Comments Solicited

A Simple Nonparametric Test for Independence

Bruce Mizrach*
Department of Economics
Rutgers University
First Draft: November 1991
This Draft: January 1995

Abstract:

A stationary stochastic process is defined to be *locally independent* if it eventually becomes independent of past realizations. I develop a simple nonparametric test for this condition. Size and power comparisons favor this statistic over the one proposed by Brock, Dechert and Scheinkman (1987) in samples under 250 observations.

JEL Classification: C14;

Keywords: U-statistics; Nonlinear dependence;

*Address for editorial correspondence: Department of Economics, Rutgers University, 301b New Jersey Hall, New Brunswick, NJ 08903. e-mail: mizrach@rci.rutgers.edu, (908) 932-8261 (voice) and (908) 932-7416 (fax). I would like to acknowledge helpful discussions with Buz Brock, Dee Dechert, David Johnson, Michael Phelan, and Philip Rothman. I also received many useful comments from seminar participants at the Department of Statistics of the Wharton School, the Bureau of Labor Statistics, and the 1992 North American Summer Meetings of the Econometric Society. I owe an especially large debt to Blake LeBaron for his multi-faceted contributions to the paper. A FORTRAN program implementing the test as well as any future revisions to this paper may be found at http://www-rci.rutgers.edu/~mizrach/ This research was initiated during my visit to the Wharton School's Finance Department.

1. Introduction

For more than a decade now, nonlinear science has been contributing important insights in a wide range of fields. The observation that many scalar mappings generate data with the spectral density of white noise held open the possibility that apparently complex phenomena had simple explanations. An earnest search for chaos in economic and financial data followed.

With this decade has come perspective. The majority of evidence suggests that economic and financial data, while nonlinear, is not low dimensional chaos. Research is now centered on understanding nonlinearities of any kind, regardless of whether they are chaotic.

Nonparametric statistical methods, because they require very weak population assumptions, have proven their versatility in this area, enabling the study of chaos and linear time series models with the same tools. The correlation integral of Grassberger and Procaccia (1984), a member of a class of nonparametric statistics known as U-statistics, has become a standard tool in the analysis of dynamical systems. Brock, Dechert and Scheinkman (BDS, 1987) cleverly adapted the correlation integral into a powerful test for independence and identical distribution. The BDS test has become the benchmark by which to judge alternative nonparametric testing procedures.

The weak population assumptions of the BDS come at a price. Typically, nonparametric statistics converge at a slower rate than parametric statistics, with the rate of convergence depending on dimension. I document this slow convergence of the BDS in a series of Monte Carlo exercises. The BDS, though asymptotically normal, has extremely high rates of Type I error. With the uniform (0,1) distribution, for example, a nominal 5% test rejects about 22% of the time in a sample of 250 observations with a two dimensional test and 35% of the time in a five dimensional test.

This paper reveals that a small amount of structural information can dramatically improve the size and power of nonparametric tests for independence. I define *local independence of order p* as the property that the unconditional probability x_{t+p} of equals the conditional probability of x_{t+p} given x_t. I then construct a distribution free test of this hypothesis for any combination of p's.

This new simple nonparametric test (SNT) has several distinct advantages. It is computationally simpler. By testing for a specific type of dependence, the SNT helps the data analyst in identifying the model; the BDS simply tells you that the data are not independent, but it does not suggest any particular alternative.

Monte Carlo simulations reveal that the SNT has much lower rates of Type I error than the BDS. These errors are also roughly equal across populations, while for the BDS, they vary as much as 50% or more between the normal and uniform distributions.

Once the test statistics are properly sized, the SNT frequently is a more powerful test for nonlinear dependence. For 5 of the 7 data generating mechanisms studied by Brock, Hsieh and LeBaron (1991), the SNT has better power. The test has also proved useful in applications. Mizrach (1996) has used the test to determine the delay time to be used in a phase space reconstruction.

The organization of the paper is as follows. In Section 1, I lay out the general theory of U-statistics leading to a development of the correlation integral. From there, I present the BDS test. Section 2 develops the SNT. Monte Carlo analysis of the size of the two tests is in Section 3, and power is examined in Section 4. Section 5 contains the two applications.

Section 6 concludes with a view towards future research.

2. U-Statistics and the BDS Test

This section begins with the basic theory of U-statistics. I proceed with three examples, two relatively straightforward, the third, quite complicated. The latter is the correlation integral which underlies the test for independence proposed by Brock, Dechert and Scheinkman (1987). Asymptotic theory for U-statistics is discussed in Section 2.2. I develop the BDS test in Section 2.3.

2.1 An Introduction to U-Statistics

I begin with a definition of the generalized sample averages known as U-statistics.[1] Let $\{x_i\}$ be a strictly stationary stochastic process taking on values in R^n with distribution function F. Let $\{X_1, ..., X_n\}$ be a sample of size n. The components include a kernel, a symmetric measurable function $h : (R^m)^j \to R$, where j and m are integers ≥ 1, and the permutation operator, $\Sigma_{n,j}$, which sums over the $\begin{pmatrix} n \\ j \end{pmatrix}$ distinct combinations of j-elements in a sample of size n. Define the canonical mapping,

$$U_n = U(X_1, X_2, ..., X_n) = \Sigma_{n,j} h(X_1, X_2, ..., X_n). \tag{2.1}$$

To demonstrate the breadth of the U-statistics, I begin with two simple examples from Serfling (1980, Ch.5). Let $j = 1, m = 1$, $h(x_i) = x_i$, $\begin{pmatrix} n \\ 1 \end{pmatrix}^{-1} = 1/n$ then

$$U(X_1, ..., X_n) = 1/n \sum_{i=1}^{n} X_i = \bar{X}. \tag{2.2}$$

or just the sample mean. Now let $j = 2, m = 1$, $\begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} = 2/n(n-1)$, and $h(x_i - x_k) = (x_i - x_k)^2/2$, one now has,

$$\begin{aligned} U(X_1, ..., X_n) &= \frac{2}{n(n-1)} \sum_{1 \leq i < k} h(X_i - X_k), \\ &= \frac{1}{(n-1)} \left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right], \end{aligned} \tag{2.3}$$

namely, the sample variance.

Our third example, which will lead us to the heart of the paper, requires a little more notation. Define the vector m-history,

$$x_t^m = (x_t, x_{t+1}, ..., x_{t+m-1}), \tag{2.4}$$

and denote its joint distribution by $F(x_t^m)$. Introduce the kernel with $j = 2$, $h : R^m \times R^m \to R$,

[1] The seminal reference is Hoeffding (1948). My notation follows Serfling (1980).

$$h(x_t^m, x_s^m) = I\left[\|x_t^m - x_s^m\| < \varepsilon\right] \equiv I(x_t^m, x_s^m, \varepsilon), \tag{2.5}$$

where $I(.)$ is the indicator (or Heaviside) function, and take $\|.\|$, for simplicity, to be the ℓ_∞ norm on R^m,

$$I(x_t^m, x_s^m, \varepsilon) = I\left[\left(\max_{0\le i\le m-1} |x_{t+i} - x_{s+i}|\right) < \varepsilon\right]. \tag{2.6}$$

The correlation integral is given by

$$C(m, \varepsilon) \equiv \int_X \int_X I(x_t^m, x_s^m, \varepsilon) dF(x_t^m) dF(x_s^m). \tag{2.7}$$

(2.7) is the expected number of m-vectors less than ε away from any given m-vector. At dimension 1, this simplifies to

$$C(1, \varepsilon) = \int_X [F(x_t + \varepsilon) - F(x_t - \varepsilon)] dF(x_t). \tag{2.8}$$

A U-statistic,

$$C(m, N, \varepsilon) \equiv \frac{2}{N(N-1)} \sum_{t=1}^{N-1} \sum_{s=t+1}^{N} I(X_t^m, X_s^m, \varepsilon), \tag{2.9}$$

is a consistent estimator of (2.8),where $N = n - m + 1$.

2.2 Asymptotic Theory for U-Statistics

The basic results in asymptotic theory for U-statistics in the time series case are due to Denker and Keller (1983). To understand their work more clearly, we need to introduce some additional notation.

Consider first a set of functions associated with each kernel, for $1 \le c \le j-1$,

$$h_c(x) \equiv \int ... \int h(x_1, x_2, ..., x_c) dF(x_2) ... dF(x_c). \tag{2.10}$$

(2.10) is just a conditional expectation for the kernel given $X_2, \ldots, X_c$,

$$E_F[h(X_1, ..., X_j) | x_2 = X_2, ..., x_c = X_c]. \tag{2.11}$$

It will be useful for our purposes to center around the unconditional mean, defining

$$\theta(F) \equiv E_F[h(X_1, ..., X_j)]. \tag{2.12}$$

Also define the conditional expectations centered about θ,

$$\tilde{h}_c(x) \equiv h_c(x) - \theta(F).$$

Denker and Keller then show that under weak dependence assumptions[2], and for a bounded kernel h such that $\sup_{t_1\ge 1,...,t_j\ge 1} E\left|h(x_{t_1}, ..., x_{t_j}\right|^{2+\delta} < \infty$, the statistic

[2]Denker and Keller (1983) assume that $\{x_t\}$ is an absolutely regular process with mixing coefficients $\beta(n)$ satisfying $\beta(n)^{\delta/(2+\delta)} = O(n^{-2+\varepsilon})$ for some $\delta > 0$, and $\varepsilon < 1/2, \sigma^2 \neq 0$. This condition is trivially satisfied for an i.i.d. process.

$$\sqrt{n}\frac{U_n - \theta(F)}{j\sigma_n} \xrightarrow{d} N(0,1), \tag{2.13}$$

where the variance is given by

$$\sigma_n^2 = j^2 \times \left(E_F[(\tilde{h}_1(X_1)^2] + 2\sum_{t=2}^{n} E_F[(\tilde{h}_1(X_1)\tilde{h}_1(X_t)] \right). \tag{2.14}$$

The derivation of the BDS statistic as well as the SNT developed in section 3 will draw heavily on this result.

2.3 The BDS Test

Brock, Dechert, and Scheinkman (1987) have developed a powerful test for independence and identical distribution based on the correlation integral we first encountered in part 2. If the sample is generated by an i.i.d. data generating mechanism, the joint distribution of a vector m-history should factor into the m (identical) marginals.

$$F(x_t^m) = \prod_{i=0}^{m-1} F(x_{t+i}) = [F(x_t)]^m. \tag{2.15}$$

By comparing $C(m, N, \varepsilon)$ for different m's, the BDS approach tests whether this factorization is correct.

Intuitively, consider the event with $m = 1$, $I[|x_t - x_s| < \varepsilon]$. Integrating with respect to all x's, one has the unconditional probability of all such events, $C(1, \varepsilon)$. Using the ℓ_∞ norm, it should then be m-times as unlikely that m events will all be less than ε,

$$Prob.\left[\left(\max_{0\le i\le m-1} |x_{t+i} - x_{s+i}|\right) < \varepsilon\right] = Prob.\left[|x_t - x_s| < \varepsilon\right]^m. \tag{2.16}$$

More formally, define

$$S(m, N, \varepsilon) = C(m, N, \varepsilon) - C(1, N, \varepsilon)^m. \tag{2.17}$$

Because the statistic (2.17) is the difference between two *sample* measures, the asymptotic theory is more complicated. Denker and Keller's theorem pertains only to the standard case with one sample and one population moment, e.g. $S(m, \varepsilon) \equiv C(m, N, \varepsilon) - C(1, \varepsilon)^m$, and cannot be applied directly.

Lacking knowledge of the population moment, BDS proceed using the "delta" method described in Serfling (1980, Ch.3). Serfling shows that the asymptotic variance of (2.17) is the same as its' Taylor expansion[3],

$$S(m, N, \varepsilon) \approx C(m, N, \varepsilon) - C(m, \varepsilon) - mC(1, \varepsilon)^{m-1}[C(1, N, \varepsilon) - C(1, \varepsilon)]. \tag{2.18}$$

Using the linear combination of U-statistics in (2.18), BDS then showed that

[3] More generally, consider some vector of sample and population statistics $z \equiv (z_1, ..., z_k)$ $N(\mu, \Sigma)$. Let $F(z)$ be some function with a non-zero differential for each component, and let $D \equiv [\partial F_i/\partial z_j|_{x=\mu}]$ then $F(z)$ is distributed $N(F(\mu), D\Sigma D')$

$$\sqrt{N}\frac{S(m,N,\varepsilon)}{\sqrt{Var\left[S(m,N,\varepsilon)\right]}} \xrightarrow{d} N(0,1), \tag{2.19}$$

where,

$$\begin{aligned} Var\left[S(m,N,\varepsilon)\right] \quad &= 2\sum_{j=1}^{m-1} K^{m-j}C(1,N,\varepsilon)^{2j} + K^m + \\ &(m-1)^2 C(1,N,\varepsilon)^{2m} - m^2 K C(1,N,\varepsilon)^{2(m-1)}, \end{aligned} \tag{2.20}$$

with

$$K = \int\int\int_X I(x_t,x_s,\varepsilon)I(x_s,x_r,\varepsilon)dF(x_t)dF(x_s)dF(x_r). \tag{2.21}$$

The statistic (2.19) has been widely applied[4] in the literature as a test for independence. Given its ability to shrug off nuisance parameters[5], the BDS has become a powerful portmanteau statistic for model specification.

3. A Simple Nonparametric Test

In this section, I propose a simpler test for independence. It has several advantages. It is computationally quicker, involving only calculations of order N, as opposed to N^2 for the BDS. The variance of the U-statistic I derive is similar to that of a binomial random variable. In Sections 3 and 4, I show that the simple test has much better size and power than the BDS in samples of less than 250 observations.

I will define a stochastic process to be *locally independent of order p* if the realization x_t provides no information about the process p periods ahead. Formally, this implies the equality of the conditional and unconditional distributions. Let $(p_1,...,p_{m-1})$ be a set of increasing integers on $[1,L], L < n-m+1$. Local independence then implies

$$Prob.\left[x_{t+p_{m-1}} < \varepsilon,...,x_{t+p_1} < \varepsilon, x_t < \varepsilon\right] = \left(Prob.\left[x_t < \varepsilon\right]\right)^m. \tag{3.1}$$

To estimate the joint, $F(x_t^m)$, and marginal, $F(x_t)$, distributions in (3.1), introduce the kernel function $h: R \to R$,

$$h(x_t) = I\left[x_t < \varepsilon\right] = \left\{ \begin{array}{ll} 1, & if\ x_t > \varepsilon \\ 0, & otherwise \end{array} \right\} \equiv I(x_t,\varepsilon). \tag{3.2}$$

The joint unconditional probability that m leads of the x's are less than ε is given by

$$\theta(m,\varepsilon) = \int_X \prod_{i=0}^{m-1} I(x_{t+p_i},\varepsilon)dF(x_t), \tag{3.3}$$

where for notational convenience I set $p_0 = 0$. A consistent estimator of (3.6) is the U-statistic

[4] In private communication, Dee Dechert reports that nearly 200 articles have employed the BDS.

[5] The nuisance parameter results pertain to the ability to apply the test to residuals from linear filters. See DeLima (1994).

$$\theta(m,N,\varepsilon) = \sum_{t=1}^{N} \prod_{i=0}^{m-1} I(X_{t+p_i},\varepsilon)/N, \tag{3.4}$$

where $N = n - \max[p_i]$.

A simple test for local independence can then be constructed using consistent estimators of the first two moments of this U-statistic.

Proposition: Let $\{x_t\}$ be locally independent for any $p_i \in [1,L]$, $i = 1,...,m-1$, $L < N$, then if $\theta(m,\varepsilon) > 0$,

$$\sqrt{N}\frac{[\theta(m,N,\varepsilon) - \theta(m-1,N,\varepsilon)\theta(1,N,\varepsilon)]}{[\theta(m-1,N,\varepsilon)\theta(1,N,\varepsilon)(1-\theta(m-1,N,\varepsilon))(1-\theta(1,N,\varepsilon))]^{0.5}} \xrightarrow{d} N(0,1). \tag{3.5}$$

Proof: The proof can be stated as a corollary to Denker and Keller (1983) after using the delta method to calculate the asymptotic variance. I will supress the ε's throughout to conserve on notation. Taking the Taylor expansion around the means of the numerator in (3.5) gives us the U-statistic,

$$\theta(m,N) - \lambda_1\theta(1,N) - \lambda_2\theta(m-1,N) + \lambda_1\theta(1) + \lambda_2\theta(m-1) - \theta(1)(m-1), \tag{3.6}$$

where $\lambda_1 = \theta(m-1)$ and $\lambda_2 = \theta(1)$. As in BDS, we may think of (3.6) as a combination of U-statistics with kernel,

$$h(u) = I(x_t^m) - \lambda_1 I(x_t) - \lambda_2 I(x_t^{m-1}). \tag{3.7}$$

Because of the one-argument kernel ($j = 1$), the conditional expectations (2.10) take a simple form,

$$\tilde{h}_1(u) = I(x_t^m) - \lambda_1 I(x_t^1) - \lambda_2 I(x_t^{m-1}) + \theta(m). \tag{3.8}$$

Now, using (3.8), calculate the leading term of the variance in (2.14),

$$\begin{aligned} E\tilde{h}_1(u_1^2) &= E\left[I(x_t^m) - 2\lambda_1 I(x_t^m) - 2\lambda_2 I(x_t^m)\right] \\ &\quad + E\left[\lambda_1^2 I(x_t) + 2\lambda_1\lambda_2 I(x_t^{m-1}) + \lambda_2^2 I(x_t^{m-1}) - \theta^2(m)\right], \\ &= \theta(m)[1 - \theta(1) + \theta(m-1) - \theta(m)]. \end{aligned} \tag{3.9}$$

For the covariance terms, because of the i.i.d assumption, we need only consider the $m-1$ overlapping terms of the m-vectors,

$$\begin{aligned} 2E\left[\sum_{t=2}^{m-1} \tilde{h}_1(u_1)\tilde{h}_1(u_t)\right] &= 2E\sum_{t=2}^{m-1}\left[I(x_1^m) - \lambda_1 I(x_1) - \lambda_2 I(x_1^{m-1}) - \theta(m)\right] \\ &\quad \times[I(x_t^m) - \lambda_1 I(x_t) - \lambda_2 I(x_t^{m-1})], \\ &= 2\theta(m)[\theta(m) - \theta(m-1)]. \end{aligned} \tag{3.10}$$

Combining (3.9) and (3.10) , we have

$$\theta(1)\theta(m-1)[(1-\theta(m-1))(1-\theta(1))]. \tag{3.11}$$

This is of course just the denominator of (3.5) for large N. This completes the calculation of the variance and the proof.□

To provide some intuition, note that the expected number of m-chains with a value of 1 in a sample of size N is

$$\mu = N\theta(m,N,\varepsilon) = \sum_{x=0}^{N} x \left(\begin{array}{c} N \\ x \end{array} \right) \theta(m,\varepsilon)^x (1-\theta(m,\varepsilon))^{N-x}. \tag{3.12}$$

The variance is given by

$$\begin{aligned} \sigma^2_{B_N} &= \mu_2' - \mu^2 = \sum_{x=0}^{N} x^2 \left(\begin{array}{c} N \\ x \end{array} \right) \theta(m,\varepsilon)^x (1-\theta(m,\varepsilon))^{N-x} - N^2\theta(m,\varepsilon)^2, \\ &= N\theta(m,\varepsilon)(1-\theta(m,\varepsilon)). \end{aligned} \tag{3.13}$$

We see that the (3.13) is simply the variance of a pair of Bernoulli trials with probability $\theta(m,\varepsilon)$ and $\theta(m-1,\varepsilon)$. While there are many conceptual advantages for the statistic (3.4), the principal advantage will be if the SNT can improve substantially over the BDS, (2.19), in finite samples. I turn to that in the next section.

4. Finite Sample Properties of the Statistics

I compare the sizes of the BDS and the simple nonparametric test (SNT) using three different distributions. In part 1, the data generating mechanism is independent normal (0,1) noise. To replicate the leptokurtic distribution found in many financial asset returns, I use, in part 2, a Student-t with 3 degrees of freedom. In part 3, I use a highly entropic distribution, the uniform (0,1).

There are judgmental inputs that go into the tests that can dramatically alter the results reported below. With the BDS, Brock, Hsieh, and LeBaron (1991) report that the size of the test is generally best when ε is set to one sample standard deviation. For the SNT, I found that setting ε equal to the sample mean usually performed best. In all exercises, I use sample sizes of $n = 25, 50, 100$ and 250, dimensions of $m = 2, 3, 4$ and 5, with 5,000 replications.

4.1 The Standard Normal Distribution

The most glaring feature of Table 1, which looks at the finite sample sizes of the BDS statistic, is the enormous rate of Type I error. For $n = 25$, the statistic rejects more than 55% of the time at $m = 2$, and almost 60% of the time at $m = 5$, for a 10% test. Only at $n = 250$ do Type I error rates fall below a 2:1 ratio. In general though, size improves as we lower m for a given n, and as we raise n for a given m.

A closer look also reveals that the BDS rejects more often in the left tail than the right. Mizrach (1994) shows that this is due to bias in both the numerator and the denominator of (2.19). Once recentered at the statistics' negative sample mean, the BDS is also skewed to the right. As a consequence, the convergence to normality is actually quicker in the left tail than in the right.

The SNT rejects too often at dimensions $m = 2, 3$,and 4. With reference to Table 2 for $m = 2$, a 10% test rejects 17.6% of the time. Conversely, at $m = 5$, the rejection rate is 4.3%. By $n = 50$, the overrejection at $m = 2$ has fallen under 13%, and the underrejection has risen to 7% at $m = 5$. At $n = 100$, we are within sampling error of the appropriate nominal sizes.

Compared to the BDS, the SNT is far more reliably sized. Using $m = 2$, the best for the BDS, the chart for the SNT at $n = 50$ compares favorably to the BDS at $n = 250$. Since the majority of economic data samples will pertain to the chart at $n = 50$, the SNT has a much wider range of applicability. We'll see whether this result carries over to distributions other than the normal.

4.2 The Student-t Distribution with 3 Degrees of Freedom

The BDS and the SNT both perform better with the Student-t than with the normal. The improvement with the BDS is somewhat larger. A 10% test at $m = 2$ rejects a little bit more than 40% of the time in Table 3 at $n = 25$. This compares to a 55% rejection rate in Table 1 with $N(0, 1)$ deviates. The convergence to normality is equally rapid. Because the BDS starts at a lower rate of Type I error, by $n = 250$, the BDS is within sampling error of its nominal size.

The $Student-t$ narrows the margin of victory, but, in sampling from a leptokurtic distribution, the SNT is still the victor nonetheless. For the high dimensions, the underrejections are less of a problem at $m = 4$ and 5, and the overrejections are reduced at $m = 2$ and 3, but only slightly. In any case, the Student-t poses no special problem for the SNT.

4.3 The Uniform (0,1) Distribution

When working with high dimensional stochastic data generating mechanisms, one often runs into the so-called "curse of dimensionality." It is simply very hard to learn something about a high dimensional population from a small sample. The phase space X here is the unit hypercube in R^m. Even with a large ε, the data can become quite sparse. The population standard deviation for the uniform is $\sqrt{1/12}$ or 0.2886. With this choice of ε, a hypercube of that length will contain only $(0.2886)^m$ of the sample. For $m = 5$, this is 0.20%. Conversely, to find 28.86% of the sample, we would need a neighborhood that contains $(0.2886)^{1/m}$ of the range. For $m = 5$, this is 78% of the range.

This problem is clearly more pronounced the more entropic is the distribution. With the normal, I numerically integrated (2.7) and found a value of 0.52 for $C(1, 1.0)$. Here, the "curse" would leave us with about 4% of the sample at $m = 5$. While we would find many more data points with a larger ε, size and (especially) power considerations favor a choice of ε at one standard deviation.

To see how the curse of dimensionality influences nonparametric tests of independence, I look at the uniform (0,1) distribution in Tables 5 and 6. For the BDS, the uniform distribution is the most problematic. For $n = 25$ and $m = 5$, the BDS rejects 82% of the time in a 10% test and 75% of the time in a 2% test. Even at $n = 250$, with $m = 2$, a 10% test still rejects 33% of the time, and at $m = 5$, it still rejects 42% of the time. The size at 250 observations resembles the performance at 50 with normal variates. The SNT proves its' mettle with the uniform distribution. There are no appreciable differences between the normal in Table 2 and the uniform in Table 6. The obvious question is whether this comes at the cost of power. I turn to that in Section 5.

5. Power of the Simple Nonparametric Test vs. the BDS

The power comparisons require a level playing field. The high rates of Type I error will make the BDS seem more powerful than it actually is. The first step in conducting a comparison of the power of the SNT and the BDS is to construct a set of finite sample critical values. In performing the Monte Carlo exercises in Tables 1 and 2, I also compiled a set of critical values that would give the BDS and the SNT the appropriate nominal sizes in finite samples.

As can be seen in Table 7, the negative bias of the BDS is very much in evidence in small samples. A 95% confidence region, asymptotically sized at ± 1.96, is instead the interval [-14.12,13.49]. At the high dimensions, the BDS produces what would be once in a lifetime realizations from a true normal distribution.

Only at $n = 250$, our largest small sample, does the BDS begin to resemble a normal distribution. A 95% confidence region has shrunk to $[-2.15, 2.24]$. Note that the convergence to normality after $n = 25$ is actually slower in the right tail than the left. Mizrach (1994) shows that this is due to a positive skewness accompanying the negative bias.

The critical values in Table 8 for $n = 25$ look similar to those of the BDS for $n = 100$. At $n = 100$, the critical values for the SNT are within sampling error of the standard normal. There is still some marginal Type I error, especially at $m = 2$. I will use finite sample corrections for both statistics throughout the power exercises.

5.1 The Data Generating Mechanisms

To facilitate comparisons with the Monte Carlo analysis of Brock, Hsieh and LeBaron (1991), I will utilize the same 7 data generating mechanisms used therein. They are (1) an AR(1) process; (2) an MA(1) process; (3) an ARCH(1) model; (4) a GARCH(1,1) model; (5) a nonlinear MA(1) process; (6) a threshold autoregressive model; and (7) the so-called "tent map." To simplify the discussion, I only look at $m = 2$, where both tests are most powerful. Solely for space reasons, I will talk about 5% tests. I will use standard normal disturbance terms throughout, though it should be noted from Table 3, that this is heavily biased in favor of the BDS.

5.1.1 First Order Autoregressive Process

The data generating mechanism is the AR(1) model.

$$x_t = 0.5x_{t-1} + \eta_t,\ \eta_t \sim\ NID(0,1). \tag{5.1}$$

At all sample sizes in Table 9 the SNT is more powerful than the BDS. This is somewhat surprising since the order of dependence here is infinite. The BDS is not getting much power from dependence at lags greater than p. In Section 5.2, I find that even with setting $p = 2$, the power of the SNT is close to that of the BDS. Score after one round, 4 for the SNT, 0 for the BDS.

5.1.2 First Order Moving Average

The data generating mechanism is the MA(1) model.

$$x_t = 0.75\eta_{t-1} + \eta_t,\ \eta_t\ \sim NID(0,1). \tag{5.2}$$

Again, by confining the alternatives to dependence at lag 1, I get far superior power to the BDS. With reference to Table 10, the advantage, in a 5% test for $n = 50$ is nearly 1.5 to 1 for the SNT; at $n = 100$, the advantage is about 30%. Both statistics are quite powerful overall, with 100% rejections at $n = 250$. I break the tie at the 1% level where the SNT has a modest advantage. Score: SNT - 8, BDS - 0.

5.1.3 An ARCH Model

The autoregressive conditional heteroscedasticity (ARCH) model is a popular parametric model for financial time series because it has a fat-tailed unconditional distribution. The process can be written as:

$$x_t = [1.0 + 0.5x_{t-1}^2]^{0.5}\eta_t,\ \eta_t\ \sim NID(0,1). \tag{5.3}$$

This is an interesting test because the data are linearly uncorrelated. With ε at the sample mean, the SNT will have no power.

The obvious thing to do is take the squares of x, as these are correlated. In this form, the SNT proved to be a powerful test. I was able to surpass the BDS in the two smaller samples in Table 11. In the larger samples, the BDS overtook the SNT.

The use of the squared kernel reveals again that a small amount of information can dramatically boost the power of the test. With rejections increasing substantially at $p = 1$ and with a squared kernel, this is a substantial step towards identifying the process. The BDS, on the other hand leaves us nearly as ignorant as we were at the start. This was a split, leaving the running score at SNT-10, BDS-2;

5.1.4 The GARCH Model

The generalized autoregressive conditional heteroscedasticity (GARCH) model is the ARMA version of the ARCH process. The squared residuals can now depend on both lagged squared residuals and lagged conditional variances. It can be written as:

$$h_t = [1.0 + 0.1x_{t-1} + 0.8h_{t-1}^2]^{0.5},\ x_t = (h_t)^{0.5}\eta_t,\ \eta_t\ \sim NID(0,1). \tag{5.4}$$

The results here were roughly similar to the ARCH case. The SNT wins easily in the small sample, rejecting 22.3% of the time at $n = 25$. The SNT is marginally nosed out at $n = 100$. Both tests reject nearly 100% of the time for $n = 250$. Running score, 12 for the SNT, 4 for the BDS.

5.1.5 A Nonlinear Moving Average Process

Several types of nonlinear moving average processes have become popular. In particular, the so-called bilinear model tries to capture the cross dependence of two moving average terms at different lags. The process can be written as:

$$x_t = 0.8\eta_{t-1}\eta_{t-2} + \eta_t,\ \eta_t\ \sim NID(0,1). \tag{5.5}$$

The power is not concentrated at one particular lag, and this proved problematic for the SNT and the BDS. Using the squares of the x's, the SNT gains victories in the two small samples, loses marginally at $n = 100$, and loses fairly sizably at $n = 250$. A split, leaving the running score at SNT - 14, BDS - 6.

5.1.6 A Threshold Autoregressive Model

Another popular parametric model for nonlinear time series involves a data induced change in regime. For these reasons, it is sometimes called a self-exciting threshold autoregressive (SETAR) model. We can write it as:

$$x_t = \left\{ \begin{array}{r} -0.5x_{t-1} + \eta_t, \quad if\ x_{t-1} > 1 \\ 0.4x_{t-1} + \eta_t, \quad if\ x_{t-1} \leq 1 \end{array} \right\},\ \eta_t \sim NID(0,1) \tag{5.6}$$

While this model is technically of first order dependence, the actual order of dependence can be much longer. The data may linger quite a while in a particular regime, until a shock induces it to cross the threshold. Neither the BDS nor the SNT does a particularly good job in small samples, rejecting at 4.2% and 2.4% respectively for $n = 25$. See Table 14. The BDS wins narrow victories at $n = 50, 100$ and 250 as well. Score, SNT - 14, BDS - 10.

5.1.7 The Tent Map

One of the prototypical chaotic maps of the interval is the so-called tent map. The tent map is a remarkably good linear random number generator. I can write the tent map as:

$$x_t = \left\{ \begin{array}{r} 2x_{t-1}, \quad if\ x_{t-1} > 0.5 \\ 2 - 2x_{t-1}, \quad if\ x_{t-1} > 0.5 \end{array} \right\}. \tag{5.7}$$

I seeded the map with a draw from a uniform random number generator on (0,1). Note that there is a singularity at 0.5.

The SNT does not have power to determine whether or not x is locally independent at the unconditional mean of $\varepsilon = 0.5$. To understand this, notice that x_t will be less than 0.5

when x_{t-1} lies on the intervals $[0.75, 1.0]$ and $[0.0, 0.25]$. Since these intervals are of equal measure, it is equally likely that x_t will be less than 0.5 when x_{t-1} is above or below the population mean.[6] I thus chose to set ε equal to the sample mean plus one-half the sample standard deviation. The BDS won victories at all four sample sizes, even at $n = 25$. This left the final score at SNT - 14, BDS - 14.

5.2 Power When the Order of Dependence is Misspecified

The size and power advantages of the SNT clearly arise from specifying the order of dependence. I take a brief look in this section at the consequences of choosing some p other than the optimal. I first looked at the AR(1) process (5.1). In Table 16, I set $p = 2$ rather than 1. The SNT still captures that the data are not i.i.d. 65% of the time in the sample with $n = 250$. With the MA(1) process (5.2), there is no memory beyond last period. Here the consequences are more dramatic. The rejection frequencies are much like those of white noise in Table 4.

This reveals both the strength and weakness of the SNT. It is useful test only if you take a stand on the order of dependence, but this information helps guide in the direction of building a model. The BDS will rarely miss nonlinear dependence in large data samples, but it is not as helpful in identifying the source of the rejection.

6. Summary of Size and Power Comparisons

Finite sample comparisons generally favor the simple nonparametric test (SNT) over the statistic of Brock, Dechert and Scheinkman (BDS, 1987). The SNT is far closer to its' nominal size in small samples. In the case of the uniform distribution, the advantage is better than 5 to 1.

The size of the BDS also varies widely across populations. For the normal distribution, an interval of $[-14.23, 13.49]$ provides a 95% confidence interval in a sample of 25. Mistakenly using this interval on a uniformly distributed population would result in Type I errors of more than 50%.The critical values for the SNT, on the other hand, are virtually indistinguishable across populations.

Having to tabulate finite sample critical values for a very large set of distributions is certainly a drawback for the BDS, but what is far more damaging is that the data analyst must make parametric assumptions to utilize the test. If enough population information is available to choose a particular table of critical values, an entirely different (and probably parametric) approach could have been utilized at the outset.

All of my power exercises conservatively utilized normally distributed errors. Even granting the BDS this handicap, 14 of 28 power comparisons favor the SNT over the BDS. In samples of 50 observations or less, the ratio is 9 to 5 in favor of the SNT.

The simple nonparametric test offers two additional advantages. The first is computational. The SNT requires calculations of order N, as opposed to N^2 for the BDS. Ordinarily in this age of cheap computing power, this advantage might be heavily discounted. Using the

[6] I thank David Johnson for this argument. Johnson and Robert McClelland (1992) develop a procedure in the spirit of Section 2 for testing the independence of regressors and disturbances.

BDS in data analysis though will frequently require bootstrap or Monte Carlo simulations, and it is here that the computational advantage will come into play.

The second advantage relates to inference. When the BDS rejects the null hypothesis of i.i.d., it usually does not lead in the direction of a particular alternative. With the SNT, the testing procedure tells you that dependence is present at the p^{th} lag of the data. This could prove to be useful knowledge in identifying a statistical model.

The SNT is not the test statistic for all situations. The BDS's resistance to nuisance parameters makes it the preferred choice for specification tests. The BDS also worked better with the chaotic tent map and the regime switching SETAR process at all sample sizes. Two questions are left for future research. The SNT must now be compared with other tests for nonlinear dependence. A second extension is to see whether a rich class of parametric nonlinear models can capture types of dependence most often found in economic and financial time series.

References

Brock, W.A. and W.D. Dechert (1989), "Statistical Inference Theory for Measures of Complexity in Chaos Theory and Nonlinear Science," in *Measures of Complexity and Chaos*, N.B. Abraham et al eds., New York: Plenum Press, 79-97.

Brock, W.A., W.D. Dechert, and J. Scheinkman (1987), "A Test for Independence Based on the Correlation Dimension," University of Wisconsin-Madison Workshop Paper 8702.

Brock, W.A., D.A. Hsieh, and B. LeBaron (1991), *Nonlinear Dynamics, Chaos and Instability*, Cambridge: MIT Press.

DeLima, P. (1994),"Nuisance Parameters in U-Statistics," Working Paper, Johns Hopkins U.

Grassberger, P. and I. Procaccia, (1984), "Dimensions and Entropies of Strange Attractors from a Fluctuating Dynamics Approach," *Physica* 13D, 34-54.

Hoeffding, W. (1948), "A Class of Statistics with Asymptotically Normal Distribution," *Annals of Mathematical Statistics* 19, 293-325.

Johnson, D. and R. McClelland (1992), "Nonparametric Tests for the Independence of Regressors and Disturbances," Working Paper, Bureau of Labor Statistics.

Mizrach, B. (1994), "Using U-statistics to Detect Business Cycle Nonlinearities," in Willi Semmler (ed.), *Business Cycles: Theory and Empirical Investigation*, Boston: Kluwer Press, 107-29.

Mizrach, B. (1996), "Determining Delay Times for Phase Space Reconstruction of the FF/DM Exchange Rate," *Journal of Economic Behavior and Organization*, forthcoming.

Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Table 1
Monte Carlo Analysis of the BDS Statistic
N(0,1) i.i.d. Random Variates*

$n = 25$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	23.94	28.19	32.67	23.35	21.09	18.25
$m = 3$	24.26	28.82	33.22	24.08	21.68	18.72
$m = 4$	25.32	30.58	35.28	23.26	20.81	18.84
$m = 5$	27.05	32.62	37.86	22.33	20.55	18.47
$N(0, 1)$	1.00	2.50	5.00	5.00	2.50	1.00

$n = 50$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	9.42	14.44	19.36	14.44	11.32	7.92
$m = 3$	9.98	15.02	20.28	14.68	11.66	8.76
$m = 4$	10.96	15.90	21.82	15.12	12.22	9.60
$m = 5$	11.78	16.76	23.18	15.94	13.24	10.56

$n = 100$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	3.52	6.92	11.58	9.76	6.70	4.00
$m = 3$	3.68	7.77	12.38	10.80	6.76	4.16
$m = 4$	3.78	8.00	13.28	10.20	7.36	4.94
$m = 5$	4.70	8.52	13.74	10.78	7.86	5.68

$n = 250$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	1.36	3.64	7.30	6.88	3.80	2.08
$m = 3$	1.34	3.64	7.46	7.10	4.78	2.60
$m = 4$	1.24	3.70	7.42	7.44	4.76	2.70
$m = 5$	1.38	3.50	8.00	7.72	4.98	2.92

*n is the sample size and m is the embedding dimension. The simulations for $n = 50, 100$ and 250 are based on 5,000 replications. For $n = 500$ and $n = 1,000$, I used 2500 replications. I set ε equal to one sample standard deviation throughout.

Table 2
Monte Carlo Analysis of the Simple Nonparametric Test
N(0,1) i.i.d. Random Variates*

$n = 25$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	5.26	7.54	9.42	8.22	4.92	2.12
$m = 3$	0.60	2.90	8.66	3.68	1.56	0.50
$m = 4$	0.10	1.12	5.22	2.40	0.76	0.18
$m = 5$	0.00	0.40	2.88	1.52	0.54	0.04
$N(0, 1)$	1.00	2.50	5.00	5.00	2.50	1.00

$n = 50$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	2.36	6.80	7.57	5.04	4.18	1.62
$m = 3$	1.50	4.12	7.76	3.84	1.76	0.78
$m = 4$	0.60	2.64	7.54	2.88	1.26	0.36
$m = 5$	0.60	0.80	3.42	2.38	1.02	0.28

$n = 100$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	2.00	5.12	7.52	5.26	3.78	1.54
$m = 3$	1.46	3.30	6.06	3.80	2.00	0.82
$m = 4$	1.30	3.76	6.82	3.50	1.56	0.62
$m = 5$	0.42	2.00	5.84	3.16	1.52	0.50

$n = 250$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	1.36	2.76	6.82	5.90	2.36	1.20
$m = 3$	1.42	2.92	5.68	5.00	2.18	0.78
$m = 4$	1.18	3.08	6.30	4.16	1.96	0.70
$m = 5$	1.30	3.10	6.92	3.40	1.76	0.72

*All exercises are based on 5,000 replications. n is the sample size and m is the dimension. ε is set equal to the sample mean.

Table 3
Monte Carlo Analysis of the BDS Statistic
Student-t with 3 Degrees of Freedom*

$n = 25$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	14.08	18.94	23.48	16.78	13.72	10.66
$m = 3$	13.71	18.26	24.17	15.79	13.23	10.40
$m = 4$	13.82	19.12	25.02	15.94	12.82	10.56
$m = 5$	12.70	18.22	24.46	15.14	12.98	10.94
$N(0, 1)$	1.00	2.50	5.00	5.00	2.50	1.00

$n = 50$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	4.66	8.20	12.89	9.28	6.26	3.90
$m = 3$	4.20	8.34	13.25	8.72	6.02	4.14
$m = 4$	3.84	7.70	13.60	9.12	6.42	4.20
$m = 5$	3.82	7.54	13.32	9.34	7.14	4.90

$n = 100$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	1.76	4.62	7.48	7.48	4.64	2.38
$m = 3$	1.78	4.36	8.78	6.88	4.48	2.48
$m = 4$	1.38	4.36	8.78	7.10	4.36	2.38
$m = 5$	1.30	3.80	7.92	7.00	4.34	2.76

$n = 250$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	1.08	3.34	6.40	5.80	2.92	1.14
$m = 3$	1.06	3.32	6.08	5.82	3.08	1.50
$m = 4$	0.94	2.74	6.20	5.76	3.18	1.66
$m = 5$	0.90	2.72	5.80	6.04	3.62	1.60

*n is the sample size and m is the embedding dimension. The simulations use 5,000 repliations. I set ε equal to one sample standard deviation.

Table 4
Monte Carlo Analysis of the Simple Nonparametric Test
Student-t with 3 Degrees of Freedom*

$n = 25$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	4.70	7.30	9.36	7.84	5.26	2.04
$m = 3$	0.88	3.18	7.72	3.42	1.42	0.52
$m = 4$	0.40	1.74	5.40	2.06	0.70	0.22
$m = 5$	0.22	1.16	3.26	1.46	0.44	0.04
$N(0, 1)$	1.00	2.50	5.00	5.00	2.50	1.00

$n = 50$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	2.50	6.34	7.66	4.98	3.82	1.68
$m = 3$	1.44	3.74	7.56	4.12	1.96	0.68
$m = 4$	0.76	2,80	7.50	2.84	1.08	0.32
$m = 5$	0.24	1.18	3.96	2.42	1.04	0.30

$n = 100$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	1.88	4.64	7.38	5.50	3.58	1.42
$m = 3$	1.18	3.44	6.04	3.78	1.88	0.80
$m = 4$	1.22	3.26	6.80	3.76	1.60	0.58
$m = 5$	0.56	2.12	5.80	2.94	1.38	0.56

$n = 250$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	1.34	3.04	6.64	5.52	2.30	1.06
$m = 3$	1.40	3.26	5.76	4.70	2.22	0.76
$m = 4$	1.10	3.06	6.20	4.16	2.08	0.78
$m = 5$	1.16	3.16	6.84	3.46	1.80	0.78

*All exercises are based on 5,000 replications. n is the sample size and m is the dimension. ε is set equal to the sample mean.

Table 5
Monte Carlo Analysis of the BDS Statistics
Uniform (0,1)*

$n = 25$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	38.89	41.42	43.99	34.21	32.40	30.39
$m = 3$	40.64	42.95	45.41	34.06	32.64	30.83
$m = 4$	42.74	44.90	47.12	33.84	32.46	31.14
$m = 5$	45.43	47.55	49.59	32.51	31.21	30.19
$N(0, 1)$	1.00	2.50	5.00	5.00	2.50	1.00

$n = 50$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	31.92	35.58	38.64	29.62	27.19	25.20
$m = 3$	34.20	37.28	39.93	30.68	28.60	26.35
$m = 4$	35.50	38.67	41.34	31.83	29.76	27.94
$m = 5$	37.43	40.38	42.77	32.35	30.67	28.95

$n = 100$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	20.56	24.62	29.04	24.14	20.94	17.94
$m = 3$	22.36	26.62	31.26	25.00	21.84	18.90
$m = 4$	24.08	26.68	32.64	25.99	23.66	20.66
$m = 5$	27.50	31.34	35.16	28.08	25.32	22.56

$n = 250$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	7.86	11.96	17.32	15.00	11.62	8.72
$m = 3$	9.54	14.10	19.22	16.70	13.26	9.80
$m = 4$	11.72	67.20	21.74	17.86	14.74	11.48
$m = 5$	14.66	19.30	23.82	19.78	15.98	13.12

*n is the sample size and m is the embedding dimension. The simulations use 5,000 repliations. I set ε equal to one sample standard deviation.

Table 6
Monte Carlo Analysis of the Simple Nonparametric Test
Uniform (0,1)*

$n = 25$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	4.76	7.18	9.36	8.12	5.12	2.12
$m = 3$	0.76	3.62	10.66	3.18	1.32	0.48
$m = 4$	0.16	1.04	5.20	2.32	0.92	0.24
$m = 5$	0.00	0.58	2.86	1.44	0.30	0.06
$N(0, 1)$	1.00	2.50	5.00	5.00	2.50	1.00

$n = 50$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	2.14	6.44	7.28	4.58	3.54	1.06
$m = 3$	1.42	3.54	7.26	3.36	1.74	0.64
$m = 4$	0.62	2.72	7.50	2.90	1.46	0.42
$m = 5$	0.06	0.52	2.26	2.32	1.04	0.22

$n = 100$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	2.22	4.98	7.40	4.76	3.02	1.42
$m = 3$	1.40	3.56	6.72	.344	1.62	0.52
$m = 4$	1.18	3.82	7.48	3.22	1.54	0.56
$m = 5$	0.36	1.84	5.66	2.98	1.22	0.50

$n = 250$	%<-2.33	%<-1.96	%<-1.64	%>1.64	%>1.96	%>2.33
$m = 2$	1.26	2.62	6.16	5.32	2.10	0.98
$m = 3$	1.20	2.76	5.62	4.32	1.90	0.72
$m = 4$	0.96	2.62	5.90	3.96	1.84	0.72
$m = 5$	1.16	3.74	6.76	3.52	1.68	0.56

*All exercises are based on 5,000 replications. n is the sample size and m is the dimension. ε is set equal to the sample mean.

Table 7
Critical Values of the BDS Statistic
N(0,1) Random Variates*

$n = 25$	0.005	0.025	0.050	0.950	0.975	0.995
$m = 2$	-73.0595	-14.1237	-8.0341	7.7601	13.4925	57.9982
$m = 3$	-87.9192	-15.8944	-8.5690	8.6918	16.7674	93.5086
$m = 4$	-100.3838	-19.2087	-8.7500	9.5042	19.2857	92.2519
$m = 5$	-130.8905	-21.3401	-10.4344	10.7292	21.7927	110.0793
$N(0, 1)$	-2.58	-1.96	-1.64	1.64	1.96	2.58

$n = 50$	0.005	0.025	0.050	0.950	0.975	0.995
$m = 2$	-6.2303	-3.5823	-2.8964	2.9174	3.7507	5.7867
$m = 3$	-6.5520	-3.6787	-3.1238	3.1383	3.9996	6.6066
$m = 4$	-6.5871	-3.9201	-3.1328	3.3287	4.3240	6.8051
$m = 5$	-6.6698	-4.0927	-3.1374	3.6838	4.7889	7.7884

$n = 100$	0.005	0.025	0.050	0.950	0.975	0.995
$m = 2$	-3.2091	-2.5096	-2.1630	2.1607	2.6403	3.6753
$m = 3$	-3.2480	-2.5151	-2.1616	2.2227	2.7991	3.9959
$m = 4$	-3.2436	-2.5700	-2.1821	2.2966	2.9293	4.2858
$m = 5$	-3.4224	-2.5748	-2.2530	2.4914	3.1222	4.4838

$n = 250$	0.005	0.025	0.050	0.950	0.975	0.995
$m = 2$	-2.6196	-2.1479	-1.8349	1.8327	2.2388	2.9627
$m = 3$	-2.5776	-2.1140	-1.8227	1.9151	2.3497	3.1784
$m = 4$	-2.0611	-2.0931	-1.8299	1.9133	2.4257	3.2927
$m = 5$	-2.6547	-2.0984	-1.8178	1.9681	2.5199	3.5256

*n is the sample size and m is the embedding dimension. The simulations use 5,000 replications. I set ε equal to one sample standard deviation.

Table 8
Critical Values for the Simple Nonparametric Test
N(0,1) Random Variables*

$n = 25$	0.005	0.025	0.050	0.950	0.975	0.995
$m = 2$	-3.3558	-2.4815	-2.3417	1.8898	2.2186	3.0239
$m = 3$	-2.4612	-2.0498	-1.7974	1.4357	1.8061	2.2966
$m = 4$	-2.2913	-1.7974	-1.6773	1.2550	1.5959	2.0498
$m = 5$	-1.8209	-1.6773	-1.4285	1.2013	1.4285	1.9704
$N(0,1)$	-2.58	-1.96	-1.64	1.64	1.96	2.58

$n = 50$	0.005	0.025	0.050	0.950	0.975	0.995
$m = 2$	-2.9354	-2.2161	-2.0498	1.7466	2.1039	2.7223
$m = 3$	-2.6017	-2.1448	-1.8584	1.5055	1.8107	2.4386
$m = 4$	-2.3685	-1.9676	-1.7915	1.4298	1.7259	2.2187
$m = 5$	-2.0210	-1.7710	-1.5667	1.2813	1.6158	2.1052

$n = 100$	0.005	0.025	0.050	0.950	0.975	0.995
$m = 2$	-2.8465	-2.1332	-2.0275	1.6461	2.0545	2.4721
$m = 3$	-2.7386	-2.1125	-1.7446	1.5155	1.8295	2.4600
$m = 4$	-2.6085	-2.0992	-1.814	1.4771	1.7616	2.4582
$m = 5$	-2.2967	-1.8905	-1.6907	1.4183	1.7375	2.3140

$n = 250$	0.005	0.025	0.050	0.950	0.975	0.995
$m = 2$	-2.7007	-1.9967	-1.6918	1.6577	1.9170	.26365
$m = 3$	-2.6015	-2.0074	-1.7131	-1.6936	1.9152	2.4711
$m = 4$	-2.6046	-2.0153	-1.7467	1.5414	1.8474	2.4543
$m = 5$	-2.5946	-2.0657	-1.7719	1.4637	1.7874	2.4993

*All exercises are based on 5,000 replications. m is the dimension of the test and n is the sample size. ε is set to the sample mean.

Table 9
Power Against AR(1) DGM
BDS vs. SNT*

BDS	10%	5%	1%
$n = 25$	19.33	10.94	2.60
$n = 50$	50.96	39.96	20.84
$n = 100$	86.02	80.14	65.78
$n = 250$	99.86	99.69	99.12

SNT	10%	5%	1%
$n = 25$	32.54	27.20	8.84
$n = 50$	58.56	54.56	33.24
$n = 100$	93.10	86.94	76.70
$n = 250$	99.94	99.90	99.10

*The statistics reported are percentage rejections in a two-sided test at the $(1-\alpha)\%$ confidence level using the finite sample critical values in Tables 7 and 8. For the BDS, I set ε equal to one sample standard deviation, and for the SNT, I use the sample mean. For both tests, I use only $m = 2$. The order of dependence for the SNT is $p = 1$. n is the sample size. The data generating mechanism (DGM) is the AR(1) process: $x_t = 0.5x_{t-1} + \eta_t,\ \eta_t \sim NID(0,1)$.

Table 10
Power Against MA(1) DGM
BDS vs. SNT*

BDS	10%	5%	1%
$n = 25$	16.82	9.81	1.01
$n = 50$	48.64	36.82	16.62
$n = 100$	83.80	78.60	61.40
$n = 250$	100.00	100.00	99.20

SNT	10%	5%	1%
$n = 25$	37.40	13.80	3.00
$n = 50$	64.00	55.00	17.00
$n = 100$	95.00	91.60	83.40
$n = 250$	100.00	100.00	100.00

*The statistics reported are percentage rejections in a two-sided test at the $(1 - \alpha)\%$ confidence level using the finite sample critical values in Tables 7 and 8. For the BDS, I set ε equal to one sample standard deviation, and for the SNT, I use the sample mean. For both tests, I use only $m = 2$. The order of dependence for the SNT is $p = 1$. n is the sample size. The data generating mechanism (DGM) is the MA(1) process: $x_t = 0.75\eta_{t-1} + \eta_t,\ \eta_t \sim NID(0, 1)$.

Table 11
Power Against ARCH(1) DGM
BDS vs. SNT*

BDS	10%	5%	1%
$n = 25$	9.39	4.64	0.82
$n = 50$	36.60	23.00	4.82
$n = 100$	74.60	65.90	46.05
$n = 250$	99.28	98.84	95.49

SNT	10%	5%	1%
$n = 25$	17.04	11.48	3.66
$n = 50$	38.86	28.54	14.54
$n = 100$	62.96	52.10	38.08
$n = 250$	93.22	90.24	76.30

*The statistics reported are percentage rejections in a two-sided test at the $(1-\alpha)\%$ confidence level using the finite sample critical values in Tables 7 and 8. For the BDS, I set ε equal to one sample standard deviation, and for the SNT, I use the sample mean. For both tests, I use only $m = 2$. The order of dependence for the SNT is $p = 1$. n is the sample size. The data generating mechanism (DGM) is the ARCH(1) process: $x_t = [1.0 + 0.5x_{t-1}^2]^{0.5}\eta_t,\ \eta_t \sim NID(0,1)$.

Table 12
Power Against GARCH(1,1) DGM
BDS vs. SNT*

BDS	10%	5%	1%
$n = 25$	8.06	3.70	0.74
$n = 50$	56.84	41.10	10.44
$n = 100$	96.20	92.40	78.00
$n = 250$	100.00	100.00	100.00

SNT	10%	5%	1%
$n = 25$	29.84	22.30	8.94
$n = 50$	60.00	51.40	35.80
$n = 100$	88.80	84.40	76.60
$n = 250$	99.80	99.60	98.20

*The statistics reported are percentage rejections in a two-sided test at the $(1-\alpha)\%$ confidence level using the finite sample critical values in Tables 7 and 8. For the BDS, I set ε equal to one sample standard deviation, and for the SNT, I use the sample mean. For both tests, I use only $m = 2$. The order of dependence for the SNT is $p = 1$. n is the sample size. The data generating mechanism (DGM) is the AR(1) process: $x_t = 0.5x_{t-1} + \eta_t,\ \eta_t \sim NID(0,1)$.

Table 13
Power Against Nonlinear MA(1) DGM
BDS vs. SNT*

BDS	10%	5%	1%
$n = 25$	0.82	0.46	0.10
$n = 50$	13.20	5.12	0.06
$n = 100$	37.32	26.78	10.82
$n = 250$	74.62	66.16	48.74

SNT	10%	5%	1%
$n = 25$	9.90	6.22	1.34
$n = 50$	21.30	13.68	5.36
$n = 100$	32.10	23.00	13.22
$n = 250$	59.30	49.20	26.98

*The statistics reported are percentage rejections in a two-sided test at the $(1-\alpha)\%$ confidence level using the finite sample critical values in Tables 7 and 8. For the BDS, I set ε equal to one sample standard deviation, and for the SNT, I use the sample mean. For both tests, I use only $m = 2$. The order of dependence for the SNT is $p = 1$. n is the sample size. The data generating mechanism (DGM) is the nonlinear moving average process: $x_t = 0.8\eta_{t-1}\eta_{t-2} + \eta_t,\ \eta_t \sim NID(0,1)$.

Table 14
Power Against SETAR DGM
BDS vs. SNT*

BDS	10%	5%	1%
$n = 25$	8.87	5.87	0.86
$n = 50$	21.71	13.52	10.88
$n = 100$	39.04	29.92	15.26
$n = 250$	74.16	64.92	48.32

SNT	10%	5%	1%
$n = 25$	11.72	2.62	0.08
$n = 50$	15.30	1.21	1.26
$n = 100$	38.46	24.64	14.28
$n = 250$	70.42	60.52	32.86

*The statistics reported are percentage rejections in a two-sided test at the $(1-\alpha)\%$ confidence level using the finite sample critical values in Tables 7 and 8. For the BDS, I set ε equal to one sample standard deviation, and for the SNT, I use the sample mean. For both tests, I use only $m = 2$. The order of dependence for the SNT is $p = 1$. n is the sample size. The data generating mechanism (DGM) is the threshold autoregressive model: $x_t = -0.5x_{t-1} + \eta_t$, if $x_{t-1} \geq 1$, or $x_t = 0.4x_{t-1} + \eta_t$, if $x_{t-1} < 1$, $\eta_t \sim NID(0,1)$.

Table 15
Power Against Tent Map DGM
BDS vs. SNT*

BDS	10%	5%	1%
$n = 25$	89.91	80.58	38.71
$n = 50$	100.00	99.96	99.80
$n = 100$	100.00	100.00	100.00
$n = 250$	100.00	100.00	100.00

SNT	10%	5%	1%
$n = 25$	49.24	44.34	12.50
$n = 50$	80.10	76.56	59.54
$n = 100$	93.74	93.18	84.82
$n = 250$	99.98	99.96	99.60

*The statistics reported are percentage rejections in a two-sided test at the $(1 - \alpha)\%$ confidence level using the finite sample critical values in Tables 7 and 8. For the BDS, I set ε equal to one sample standard deviation, and for the SNT, I use the sample mean plus one-hlaf the sample standard deviation. For both tests, I use only $m = 2$. The order of dependence for the SNT is $p = 1$. n is the sample size. The data generating mechanism is the so-called "tent map": $x_t = 2x_{t-1}$, if $x_{t-1} < 0.5$, or $x_t = 2-2x_{t-1}$, if $x_{t-1} > 0.5$.

Table 16
Power When Order of Independence in Mis-Specified*

AR(1)	10%	5%	1%
$n = 25$	18.46	9.00	2.58
$n = 50$	24.36	17.16	6.80
$n = 100$	41.60	28.14	14.98
$n = 250$	73.84	64.50	44.48

MA(1)	10%	5%	1%
$n = 25$	11.84	5.54	1.10
$n = 50$	11.76	7.84	2.74
$n = 100$	12.60	6.82	1.98
$n = 250$	13.00	7.00	1.66

*The statistics reported are percentage rejections in a two-sided test at the $(1 - \alpha)\%$ confidence level using the finite sample critical values in Table 8. The DGM's are the same as used in Table 9 for the AR(1) and Table 10 for the MA(1). I have "mistakenly" set $p = 2$ in each exercise. n is the sample size.