Business Cycles: Theory and Empirical Methods

Edited by
Willi Semmler
New School for Social Research

Kluwer Academic Publishers
Boston / Dordrecht / London

14 USING U-STATISTICS TO DETECT BUSINESS CYCLE NONLINEARITIES

Bruce Mizrach

Nonlinearity is an omnipresent factor in economics. Monetary and fiscal policies change regime, exchange rates change from flexible to fixed and back again. Time series analysts have not been unaware of these nonlinearities; many simply felt that from the perspective of modelling, linear approximations were sufficient. A large part of this malign neglect came from the fact that the nonlinearities remained hidden from the conventional technical apparatus. Many aggregate macro time series are uncorrelated, leading to the conclusion that they were simply random.

The belief that hidden structure might be lurking in the data was strengthened by the growth of the literature on nonlinear dynamics. Many simple, deterministic data generating mechanisms produce time series with the spectra of white noise.

The need for tests to detect a wide range of temporally dependent alternatives was first approached using an apparatus known as the correlation integral. Grassberger and Procaccia (1984) first introduced this technique as a means of estimating fractal dimension.

Brock, Dechert, and Scheinkman (BDS, 1987) put the correlation integral on a firm statistical footing. They recognized that the integral was of the class of U-statistics first rigorously analyzed by Hoeffding (1948) and extended to the time series case by Denker and Keller (1983). This theory is developed in Section 1.

BDS devised a test for independence and idential distribution of a time series which has motivated this work. In Section 2, I discuss the BDS test. The statistic, because of its weak assumptions and resistance to nuisance parameters, has found wide application in the nonlinear time series literature.

The weakness of all tests involving the correlation integral has been their performance in small finite samples. In Monte Carlo analysis in Section 3, I find the size to be extremely poor in samples under 500 observations. This large a sample size requirement has prevented application of the BDS to most economic data, for which there are simply too few observations. At $N = 50$, a conventional economic sample size, the BDS rejects five times more frequently than it should at a conventional 5 percent test.

Brock and Dechert (1989) speculate that "astute choices of kernel function . . . should improve the ability of this [the BDS] test." Their intuition is indeed correct. In Section 4, I review the development a new test proposed in Mizrach (1991), which is computationally simpler and more accurately sized in finite samples.

Mizrach's simple nonparametric test (SNT) statistic extends the range over which the U-statistics can applied. I document in Section 5 that for samples as small as 50 observations, the statistic is within 30 percent of its appropriate nominal size; with the BDS, it takes five times as many observations to achieve a similar size.

Mizrach (1991) suggests that this finite sample improvement does not weaken the power of the BDS. I turn to an empirical application in the last part of the paper to explore this conjecture. I survey the nonlinear time series literature in Section 6 and find that very little work has been done on European business cycles. Consequently, I undertake a study of aggregate macroeconomic time series for France, Italy and Germany. I look at the real (M3) money supply, indices of industrial production, real wages and prices, unemployment, and the current account deficit. Both the BDS and the SNT discover non-GARCH nonlinearities in a number of series.

A concluding section reassesses the SNT and BDS in the context of business cycle applications.

1. U-Statistics: An Introduction

In Section 1.1, I begin with some notational preliminaries and definitions needed for expository purposes. In a second part, I move on to Denker and Keller's extension of asymptotic theory for U-statistics to the time series case.

1.1. Some Preliminaries

U-statistics[1] are generalizations of sample averages. The components include a kernel, a symmetric measurable function $h{:}R^m \to R$, and the permutation operator, $\Sigma_{n,m}$, which sums over the $\binom{n}{m}$ distinct combinations of m-elements in a sample of size n.

Let $\{x_i\}$ be a strictly stationary stochastic process with distribution function F, and let $\{X_1, \cdots, X_n\}$ be a sample of size n. Define the canonical mapping,

$$U_n = U(X_1, X_2, \cdots, X_n) = \binom{n}{m}^{-1}\Sigma_{n,m}\, h(X_1, X_2, \cdots, X_n). \quad (1.1)$$

The U-statistics are a broad class. I begin with two simple examples. Let $m = 1$, $h(x_i) = x_i$, $\binom{n}{m}^{-1} = 1/n$, then

$$U(X_1, \cdots, X_n) = 1/n\sum_{i=1}^{n} X_i = \overline{X} \quad (1.2)$$

or just the sample mean. Now let $m = 2$, $\binom{n}{2}^{-1} = \frac{2}{n(n-1)}$, and $h(x_i - x_j) = (x_i - x_j)^2/2$, one now has,

$$U(X_1,\cdots, X_n) = \frac{2}{n(n-1)} \sum_{1\le i<j} h(X_i - X_j)$$

$$= \frac{1}{(n-1)}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right], \quad (1.3)$$

namely, the sample variance.

1.2. Asymptotic Theory for U-Statistics: The Time Series Case

Hoeffding (1948) was the first to show that U-statistics could be approximated using a projection representation as the sum of i.i.d. random variables. For the analysis of the correlation integral, we will need to consider a vector valued version of (1.1). Let $x_t^m \in X^m$ be a random vector in R^m, and let $F(x_t^m)$ be its joint distribution. Define the new kernel function, $h{:}R^m \times R^m \cdots \times R^m \to R$,

$$h_1(x_1^m) = \int_x \ldots \int_x h(x_1^m, \cdots, x_N^M) \prod_{i=2}^N dF(x_i^m). \tag{1.4}$$

Serfling (1980) notes that the new kernel is just a conditional expectation,

$$h_1(x_1^m) = E[h(X_1^m, \cdots, X_N^m)|X_1^m = x_1^m], \tag{1.5}$$

where the expectation is taken with respect to the distribution function, F. Now taking the expected value of (1.5), define

$$\theta = h_1(x_1^m) = E[h(X_1^m, \cdots, X_N^M)|X_1^m = x_1^m]. \tag{1.6}$$

The difference between (1.5) and (1.6),

$$g_1(x_1^m) = h_1(x_1^m) - \theta, \tag{1.7}$$

has zero expectation, and a martingale representation that is used quite often for the central limit theory of U-statistics. Note that

$$E[g_1(X_1^m)] = E[g_2(x_1^m, X_2^m)] = 0. \tag{1.8}$$

The projection of a U-statistic is defined as

$$\hat{U}_N = \sum_{i=1}^N h_1(x_i^m) - (m-1)\theta. \tag{1.10}$$

Serfling then proves that the asymptotic theory for U_N and $\hat{U}_N$ are essentially equivalent, that is,

$$E[U_N - \hat{U}_N]^v = O(N^{-v}), \tag{1.10}$$

assuming the relevant moments exist.

The projection method has greatly simplified the central limit theory for U-statistics. Denker and Keller (1983) have shown that the variance of a U-statistic is given by

$$\sigma_{U_N}^2 = E[(\sum_{t=1}^N g(X_t^m))^2], \tag{1.11}$$

and under weak mixing assumptions[2] that

$$\frac{N}{m\sigma_{U_N}}(U_N - \theta) \xrightarrow{d} N(0, 1). \tag{1.12}$$

The Denker-Keller result has proved instrumental in the dynamical systems literature. In particular, it is the basis for the test for independence introduced by Brock, Dechert and Scheinkman (1987) to which we know turn.

2. The Correlation Integral and the BDS Test

In the literature on nonlinear analysis, data from dynamical systems have frequently been studied using the correlation integral. Originally proposed by Grassberger and Procaccia (1984) as a means of determining the fractal dimension of an attracting set, this estimator can also be classified as a U-statistic. Consider again vector m-histories,

$$x_t^m = (x_t, x_{t+1}, \cdots, x_{t+m-1}), \tag{2.1}$$

with joint distribution $F(x_t^m)$. Introduce now the kernel,

$$h(x_t^m, x_s^m) = I[\|x_t^m - x_s^m\| < \varepsilon] \equiv I(x_t^m, x_s^m, \varepsilon). \tag{2.2}$$

where I is the indicator (or Heaviside) function, and we take $\|.\|$, for simplicity, to be the max norm,

$$I(x_t^m, x_s^m, \varepsilon) = I[\max_i \prod_{i=0}^{m-1} |x_{t+i} - x_{s+i}| < \varepsilon]. \tag{2.3}$$

The correlation integral is given by

$$C(m, \varepsilon) = \int_X \int_X I(x_t^m, x_s^m, \varepsilon) dF(x_t^m) dF(x_s^m). \tag{2.4}$$

At dimension 1, this simplifies to

$$C(1, \varepsilon) = \int_X \int_X \left[F(x_t - \varepsilon) - F(x_t - \varepsilon)\right] dF(x_t). \tag{2.5}$$

A consistent estimator of (2.4) is the U-statistic,

$$C(m, N, \varepsilon) \equiv \frac{2}{N(N-1)} \sum_{t=1}^{N-1} \sum_{s=t+1}^{N} I(X_t^m, X_s^m, \varepsilon), \tag{2.6}$$

where $N = n - m + 1$. (2.6) is the expected number of m-vectors less than ε away from any given m-vector.

The analysis of statistics based on the correlation integral begins with the assumption that the X's are independent and identically distributed. Define $S(m, N, \varepsilon) \equiv C(m, N, \varepsilon) - C(1, N, \varepsilon)^m$. BDS show that

$$\lim_{N \to \infty} S(m, N, \varepsilon) = 0. \tag{2.7}$$

This follows from the fact that each of the events in the indicator function is a binomial trial with probability $\theta = C(1, N, \varepsilon)$,

$$\text{Prob.}[(\max_i \prod_{i=0}^{m-1} |x_{t+i} - x_{s+i}| < \varepsilon] = \text{Prob.}[|x_t - x_s| < \varepsilon]^m. \tag{2.8}$$

The expression for the variance (2.5) also is more tractable when the data are i.i.d. For the kernel used in (2.2),

$$\mathrm{Var}[S(m,N,\varepsilon)] = 4[2\sum_{j=1}^{m-1} K^{m-j}C(1,N,\varepsilon)^j + K^m + (m-1)^2 C(1,N,\varepsilon)^{2m} - m^2 K C(1,N,\varepsilon)^{2(m-1)}], \quad (2.9)$$

where

$$K = \int_X\int_X\int_X I(x_t, x_s, \varepsilon) I(x_s, x_r, \varepsilon) dF(x_s) dF(x_r). \quad (2.10)$$

A consistent estimator of K is

$$K(N,\varepsilon) = \frac{6}{N(N-1)(N-2)} \sum_{t=1}^{N-2}\sum_{s=t+1}^{N-1}\sum_{r=s+1}^{N} I(X_t, X_s, \varepsilon) I(X_s, X_r, \varepsilon). \quad (22.1)$$

Noting the independence at dimension 1 of the m-vectors,

$$E[I(x_t, x_s)I(x_s, x_r)] = E[I(x_t, x_s)]^2, \quad (2.12)$$

a computationally simpler, but still consistent estimator is,

$$K(N,\varepsilon) = \sum_{t=1}^{N}\left(\sum_{s\neq t}^{N} I(X_t, X_s, \varepsilon)\right)^2 / [N(N-1)(N-2)]. \quad (2.13)$$

It has been shown by Brock, Dechert and Scheinkman (1987) that

$$\sqrt{N}\frac{S(m,N,\varepsilon)}{\sqrt{\mathrm{Var}[S(m,N,\varepsilon)]}} \xrightarrow{d} N(0,1) \quad (2.14)$$

This has become a widely applied statistic in the literature as a test for independence. Given its ability to shrug off nuisance parameters,[3] the BDS has become a powerful portmanteau statistic for model specification.

3. The Finite Sample Properties of the BDS

While the BDS has found wide application, a principal problem with using the statistic in data analysis is that it is poorly sized in finite samples. I document the empirical distribution in 3.1 for the normal distribution[4] and find high rates of Type I error in small samples.

One rather striking result is the negative bias of the statistic in samples under 250 observations. In the left tail of the empirical distribution, the size of the BDS deviates substantially more from the nominal size than it does in the right tail. Controlling for the bias, the statistic is actually skewed to the right though. Some theoretical discussion of the finite sample bias is pursued in Section 3.2.

3.1. Monte Carlo Analysis: The Standard Normal Distribution

I look at sample sizes of 50, 100, 250, 500, and 1,000 observations, and dimensions from $m = 2$ to 5. Brock, Hsieh and LeBaron (1992) found that the size of the BDS was least distorted setting ε equal to one sample standard deviation; I used that value throughout the paper. N(0, 1) i.i.d. pseudo-random numbers were generated using the IMSL subroutine RNNOF with the largest congruential multiplier.[5] The generator is seeded so that the same random numbers are used throughout all the Monte Carlo exercises in the paper. Results may be found in Table 1.

The pattern of the first three entries of Table 1 can be summarized fairly succinctly. For samples under 500 observations, the smaller the sample size, the larger the empirical size of the test. This holds true at all dimensions m. For a given sample size, the larger is m, the greater is the rate of Type I error.

The distortions at large m are due to the so-called "curse of dimensionality." Setting ε equal to one sample standard deviation should yield, on average, the population expectation (obtained by numerical integration) for $C(1, \varepsilon)$ of 0.52. At $m = 5$ though, we should expect to find only $(.52)^m$ of the sample, or merely 3.8 percent of all observations.[6]

The magnitude of the Type I error rates is substantial. At $N = 50$, a nominal 10 percent tests rejects more than 33 percent of the time at $m = 2$, and almost 40 percent of the time at $m = 5$. Nominal 5 percent tests at $N = 50$ reject more than five times too often at $m = 2$, and six times too often at $m = 5$. Though not reported, the size distortions are even more substantial for more entropic distributions like the uniform.

The empirical sizes of the test do drop fairly rapidly. At $N = 250$, the 10 percent test is sized at less than 16 percent for all choices of m. Similarly, no 5 percent test is larger than 8.5 percent. For $N = 500$ and $N = 1000$, the BDS is a highly reliable statistic, even in the extreme tails.

3.2. An Analysis of the Bias and Skewness in the BDS

Our discussion to this point has focused on the nominal size of symmetric confidence intervals. The portrait of Table 1 is far more complicated if we look at the left and right tails of the statistic separately.

Table 1. Monte Carlo Analysis of the BDS Statistic* Standard Normal Distribution.

$N = 50$	% < –2.33	% < –1.96	% < –1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0942	0.1444	0.1936	0.1444	0.1132	0.0792
$m = 3$	0.0998	0.1502	0.2028	0.1468	0.1166	0.0876
$m = 4$	0.1096	0.1590	0.2182	0.1512	0.1222	0.0960
$m = 5$	0.1178	0.1676	0.2318	0.1594	0.1324	0.1056
$N(0, 1)$	0.0100	0.0250	0.0500	0.0500	0.0250	0.0100
$N = 100$	% < –2.33	% < –1.96	% < –1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0352	0.0692	0.1158	0.0976	0.0670	0.0400
$m = 3$	0.0368	0.0777	0.1238	0.1080	0.0676	0.0416
$m = 4$	0.0378	0.0800	0.1328	0.1020	0.0736	0.0494
$m = 5$	0.0470	0.0852	0.1374	0.1078	0.0786	0.0568
$N = 250$	% < –2.33	% < –1.96	% < –1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0136	0.0364	0.0730	0.0688	0.0380	0.0208
$m = 3$	0.0134	0.0364	0.0746	0.0710	0.0478	0.0260
$m = 4$	0.0124	0.0370	0.0742	0.0744	0.0476	0.0270
$m = 5$	0.0138	0.0350	0.0800	0.0772	0.0498	0.0292
$N = 500$	% < –2.33	% < –1.96	% < –1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0088	0.0280	0.0584	0.0560	0.0312	0.0160
$m = 3$	0.0092	0.0320	0.0672	0.0604	0.0356	0.0164
$m = 4$	0.0104	0.0284	0.0628	0.0640	0.0376	0.0184
$m = 5$	0.0084	0.0284	0.0648	0.0652	0.0408	0.0216
$N = 1000$	% < –2.33	% < –1.96	% < –1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0092	0.0308	0.0640	0.0544	0.0284	0.0132
$m = 3$	0.0112	0.0288	0.0592	0.0536	0.0324	0.0148
$m = 4$	0.0124	0.0300	0.0600	0.0632	0.0344	0.0132

* N = the sample size, and m = embedding dimension. The simulations for N = 50, 100, 250 are based on 5000 replications. For N = 500 and N = 1000, I used 2500 replications. ε is one sample standard deviation.

For $N = 50$ and 100, the BDS statistic is biased down from zero. The statistic is about 1.5 times as likely to violate a lower 90 percent or 95 percent critical value, as the corresponding upper critical value. This bias is more pronounced at higher dimensions. At $N = 50$ and $m = 2$, 19.36 percent of the time the statistic takes on a value less than –1.64, but only 14.44 percent of the time does it exceed 1.64. At $m = 5$, the bias has expanded to a 23.18 percent rejection at –1.64 and 15.94 percent at 1.64.

At $N = 250$, the BDS is essentially balanced for 10 percent tests. The rejection frequencies are within sampling error, though still a bit too large

Table 2. Critical Values for the BDS Statistic* $N(0, 1)$ Random Variables

$N = 50$	0.010	0.025	0.050	0.950	0.975	0.990
$m = 2$	–6.2303	–3.5823	–2.8964	2.9174	3.7507	5.7867
$m = 3$	–6.5220	–3.6787	0.1238	3.1383	3.9996	6.6066
$m = 4$	–6.5871	–3.9201	0.1328	3.3287	4.3240	6.8051
$m = 5$	–6.6698	–4.0927	0.1374	3.6838	4.7889	7.7884
$N(0, 1)$	–2.33	–1.96	–1.64	1.64	1.96	2.33
$N = 100$	0.010	0.025	0.050	0.950	0.975	0.990
$m = 2$	–3.2091	–2.5096	–2.1630	2.1607	2.6403	3.6753
$m = 3$	–3.2480	–2.5151	–2.1616	2.2227	2.7991	3.9559
$m = 4$	–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
$N = 250$	0.010	0.025	0.050	0.950	0.975	0.990
$m = 2$	–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
$m = 4$	–2.6011	–2.0931	–1.8299	1.9133	2.4257	3.2927
$m = 5$	–2.6547	–2.0894	–1.8178	1.9681	2.5199	3.5256

* N is the sample size, m is the dimension of the test. Along the top row are the alpha levels, and below them are the empirical critical values that size the test to the appropriate alpha level. ε is one sample standard deviation. I use 5,000 replications.

overall. For 5 percent tests, the imbalance actually shifts slightly to the right for $N \geq 250$.

Rejections in the right tail seem to involve very strong outliers. The 1 percent tests in the right tail should reject only one-fifth as often as the 5 percent test. For all values of m and $N < 1{,}000$, a 1 percent test is instead one-half as likely to reject as a 5 percent test. The strong outliers in the right tail contribute to a rather counterintuitive result. Although the statistic rejects more frequently in the left tail than the right, the empirical critical values needed to properly size the BDS are actually *larger* in the right tail than the left. In Table 2, I provide the empirical quantiles of the BDS for $N = 50$, 100 and 250 for the standard normal distribution. With only one exception ($N = 50$, $m = 2$ for a 2 percent test), the critical values are larger in the right tail of the distribution.

Clearly something more complicated is going on than a monotone convergence to normality. At small sample sizes, the skewness in the BDS is dominated by the bias in the numerator. The sample average of $S(m, N, \varepsilon) - C(m, N, \varepsilon) - C(1, N, \varepsilon)^m$ is negative. Note that $E[C(1, N, \varepsilon)] = C(1, \varepsilon)$, and $E[C(m, N, \varepsilon)] = C(1, \varepsilon)^m$. Therefore,

$$E[C(1, N, \varepsilon) - C(1, \varepsilon)^m] = 0. \qquad (3.1)$$

Table 3. Bias and Skewness Diagnostics*.

$N = 50$	Numerator	# > 0	# < 0	BDS	Pos.Std.Dev.	Neg.Std.Dev.
$m = 2$	−0.0015	2149	2851	−0.1883	0.0496	0.0509
$m = 3$	−0.0022	2101	2899	−0.2925	0.0591	0.0611
$m = 4$	−0.0022	2032	2968	−0.2820	0.0523	0.0562
$m = 5$	−0.0017	2006	2994	−0.3761	0.0410	0.0455
$N = 100$	Numerator	# > 0	# < 0	BDS	Pos.Std.Dev.	Neg.Std.Dev.
$m = 2$	−0.0007	2196	2804	−0.1254	0.0530	0.0534
$m = 3$	−0.0010	2146	2854	−0.1461	0.0626	0.0634
$m = 4$	−0.0010	2074	2926	−0.1608	0.0553	0.0570
$m = 5$	−0.0008	2039	2961	−0.1585	0.0432	0.0449
$N = 250$	Numerator	# > 0	# < 0	BDS	Pos.Std.Dev.	Neg.Std.Dev.
$m = 2$	−0.0002	2271	2729	−0.0722	0.0547	0.0546
$m = 3$	−0.0003	2223	2777	−0.0849	0.0645	0.0644
$m = 4$	−0.0003	2219	2781	−0.0935	0.0572	0.0569
$m = 5$	−0.0003	2176	2824	−0.0964	0.0444	0.0442
$N = 500$	Numerator	# > 0	# < 0	BDS	Pos.Std.Dev.	Neg.Std.Dev.
$m = 2$	−0.0002	442	558	−0.0655	0.0549	0.0549
$m = 3$	−0.0002	452	548	−0.0759	0.0646	0.0646
$m = 4$	−0.0002	464	536	−0.0690	0.0571	0.0571
$m = 5$	−0.0001	458	542	−0.0627	0.0441	0.0442

Numerator = $C(m, N, 1.0) - C(1, N, 1.0)^m$ Std.Dev. = $\sqrt{\text{Eq.}(2.13)}$

* N is the sample size. For $N = 50$, 100, and 250, the sample averages are for 5000 replications. For $N = 500$, there are 1000 replications. The first statistic is the sample average of the numerator. The next two statistics are the number of the times the numerator was positive and negative. Pos.Std.Dev. is the average standard deviation when the numerator is positive, and Neg.Std.Dev. is the average when its negative. I set $\varepsilon = 1.0$. The input is $N(0, 1)$ noise.

From Jensen's inequality though,

$$E[C(1, N, \varepsilon)^m] > (E[C(1, N, \varepsilon)])^m = C(1, \varepsilon)^m. \quad (3.2)$$

Substituting back into (3.1),

$$E[C(m, N, \varepsilon) - C(1, N, \varepsilon)^m] < E[C(m, N, \varepsilon) - C(1, \varepsilon)^m] = 0, \quad (3.3)$$

yielding the negative bias in the numerator of (2.14)[7].

As shown in Table 3, this bias[8] is fairly substantial at $N = 50$, and $N = 100$. Even at $m = 2$, nearly 60 percent of the statistics are negative. The bias diminishes fairly rapidly and is no greater than -3.5×10^{-4} for $N > 250$.

Once we divide by the standard deviation, the bias in the BDS statistic itself is amplified. At $N = 50$ and $m = 5$, the sample mean of the BDS statistic in 5,000 replications is −0.3761. The bias falls by half when the sample size is doubled, and for $N > 250$, the bias in the BDS in greater than −0.1.

At the small sample sizes, the standard deviation is actually mitigating the bias. Note that for $N = 50$ and $N = 100$, the standard deviation when $S(m, N)$ is positive is smaller than when it is negative. While the statistic rejects far too often in both tails, the negative skewness would be even more pronounced if the variances were balanced. The use of $C(1, N)^m$ in place of the $C(m, N)$ in constructing the variance diminishes the finite sample bias of the BDS.

When $S(m, N)$ exceeds zero, it is either because $C(m, N)$ is exceptionally big or because $C(1, N)$ is exceptionally small. It is the latter case which effects the variance. I numerically integrated the correlation integral $C(1, \varepsilon)$ with $\varepsilon = 1.0$, and arrived at an estimate of 0.5204. At $N = 50$ with $m = 2$, the sample average of $C(1, 50)$ is 0.5172 when $S(m, N)$ is positive and 0.5231 when it is negative. $K(N)$ moves monotonically with C. The sample average of $K(50)^m$ is 0.0901 when the BDS is positive and 0.0947 when it's negative. A similar pattern holds for both C and K at all dimensions in the two small samples.

A detailed decomposition of the variance reveals why the standard deviation falls when $S(m, N)$ is positive. In the variance, there are four terms in (2.9) of order $o(C^m)$ or larger. While one of the four terms enters with a negative sign, $-m^2KC^{2m-2}$, the other three dominate. When the BDS takes on the rare positive value, the estimator for the variance magnifies a small realization of $C(1, N)$ exponentially. At $N = 50$, the standard deviation is about 3 percent smaller at $m = 2$ and 11 percent smaller at $m = 5$. By $N = 100$, the differences have fallen below 5 percent.

These outlying realizations explain the pattern of critical values in Table 2 as well. For the BDS to be positive, ceteris parabus, $C(1, N)$ must be unusually small, but this in turn biases the variance downward. The combination of the makes the ratio of 1 percent to 5 percent right tail rejections quite close (about 1:2, rather 1:5). Once centered at its' sample mean, the BDS is actually skewed to the *right*, not to the left. Consider $m = 2$ where the bias in the numerator is the smallest. At $N = 50$, the skewness is 0.3647, and at $N = 250$ it is still 0.3318. The skewness in the standard deviation diminishes more quickly, falling from 1.1317 at $N = 50$ to 0.0745 at $N = 250$. Fat tails on both left and right give a leptokurtic distribution, with excess kurtosis of 1.1962 at $N = 50$ and 0.1455 at $N = 250$.

4. The Simple Nonparametric Test (SNT)

In this section, I propose a new U-statistic similar to the BDS. It has three main advantages. It is computationally simpler, involving only calculations of order N, as opposed to N^2 for the standard BDS. The variance of the U-statistic I derive is similar to that of a binomial random variable. Third, it will be documented in Section 5, the statistic is properly sized in samples as small as 50.

Our principal modification to the BDS is the kernel function. Replace (2.2) with the function $h{:}R \to R$,

$$h(x_t) = I[x_t < \varepsilon] = \begin{Bmatrix} 1, \text{ if } x_t < \varepsilon \\ 0, \text{ otherwise} \end{Bmatrix} \equiv I(x_t, \varepsilon). \quad (4.1)$$

The correlation integral at dimension m is given by

$$\theta(m, \varepsilon) = \int_X \prod_{i=1}^{m-1} I(x_{t+i}, \varepsilon)\, dF(x_{t+i}). \quad (4.2)$$

A consistent estimator of (4.2) is

$$\theta(m, N, \varepsilon) = \sum_{t=1}^{N} \prod_{i=0}^{m-1} I(X_{t+i}, \varepsilon)/N. \quad (4.3)$$

Since the kernel (4.1) depends only on a single realization, the correlation integral sums, under the assumption of i.i.d., chains of m-independent events. This simplifies the analysis considerably.

The expected number of m-chains with a value of 1 in a sample of size N is

$$\mu = N\theta(m, \varepsilon) = \sum_{x=0}^{N} x \binom{N}{x} \theta(m, \varepsilon)^x (1 - \theta(m, \varepsilon))^{N-x} \quad (4.4)$$

The variance is given by

$$\sigma^2_{B_N} = \mu_2' - \mu^2 = \sum_{x=0}^{N} x^2 \binom{N}{x} \theta(m, \varepsilon)^x (1 - \theta(m, \varepsilon))^{N-x} - N^2\theta(m, \varepsilon)^2$$

$$= N\theta(m, \varepsilon)(1 - \theta(m, \varepsilon)). \quad (4.5)$$

A simple test for i.i.d. can then be constructed using consistent estimators of these two moments. Mizrach (1991) proves that the following statistic has an asymptotic normal distribution,

$$\sqrt{N} \frac{[\theta(m, N, \varepsilon) - \theta(m-1, N, \varepsilon)\theta(1, N, \varepsilon)}{[\theta(m-1, N, \varepsilon)\theta(1, N, \varepsilon)(1 - \theta(m-1, N, \varepsilon))(1 - \theta(1, N, \varepsilon))} \xrightarrow{d} N(0, 1). \quad (4.6)$$

Table 4. Monte Carlo Analysis of SNT* $N(0,1)$ Random Variables.

$N = 50$	% < −2.33	% < −1.96	% < −1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0236	0.0680	0.0757	0.0504	0.0418	0.0162
$m = 3$	0.0150	0.0412	0.0776	0.0384	0.0176	0.0078
$m = 4$	0.0060	0.0264	0.0754	0.0288	0.0126	0.0036
$m = 5$	0.0060	0.0080	0.0342	0.0238	0.0102	0.0028
$N(0, 1)$	0.0100	0.0250	0.0500	0.0500	0.0250	0.0100
$N = 100$	% < −2.33	% < −1.96	% < −1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0200	0.0512	0.0752	0.0526	0.0378	0.0154
$m = 3$	0.0146	0.0330	0.606	0.0380	0.0200	0.0082
$m = 4$	0.0130	0.0376	0.0682	0.0350	0.0156	0.0062
$m = 5$	0.0042	0.0200	0.0584	0.0316	0.0152	0.0050
$N = 250$	% < −2.33	% < −1.96	% < −1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0136	0.0276	0.0682	0.0590	0.0236	0.0120
$m = 3$	0.0142	0.0292	0.0568	0.0500	0.0218	0.0078
$m = 4$	0.0118	0.0308	0.0630	0.0416	0.0196	0.0070
$m = 5$	0.0130	0.0310	0.0692	0.0340	0.0176	0.0072
$N = 500$	% < −2.33	% < −1.96	% < −1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0146	0.0310	0.0538	0.0472	0.0300	0.0106
$m = 3$	0.0130	0.0324	0.0576	0.0460	0.0226	0.0108
$m = 4$	0.0130	0.0294	0.0614	0.0428	0.0224	0.0088
$m = 5$	0.0128	0.0310	0.0614	0.0400	0.0194	0.0080
$N = 1000$	% < −2.33	% < −1.96	% < −1.64	% > 1.64	% > 1.96	% > 2.33
$m = 2$	0.0120	0.0256	0.0572	0.0474	0.0212	0.0096
$m = 3$	0.0128	0.0310	0.0602	0.0484	0.0246	0.0082
$m = 4$	0.0138	0.0310	0.0606	0.0430	0.0196	0.0072
$m = 5$	0.0114	0.0314	0.0618	0.0436	0.0188	0.0078

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

While we have stated many advantages for the statistic (4.6) at the outset, the principal advantage will be if (4.6) can improve substantially over the standard construct, (2.14), in finite samples. I turn to that in the next section.

5. Finite Sample Properties of the SNT

The Monte Carlo analysis of the SNT (4.6) follows the same design as in Section 3. Results are in Table 4. I found in Mizrach (1991) that good power was achieved setting ε equal to the sample mean. The random number generator provides the identical random numbers used in Tables 1 and 2. Since calculations here are of order N, I have used 5,000 replications at all sample sizes.

For all sample sizes and dimensions, the SNT is more accurately sized. This is especially true in the two small samples. At $N = 50$, $m = 2$, (4.6) rejects 12.6 percent of the time at ± 1.64, compared to 33.8 percent for the BDS. At $m = 4$, the advantage is 10.4 percent to 58.5 percent for the BDS. At $N = 100$, the SNT still enjoys a nearly 3:1 advantage.

In 5 percent tests, the new statistic measures in at 10.9 percent at $N = 50$ with $m = 2$, and 8.9 percent at $N = 100$. The corresponding results for the BDS from Table 1 are 25.8 percent and 13.6 percent respectively. The advantage in both small samples is again nearly threefold.

The SNT betters the standard BDS at all the remaining sample sizes, though the margin of victory grows slimmer. For $N > 250$, the differences can be attributed solely to sampling error. Overall, these modifications result in a fairly dramatic improvement in finite sample performance.

I now turn to the empirical part of the paper, taking a brief detour to survey the existing literature.

6. A Brief Survey of the Literature

While there have been a number of studies on U.S. economic and financial time series,[9] there have been few examinations of business cycle data from a nonlinear perspective for Europe. This gap in the literature is especially large because many observers have claimed that U.S. data may be atypical. Blanchard and Summers (1986) note differences in labor markets in their explanation of high rates of European unemployment. de Jong and Shepard (1986) cite differences in the market structure of the United States.

Frank, Gencay, and Stengos (1988) were the first to search for "chaos" in European macroeconomic aggregates. The authors found evidence of low dimensional structure in Italian, British, and West German national income series after filtering with ARMA and GARCH models. Frank, Sayers and Stengos (1992) have isolated nonlinear structure in Canadian unemployment rates. Stevenson, Jones and Manning (1992) employ threshold models for unemployment data in the United Kingdom.

Exchange rate modeling has been a particularly fruitful area for nonlinear approaches. The leptokurtic ("fat-tailed") distributions of spot exchange rates have been analyzed by Hsieh (1989). Mizrach (1992) uses nearest neighbor methods to forecast EMS exchange rates. De Gooijer (1989) looks at world stock returns. None of these papers though has looked beyond GARCH effects.

In summing up the U.S. evidence, it seems fair to conclude that nonlinearities not captured by the GARCH filter are in macroeconomic time series. In the next section, I see if the United states evidence extends to Europe or whether the United States is indeed atypical.

7. An Application to European Macroeconomic Time Series

This application attempts to answer whether the size improvements of the SNT come at the expense of power. I compare the tests using European macroeconomic time series.

7.1. Data

I collected monthly series for France, Germany, and Italy for the period March 1979 to June 1992. This is the period in which many countries throughout Europe returned to a fixed exchange rate regime within the European Monetary System (EMS). I look at the broad M3 equivalent money supply, deflated by the GDP price deflator, indices of unit labor costs and industrial production, the GDP deflator and the unemployment rate. The data are from public sources in each country and collected at the Federal Reserve Bank of New York.

A previous study by Mizrach (1992) had difficulty detecting any fundamental nonlinearities in exchange rates from that period. It remains to be seen whether they show up in the macro data.

7.2. Empirical Results

To analyze nonlinear dependence, I first wanted to remove any *linear* dependence in the data. I fit ARMA $(p, 1, 0)$ models, choosing p so as to minimize the Bayesian Information Criterion (BIC),

$$\log(\hat{\sigma}^2) = p \log(N)/N, \tag{7.1}$$

where $\hat{\sigma}^2$ is the estimated variance and N is the sample size. The BIC penalizes the likelihood for extra parameters.

In Table 5, I use the BDS and SNT to analyze the *residuals* from the

Table 5. Tests for Nonlinear Dependence* AR(p) Filtering

France				
Variable	BDS	SNT	ARCH [q]	Filter
Real M3	1.461	2.006^	0.005 [q=1]	AR(2)
Indus. Prod.	4.814	3.478^	12.829 [q=2]	AR(2)
Current Acct.	2.426	1.621	22.128 [q=3]	AR(1)
Unit Labor	−1.608	3.833	18.128 [q=3]	AR(3)
Prices	−1.013	0.045	0.001 [q=1]	AR(2)
Unemployment	—	—	—	—
Germany				
Variable	BDS	SNT	ARCH [q]	Filter
Real M3	7.650	4.133	0.025 [q=1]	AR(1)
Indus. Prod.	3.741	2.396^	10.801 [q=1]	AR(3)
Current Acct.	−1.331	0.071	9.377 [q=2]	AR(4)
Unit Labor	1.456	0.414	10.532 [q=2]	AR(3)
Prices	0.049	−0.006	0.824 [q=1]	AR(9)
Unemployment	−3.607	−3.146^	0.290 [q=1]	AR(1)
Italy				
Variable	BDS	SNT	ARCH [q]	Filter
Real M3	0.018	1.971^	0.615 [q=1]	AR(2)
Indus. Prod.	−0.049	0.015	0.009 [q=1]	AR(2)
Current Acct.	−1.331	−3.136	1.110 [q=1]	AR(3)
Unit Labor	1.456	−0.069	0.015 [q=1]	AR(2)
Prices	0.049	−0.044	1.054 [q=1]	AR(9)
Unemployment	−3.607	−3.026	37.559 [q=9]	AR(8)

* I used the Bayesian Information Criterion to select the order of the AR filter. I set the dimension $m = 2$. 95 percent critical values for the BDS at this sample size are approximately −2.51, 2.64. At this sample size, you can use the asymptotic distribution of the SNT, the $N(0, 1)$, indicating ± 1.96 for critical values. The ARCH statistic is Nr^2, which has an asymptotic $X^2(q)$ distribution. A ^ indicates the residuals were squared in the SNT test.

linear ARMA filters. The BDS and SNT find temporal dependence in French industrial production, German real money balances, industrial production and unemployment, and Italian unemployment. The SNT finds nonlinear dependence in four series that the BDS does not: French real money balances and unit labor costs, and Italian real money balances and the current account. The BDS rejects i.i.d. for the French current account when the SNT does not.

It might be argued that the U-statistics are simply picking up dependence

Table 6. Tests for Nonlinear Dependence* ARCH(q) Filtering

France			
Variable	BDS	SNT	Filter
Indus. Prod.	2.527	3.426^	ARCH(2)
Current Acct.	3.399	3.864^	ARCH(3)
Unit Labor	−0.607	4.134	ARCH(3)
Germany			
Variable	BDS	SNT	Filter
Indus. Prod.	1.229	−0.107	ARCH(1)
Current Acct.	−0.081	−0.691	ARCH(2)
Unit Labor	5.021	5.106^	ARCH(2)
Italy			
Variable	BDS	SNT	Filter
Unemployment	−0.153	3.572	ARCH(9)

* I used the Bayesian Information Criterion to select the order of the ARCH filter. I set the dimension $m = 2$ for the U-statistics. 95 percent critical values for the BDS at this sample size are approximately −2.51, 2.64. At this sample size, you can use the asymptotic distribution of the SNT, the $N(0, 1)$, indicating ± 1.96 for critical values. A ^ indicates the residuals were squared in the SNT test.

in second moments from autoregressive conditional heteroscedasticity (ARCH). I used the Lagrange multiplier test for ARCH(q) dependence. The statistic, Nr^2, where N is the sample size and r^2 the goodness of fit for an OLS regression on q ARCH lags along with a constant, is distributed $X^2(q)$. I find ARCH dependence in three French series, three German series, and one Italian series. Only in one case, German labor costs, do I find ARCH effects when neither the BDS nor the SNT reject. The ARCH-LM test fails to find dependence in three cases that are caught by the U-statistics.

I wanted to find out how many of the rejections in Table 5 were the results of non-ARCH temporal dependence. I fit ARCH(q) models to the AR(p) residuals and re-tested the ARCH residuals using the U-statistics. I only fit ARCH models to those series that had significant ARCH effects in the LM-test, and I chose the order q of the ARCH regression using the BIC. Results of the second round of tests are in Table 6.

The BDS and SNT both reject for the French current account (which had been previously missed by the SNT), and German labor costs. The last case is interesting since neither the BDS nor the SNT rejected with the

ARMA residuals. Dependence in German industrial production and current account does seem to be due to ARCH effects because neither the SNT nor the BDS find additional temporal dependence.

The SNT again shows its power with the ARCH residuals. It detects three cases of nonlinear dependence that the BDS misses: French industrial production and labor costs and Italian unemployment. Conversely, the BDS does not catch any nonlinearity that the SNT misses.

8. Conclusion

By achieving reasonable results in samples as small as 50, the new simple nonparametric test (SNT) extends the useful range of U-statistic inference. Economic data, even those at an annual frequency, can be analyzed using the asymptotic distribution of the SNT.

Our analysis of the bias and skewness in the BDS should prove useful in understanding the power of tests based on the correlation integral. A negative bias though is bad for detecting processes with serial correlation, like many economic time series.

Mizrach (1991) examines the power of the SNT for a number of data generating mechanisms. This paper extends that work by looking at real economic data from the EMS. The SNT successfully uncovers nonlinearity in six cases that the BDS misses, while the BDS only captures one case that the SNT does not. Our real world experiments bear out Mizrach's conjecture that the SNT offers improvements in both size *and* power.

Acknowledgments

I would like to acknowledge several useful discussions on the BDS with Dee Dechert, Blake LeBaron, and Philip Rothman. I would also like to thank Willi Semmler for helpful comments on an earlier draft. The views presented are solely those of the author and do not reflect the views of the New York Federal Reserve Bank or the Board of Governors of the Federal Reserve.

Notes

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

2. Denker and Keller (1983) prove the theorem providing that any one of the following three conditions holds: (i) $\{X_N\}$ is uniformly mixing in both directions of time, $\sigma^2_{U_N} \to \infty$, and

for some $\delta > 0$, $\sup_{1 \le t_1 < t_2 < \cdots < t_m} E|h(x_{t_1}, \cdots, x_{t_m})|^{2+\delta}; < \infty$; (ii) $\{X_N\}$ is uniformly mixing in both directions of time with mixing coefficients satisfying $\Sigma\phi(N) < \infty$, $\sigma^2 \neq 0$, and $\sup_{1 \le t_1 < t_2 < \cdots < t_m} E|h(x_{t_1}, \cdots, x_{t_m})|^2 < \infty$; (iii) $\{X_N\}$ is absolutely regular with coefficients $\beta(N)$ satisfying $\Sigma\beta(N)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$, $\sigma^2 \neq 0$, and $\sup_{1 \le t_1 < t_2 < \cdots < t_m} E|h(x_{t_1}, \cdots, s_{t_m})|^{2+\delta} < \infty$.

3. Brock and Dechert (1989) show that (2.14) holds for residuals from a $\sqrt{N}$ consistent estimate of a nonlinear autoregressive process.

4. A wider Monte Carlo investigation with qualitatively similar results can be found in the monograph of Brock, Hsieh and LeBaron (1992) and also in Mizrach (1991).

5. IMSL generates random numbers using a multiplicative congruential method. The form of the generator is $x_t = cx_{t-1} \bmod(2^{31} - 1)$. Results differed substantially with choice of c, but RNNOF had the best small sample properties with $c = 950706376$ and no shuffling. Fishman and Moore (1986) discuss the optimality properties of this choice of c.

6. The curse also arises in nonparametric density estimation where rates of convergence for kernel estimators typically depend on the dimension of the distribution. See Stone (1980) for some theoretical results.

7. I thank Dee Dechert for this point.

8. I fixed $\varepsilon = 1.0$ in Table 3, so I will hereafter suppress the argument.

9. See C. L. Sayers' survey, "Testing for Chaos and Nonlinearities in Macroeconomic Time Series," in this volume.

References

Blanchard, O., and Summers, L. 1986. "Hysteresis and the European Unemployment Problem." *NBER Macroeconomics Annual 1986* Cambridge: MIT Press.

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

Brock, W. A., and Dechert, W. D. 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., Hsieh, D. A., and LeBaron, B. 1992. *Nonlinear Dynamics, Chaos and Instability*. Cambridge: MIT Press.

De Gooijer, J. G. 1989. "Testing Non-Linearities in World Stock Market Prices." *Economics Letters* 31, 31–35.

De Jong, H. W., and Shepard, W. G. (eds.), 1986. *Mainstreams in Industrial Organization*, Books 1 and 2. Dordrecht: Nijhoff.

Denker, M., and Keller, G. 1983, "On U-Statistics and v. Mises Statistics for Weakly Dependent Processes." *Z. Warsch. verw. Gebiete* 65, 505–22.

Denker, M., and Keller, G. 1986. "Rigorous Statistical Procedures for Data from Dynamical System." *Journal of Statistical Physics* 44, 1986, 67–93.

Fishman, G. F., and Moore III, L. R. 1986. "An Exhaustive Analysis of Multiplicative

Congruential Random Number Generators with Modulus $2^{31} - 1$." *SIAM Journal on Scientific and Statistical Computing* 7, 24–45.
Frank, M. Z., Gencay, R. and Stengos, T. 1988. "International Chaos?" *European Economic Review* 32, 1569–1584.
Frank, M. Z., Sayers, C. L., and Stengos, T. 1992. "Evidence Concerning Nonlinear Structure in Canadian Provincial Unemployment Rates." *Structural Change and Economic Dynamics*, forthcoming.
Grassberger, P., and Procaccia, I. 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.
Hsieh, D. 1989. "Testing for Nonlinearity in Daily Foreign Exchange Rate Changes." *Journal of Business* 62, 339–68.
Mizrach, B. 1991. "A Simple Nonparametric Test for Independence." Working Paper, Department of Finance, The Wharton School.
Mizrach, B. 1992. "Multivariate Nearest-neighbour Forecasts of EMS Exchange Rates." *Journal of Applied Econometrics* 7 Supplement, S151–63.
Sayers, C. L. 1992. "Testing for Chaos and Nonlinearities in Macroeconomic Time Series." Working Paper, University of Houston.
Serfling, R. 1980. *Approximation Theorems of Mathematical Statistics*, New York: Wiley.
Stevenson, M., Jones, D., and Manning, N. 1992. "The Unemployment rate and the Business Cycle in Britain: An Aggregate and Regional Analysis." Working Paper, University of Technology, Sydney, Australia.
Stone, R. 1980. "Optimal Rates of Convergence for Nonparametric Estimators." *Annals of Statistics* 8, 1348–60.

15 THE TIME REVERSIBILITY TEST WITH APPLICATION TO FINANCIAL DATA

Philip Rothman

1. Introduction

Asset market data provide information on intertemporal marginal rates of substitution and transformation. A great deal of work has been done studying the time series behavior of returns on bonds and stocks and the implications of such behavior as a check on the consistency of intertemporal asset pricing models. Most research has focused on tests of consumption-based asset pricing models. More recently, Cochrane (1991) pioneered the development and testing of production-based asset pricing models.

Given this important and voluminous literature produced by financial economists, Cochrane and Hansen (1992) noted a stylized fact of equilibrium business cycle research which they called "surprising." Specifically, asset prices are often ignored in assessing and evaluating the performance of dynamic macroeconomic models. They argued that new utility functions developed in the finance literature to account for anomalous features of asset pricing data may substantially change the dynamic properties of equilibrium business cycle models and the welfare analysis carried out with them. As such, advances made in understanding the time series behavior of asset prices can have very important implications for macroeconomic modeling.