifications.

In summary, most test statistics correctly indicate the presence or lack of serial independence in the residuals of the fitted models. One notable exception, however, is the test statistic  $\tilde{M}_m^*$  that gives rise to far too high  $p$ -values in all cases. Moreover, from the  $p$ -values reported for residuals of the AR(2) models we see that in the cases of  $\tilde{S}_{n,0.3}^*$  and  $\tilde{T}_n^*$  no evidence of residual serial dependence is detected when  $m = 2$ , while for  $m = 4$  and 6 the  $p$ -values of these test statistics are smaller than the nominal 5% level. Thus, it is recommended not to rely completely on low-dimensional test results. Finally, we like to mention that the results (not displayed here) of the well-known Ljung-Box and McLeod-Li portmanteau test statistics provided no evidence of inadequacy of the fitted time series models.

### 6. Concluding remarks

This paper has developed two non parametric portmanteau tests for detecting time series non linearities in high dimensions using the correlation integral. The tests statistics are consistent

and asymptotically normally distributed. The finite-sample rejection probabilities of the tests do not differ too much from their asymptotic levels. Our tests are invariant under continuous monotonic transformation of data. Moreover, no unrealistic data requirements are needed when the dimension  $m$  increases.

On the whole, our simulation results indicate that in commonly used samples, the  $\tilde{\mathcal{I}}_n(m)$  test performs similarly, and sometimes better, as the best performing RBDS test statistic  $\tilde{T}_n$ . An additional advantage is that our test statistics can be relatively easily computed by adjusting the fast MATLAB code for computing the traditional BDS statistic of Kanzler (1999), whilst  $\tilde{T}_n$  is computationally demanding when  $n > 200$ , and is therefore not viable in practice.

### Acknowledgments

We are grateful to a referee for his helpful comments and suggestions that rendered a much improved version. We wish to thank Kilani Ghoudi for providing us with the C code of the RBDS tests, and Yong-mia Hong for sending us the Gauss code of the Hong–White test statistic.

### Funding

This work is supported in part for Yuan by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under Award Number G12MD007597.

### Appendix: Proofs

#### Proof of Theorem 1.

(i) We first show

$$\sqrt{n} \begin{pmatrix} C_{n,\delta}(m) - C^m \\ C_{n,\delta}(1) - C \end{pmatrix} \xrightarrow{D} \mathcal{N}(0, \Omega), \text{ where } \Omega = 4 \begin{pmatrix} \tau_m^2 & mC^{m-1}\tau^2 \\ mC^{m-1}\tau^2 & \tau^2 \end{pmatrix}, \quad (\text{A.1})$$

where  $\tau_m^2 = K^m - C^{2m} + 2 \sum_{i=1}^{m-1} (K^{m-i}C^{2i} - C^{2m})$ .

Since  $(C_{n,\delta}(m), C_{n,\delta}(1))^T$  is a vector  $U$ -statistic with non degenerate kernel, so it is asymptotical normal, we only need to find its asymptotic marginal distribution and asymptotic covariance. For this, let  $\eta_m(z) = E_{\mathbb{H}_0} \mathbf{1}(\|z - Z_2\| \leq \delta)$ , and  $\tau_m^2 = \text{Var}_{\mathbb{H}_0}(\eta_m(Z_1))$ , then  $0 < \tau_m^2 < \infty$ . In fact, by Theorem 2.1 of Brock et al. (1996), which in turn is based on Theorem 1c of Denker and Keller (1983), we have

$$\sqrt{n}(C_{n,\delta}(m) - C^m) \xrightarrow{D} \mathcal{N}(0, 4\tau_m^2)$$

and

$$\sqrt{n}(C_{n,\delta}(1) - C) \xrightarrow{D} \mathcal{N}(0, 4\tau^2).$$

We only need to check the asymptotic covariance of  $\sqrt{n}(C_{n,\delta}(m) - C^m)$  and  $\sqrt{n}(C_{n,\delta}(1) - C)$ . Expanding the product  $\sqrt{n}(C_{n,\delta}(m) - C^m)\sqrt{n}(C_{n,\delta}(1) - C)$  and leaving out terms with zero expectation, we have

$$\begin{aligned} \text{Cov}(\sqrt{n}(C_{n,\delta}(m) - C^m), \sqrt{n}(C_{n,\delta}(1) - C)) &= n \frac{1}{\binom{n-m+1}{2}} \frac{1}{\binom{n}{2}} \times \left( (n-m) \right. \\ &\times m E \left[ \left( \mathbf{1}(|Y_1 - Y_2| \leq \delta, |Y_2 - Y_3| \leq \delta, \dots, |Y_m - Y_{m+1}| \leq \delta) - C^m \right) \right] \end{aligned}$$

$$\begin{aligned}
& \times \left( \mathbf{1}(|Y_1 - Y_2| \leq \delta) - C \right) \Big] \\
& + (n - m) \sum_{j=1}^{m-1} j E \left[ \left( \mathbf{1}(|Y_1 - Y_2| \leq \delta, |Y_2 - Y_3| \leq \delta, \dots, |Y_m - Y_{m+1}| \leq \delta) - C^m \right) \left( \mathbf{1}(|Y_1 - Y_3| \leq \delta) - C \right) \right] \\
& + (n - m)m(n - m - 1) E \left[ \left( \mathbf{1}(|Y_1 - Y_2| \leq \delta, |Y_2 - Y_3| \leq \delta, \dots, \right. \right. \\
& \left. \left. \times |Y_m - Y_{m+1}| \leq \delta) - C^m \right) \left( \mathbf{1}(|Y_1 - Y_{m+2}| \leq \delta) - C \right) \right] \\
& + (\text{other terms in which the } Z_i\text{'s and } Z_j\text{'s have common indices} \\
& \text{for their components}) \\
& + \sum_{j=0}^{n-2m-1} (n - 2m - j)m \\
& \times E \left[ \left( \mathbf{1}(|Y_1 - Y_{m+1}| \leq \delta, |Y_2 - Y_{m+2}| \leq \delta, \dots, |Y_m - Y_{2m}| \leq \delta) - C^m \right) \right. \\
& \left. \times \left( \mathbf{1}(|Y_1 - Y_{m+1}| \leq \delta) - C \right) \right] \\
& + \sum_{j=0}^{n-2m-1} (n - 2m - j) \left( \binom{m}{2} - m \right) \\
& \times E \left[ \left( \mathbf{1}(|Y_1 - Y_{m+1}| \leq \delta, |Y_2 - Y_{m+2}| \leq \delta, \dots, |Y_m - Y_{2m}| \leq \delta) - C^m \right) \right. \\
& \left. \times \left( \mathbf{1}(|Y_1 - Y_2| \leq \delta) - C \right) \right] \\
& + \sum_{j=0}^{n-2m-1} (n - 2m - j) 2m(n - 2m) \\
& \times E \left[ \left( \mathbf{1}(|Y_1 - Y_{m+1}| \leq \delta, |Y_2 - Y_{m+2}| \leq \delta, \dots, |Y_m - Y_{2m}| \leq \delta) - C^m \right) \right. \\
& \left. \times \left( \mathbf{1}(|Y_1 - Y_{2m+1}| \leq \delta) - C \right) \right] \\
& \sim n \frac{1}{\binom{n-m+1}{2}} \frac{1}{\binom{n}{2}} 2m(n - 2m) \sum_{j=1}^{n-2m} j \\
& \times E \left[ \left( \mathbf{1}(|Y_1 - Y_{m+1}| \leq \delta, |Y_2 - Y_{m+2}| \leq \delta, \dots, |Y_m - Y_{2m}| \leq \delta) - C^m \right) \right. \\
& \left. \times \left( \mathbf{1}(|Y_1 - Y_{2m+1}| \leq \delta) - C \right) \right] \\
& = n \frac{1}{\binom{n-m+1}{2}} \frac{1}{\binom{n}{2}} m(n - 2m)^2 (n - 2m + 1)
\end{aligned}$$

$$\begin{aligned} & \times E \left[ \left( \mathbf{1}(|Y_1 - Y_{m+1}| \leq \delta, |Y_2 - Y_{m+2}| \leq \delta, \dots, |Y_m - Y_{2m}| \leq \delta) - C^m \right) \right. \\ & \left. \times \left( \mathbf{1}(|Y_1 - Y_{2m+1}| \leq \delta) - C \right) \right] \\ & \rightarrow 4m[P(|Y_1 - Y_{m+1}| \leq \delta, |Y_1 - Y_{2m+1}| \leq \delta) \\ & \quad \times P(|Y_2 - Y_{m+2}| \leq \delta, \dots, |Y_m - Y_{2m}| \leq \delta) - C^{m+1}] \\ & = 4m[P(|Y_1 - Y_2| \leq \delta, |Y_1 - Y_3| \leq \delta)C^{m-1} - C^{m+1}] = 4mC^{m-1}(\gamma - C^2). \end{aligned}$$

Observe that  $\gamma = E_{Y_1}[P(|Y_2 - Y_1| \leq \delta, |Y_3 - Y_1| \leq \delta)|Y_1] = E_{Y_1}P^2(|Y_2 - Y_1| \leq \delta|Y_1) = K$ . Using the relationship  $\tau^2 = K - C^2$  we have (A.1).

Now let  $g(x, y) = \ln(x/y^m)$  and  $G(x, y) = (\partial g/\partial x, \partial g/\partial y) = (1/x, -m/y)$ . We have  $g(C^m, C) = 0$ . By (A.1) and the delta method (e.g. Serfling, 1980), we get

$$\sqrt{n}\mathcal{I}_{n,\delta}(m) = \sqrt{n}(g(C_{n,\delta}(m), C_{n,\delta}(1)) - g(C^m, C)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(m, \delta)),$$

where  $\sigma^2(m) = G(C^m, C)\Omega G^T(C^m, C) = 4\left(\frac{\tau_m^2}{C^{2m}} + \frac{m^2\tau^2 + 2m^2(C^2 - K)}{C^2}\right) = 4\left(\frac{\tau_m^2}{C^{2m}} + \frac{m^2(C^2 - K)}{C^2}\right)$ .

- (ii) Since the estimator is an  $U$ -statistic with an indicator kernel, with  $\delta \in (0, \delta_0)$  for some small  $\delta_0 > 0$ , so  $\{\mathcal{I}_{n,\delta}(m) : \delta \in (0, \delta_0)\}$  is a P-Donsker class (van der Vaart and Wellner, 1996), and so the weak convergence in Part (i) is in the  $l^\infty$  sense, thus we can take the limit  $\delta \rightarrow 0$  on both sides of the expression in Theorem 1 (i), and get

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \sqrt{n}\mathcal{I}_{n,\delta}(m) = \sqrt{n}(g(C_{n,\delta}(m), C_{n,\delta}(1)) - g(C^m, C)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(m))$$

with  $\sigma^2(m) = \lim_{\delta \rightarrow 0} \sigma^2(m, \delta)$ , and the result does not depend on the rate at which  $\delta \rightarrow 0$ . In the above argument, we use the fact that, the theorems for Donsker class in van der Vaart and Wellner (1996) for empirical process are in fact valid for  $U$ -statistics based on i.i.d. observations. □

**Proof of Theorem 2.** Let  $\xrightarrow{D}$  denote weak convergence on  $C[a, b]$ , and  $W(\cdot)$  be the Gaussian process on  $[a, b]$  with  $E[W(\delta)] = 0$  and  $Cov(W(\delta_1), W(\delta_2)) = \sigma^2(m, \delta_1 \wedge \delta_2)$ . For clarity we denote  $\mathcal{I}_{n,\delta}(m)$  as  $\mathcal{I}_n(m, \delta)$ . We first prove

$$\sqrt{n}\mathcal{I}_n(m, \cdot) \xrightarrow{D} W(\cdot). \tag{A.2}$$

In fact, by Theorem 1 (i) and similarly in its proof, we see that any finite dimensional distribution of  $\sqrt{n}\mathcal{I}_n(m, \cdot)$  weakly converges to those of  $W(\cdot)$ , so we only need to check tightness of the sequence  $\{X_n(\cdot)\} \equiv \{\sqrt{n}\mathcal{I}_n(m, \cdot)\}$  on  $C[a, b]$ , and by Theorem 12.3 in Billingsley (1968, p. 95), we only need to show that (the original result is for  $C[0, 1]$ , and it is the same for  $C[a, b]$ )

- (i) The sequence  $\{X_n(0)\}$  is tight.
- (ii) There exist constants  $\gamma \geq 0$  and  $\alpha > 1$  and a non decreasing, continuous function  $G$  on  $[a, b]$  such that for all  $\delta_1, \delta_2 \in [a, b]$ ,

$$P(|X_n(\delta_1) - X_n(\delta_2)| \geq \lambda) \leq \frac{1}{\lambda^\gamma} |G(\delta_1) - G(\delta_2)|^\alpha. \tag{A.3}$$

In fact, define  $\ln(0/0) = c = \lim_{\delta \rightarrow 0} \ln(C(m, \delta)/C^m(1, \delta))$ , then  $\{X_n(0)\} \equiv \{c\}$ , a constant sequence, so (i) is true. Also, by Theorem 1,  $X_n(\delta_1) - X_n(\delta_2) \xrightarrow{D} \mathcal{N}(0, \rho^2)$ , with  $\rho^2 = \rho^2(\delta_1, \delta_2) = \sigma^2(m, \delta_1) + \sigma^2(m, \delta_2) - 2\sigma^2(m, \delta_1 \wedge \delta_2) = \sigma^2(m, \delta_1 \vee \delta_2) - \sigma^2(m, \delta_1 \wedge \delta_2) = |\sigma^2(m, \delta_1) - \sigma^2(m, \delta_2)|$  and  $\delta_1 \vee \delta_2 = \max(\delta_1, \delta_2)$ . By definition of  $C_{n,\delta}(m)$ , its variation is an increasing function of  $\delta$ , that is,  $\sigma^2(m, \cdot)$  is an increasing function

and apparently it is continuous. Thus, there is an increasing function  $G$  on  $[a, b]$  satisfying  $\rho^2(\delta_1, \delta_2) \leq |G(\delta_1) - G(\delta_2)|^2$  for all  $\delta_1, \delta_2 \in [a, b]$ . Thus, for large  $n$ ,

$$P(|X_n(\delta_1) - X_n(\delta_2)| \geq \lambda) \leq \frac{1}{\lambda^2} |G(\delta_1) - G(\delta_2)|^\alpha$$

and (ii) is satisfied and so (A.2) is true.

Now, let  $J(r) = \int_a^b r(\delta)h(\delta)d\delta$  for  $r \in C[a, b]$ , then  $\forall r, s \in C[a, b]$ ,  $|J(r) - J(s)| \leq \int_a^b ||r - s||h \rightarrow 0$  as  $||r - s|| \rightarrow 0$ , that is,  $J(\cdot)$  is a continuous function on  $C[a, b]$ . Thus, by (A.3) and the basic property of weak convergence, we have

$$\sqrt{n}\tilde{\mathcal{I}}_n(m) = \sqrt{n} \int_a^b \mathcal{I}_{n,\delta}(m)h(\delta)d\delta = J(\sqrt{n}\mathcal{I}_n(m, \cdot)) \xrightarrow{D} J(W) = \int_a^b W(\delta)h(\delta)d\delta.$$

Lastly, divide  $[a, b]$  into  $k$  intervals each of equal length  $\Delta$ , let  $\delta_i$  be a point in the  $i$ th interval, then  $J(W) = \lim_{\Delta \rightarrow 0} \sum_{i=1}^k W(\delta_i)h(\delta_i)\Delta := \lim_{\Delta \rightarrow 0} J_k$ . Since  $W(\delta_i) \sim \mathcal{N}(0, \sigma^2(m, \delta_i))$  and

$$\text{Var}(J_k) = \sum_{i=1}^k \sum_{j=1}^k \sigma^2(m, \delta_i \wedge \delta_j)h(\delta_i)h(\delta_j)\Delta^2 := V_{k,h}^2(m),$$

we have  $J_k \sim \mathcal{N}(0, V_{k,h}^2(m))$ , with characteristic function  $s_k(t) = \exp(-t^2 V_{k,h}^2(m)/2) \rightarrow \exp(-t^2 V_h^2(m)/2)$  which is the characteristic function for  $N(0, V_h^2(m))$ , and where

$$V_h^2(m) = \lim_{k \rightarrow \infty} V_{k,h}^2(m) = \int_a^b \int_a^b \sigma^2(m, s \wedge t)h(s)h(t)dsdt.$$

Thus, we get  $J(W) \sim \mathcal{N}(0, V_h^2(m))$ . □

## References

- Billingsley, P. (1968). *Convergence in Probability Measures*. New York: Wiley.
- Brock, W. A., Dechert, W. D., LeBaron, B., Scheinkman, J. A. (1996). A test for independence based on the correlation dimension. *Econ. Rev.* 15:197–235.
- Brock, W. A., Hsieh, W. D., LeBaron, B. (1991). *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*. Cambridge, MA: MIT Press.
- Delgado, M. A. (1996). Testing serial independence using the sample distribution function. *J. Time Ser. Anal.* 17:271–285.
- Denker, M., Keller, G. (1983). On U-statistics and von Mises statistics for weakly dependent process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64:505–522.
- Ferguson, T. S., Genest, C., Hallin, M. (2000). Kendall's tau for serial independence. *Can. J. Stat.* 28:587–604.
- Genest, C., Ghoudi, K., Rémillard, B. (2007). Rank-based extensions of the Brock, Dechert, and Scheinkman test. *J. Am. Stat. Assoc.* 102:1363–1376.
- Grassberger, P., Procaccia, I. (1984). Dimensions and entropies of strange attractors from a fluctuating dynamics approach. *Physica* 13D:34–54.
- Hallin, M., Puri, M. L. (1992). Rank tests for time series analysis: a survey. In: Brillinger, D. R., Parzen, E., Rosenblatt, M., eds. *New Directions in Time Series Analysis, Part I* (pp. 111–153). New York: Springer.
- Hong, Y., White, H. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* 73:837–901.
- Kanzler, L. (1999). Very fast and correctly sized estimation of the BDS statistic. University of Oxford. Available at: [http://papers.ssrn.com/paper.taf?abstract\\_id=151669](http://papers.ssrn.com/paper.taf?abstract_id=151669).
- Pompe, B. (1993). Measuring statistical dependencies in a time series. *J. Stat. Phys.* 73(3/4):587–610.
- Prichard, D., Theiler, J. (1995). Generalized redundancies for time series analysis. *Physica* D84:476–493.
- Robinson, P. M. (1991). Consistent nonparametric entropy-based testing. *Rev. Econ. Stud.* 58:437–453.

- Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
- Skaug, H. J., Tjøstheim, D. (1993). Nonparametric tests of serial dependence. In: Subba Rao, T., ed. *Developments in Time Series Analysis* (pp. 207–229). London: Chapman and Hall.
- Skaug, H. J., Tjøstheim, D. (1996). Measures of distance between densities with application to testing for serial independence. In: Robinson, P., Rosenblatt, M., eds. *Time Series Analysis in Memory of E. J. Hannan* (pp. 363–377). New York: Springer.
- Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
- van der Vaart, A., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes, with Applications to Statistics*. New York: Springer.