A TEST FOR INDEPENDENCE BASED ON THE CORRELATION DIMENSION

W. A. Brock

Department of Economics
University of Wisconsin

W. D. Dechert

Department of Economics
University of Houston

J. A. Scheinkman

Department of Economics
University of Chicago

B. LeBaron

Department of Economics
University of Wisconsin

ABSTRACT

This paper presents a test of independence that can be applied to the estimated residuals of any time series model that can be transformed into a model driven by independent and identically distributed errors. The first order asymptotic distribution of the test statistic is independent of estimation error provided that the parameters of the model under test can be estimated $\sqrt{n}$–consistently. Because of this, our method can be used as a model selection tool and as a specification test. Widely used software[1] written by Dechert and LeBaron can be used to implement the test. Also, this software is fast enough that the null distribution of our test statistic can be estimated with bootstrap methods. Our method can be viewed as a nonlinear analog of the Box–Pierce Q statistic used in ARIMA analysis.

---

[1] Available via Gopher at: `gopher.econ.wisc.edu`

1   Introduction

The original BDS[2] paper was motivated by the special problems raised by designing a test on time series data to detect whether such data came from a (possibly noisy) chaotic data generation process. The BDS statistic has its origins in the recent work on deterministic nonlinear dynamics and chaos theory. Not only is the resulting test useful in detecting deterministic chaos, but it also serves as a residual diagnostic that can be used to test the "goodness of fit" of an estimated model. Furthermore the test statistic has the convenient property that the first order asymptotics are unaffected by estimation error for the cases discussed below.

In this paper we present a non parametric method for testing for serial dependence and nonlinear structure in a time series. The method can be applied to forecast errors of fitted models. The null hypothesis that is tested for is that a time series sample comes from a data generating process that is IID.[3] The alternative hypothesis is not specified.While there are alternative hypotheses against which this test would have very little power, we present some Monte Carlo results in section 5 for some of the alternative hypotheses against which the test has good power.

In practice the test has been used to examine whether model errors are IID by fitting a model (e.g., models (1.1) and (1.2) below) and testing the estimated errors of the model for IID. Here model errors are to be interpreted as the ultimate stochastic driving process of the model. It is also possible to bootstrap the distribution of the test statistic under the null model (e.g., Theiler and Eubank (1993)).

In section 3 we show how to use the test on a model of the form

$$y_t = f(x_t, b, u_t), \tag{1.1}$$

where $\{x_t\}$ is a strictly stationary vector of observable variables which may include lagged y's; b is a vector of parameters that can be estimated $\sqrt{n}$–consistently (e.g., maximum likelihood estimators); $\{u_t\}$ is an IID process with zero mean and finite variance; and each u_t is independent of x_s for $s \leq t$. We will assume that f in equation (1.1) can be inverted so that

$$u_t = g(y_t, x_t, b). \tag{1.2}$$

This test can be applied to these residuals. If the model in equation (1.1) is correct, then the estimated residuals will (asymptotically) pass the test for IID. A failure to pass the test is an indication that the model is misspecified.

This test covers models of the form:

---

[2] Brock, Dechert, and Scheinkman (1987). The paper in the Review is a revision of Brock, Dechert, Scheinkman, and LeBaron (1991), which in turn is based on the original BDS paper.

[3] Throughout we will use the abbreviation IID for independent and identically distributed and MDS for martingale difference sequence.

$$y_t = \mu(x_t, a) + \sigma(x_t, b)u_t, \tag{1.3}$$

where a and b are vectors of parameters that can be estimated $\sqrt{n}$–consistently. This case includes the popular Generalized Auto Regressive Conditionally Heteroscedastic in Mean (GARCH-M) models of Bollerslev (1986) and Engle (1982) which have been applied to finance and economics by many writers, e.g. French, Schwert and Stambaugh (1987) and Hsieh (1987). These methods can be used to produce a diagnostic test for such models.

In many applications the first order asymptotic distribution of the statistic is independent of estimation error provided certain sufficient conditions hold. For example, Brock and Potter (1993) show that our diagnostic test is asymptotically justified on an appropriate transformation of GARCH type models provided that the conditional mean is known. This may be useful in financial applications such as high frequency data, where, to a first approximation, the conditional mean component is small relative to the noise component. See Bollerslev, Engle, and Nelson (1994, p. 2988) for this important point.

In the recent past, many researchers have used and analyzed the methods of this paper. Examples include, Abhyankar et al. (1995), Barnett, Gallant, and Hinich (1993), Barnett et al. (1994), Bollerslev, Engle, and Nelson (1994), Chavas and Holt, (1991), de Grauwe, Dewachter, and Embrechts (1993), Hiemstra and Jones (1994), Jaditz (1994), Krager, H., (1992), McClelland and Johnson (1993, 1994), Pesaran and Potter (1992), Potter (1995), Terasvirta, Tjostheim, and Granger (1994), Theiler and Eubank (1993), and Wolff (1995).[4]

There are a number of reasons why the BDS statistic has become so widely used. First, it can be applied as a goodness of fit test to any model which can be transformed into a model with additive IID errors and whose parameters can be estimated $\sqrt{n}$–consistently. Unlike many methods no correction term appears for the estimation of nuisance parameters. This invariance property is very useful in practice since no additional estimation is required for a given problem.[5] Hence no correction term to the first order asymptotic distribution is needed when estimated residuals are used. For this reason it is convenient to think of our test statistic as a nonlinear analog to the familiar Box–Pierce Q statistic in linear analysis, i.e., the diagnostic checking of fitted ARIMA models.[6]

Second, our test statistic is a function of U–statistics. This function measures the extra predictability in the data over and above that contained in the null. This property makes our

---

[4] Our methods have also attracted the attention of writers such as Casti (1992), Peters (1994), and Vaga (1994), who write for a broader audience than traditional academics.

[5] The first order asymptotic distribution of a residual diagnostic test statistic is derived by writing the statistic as a function of the parameter vector, expanding this function in a Taylor series around the true value of the parameter vector under the null, and discarding higher order terms. Typically, an extra term appears in this expansion due to the failure of the difference between the estimated residuals and the true residuals to go to zero (when multiplied by the square root of the sample length, n) as n tends to infinity. But, for our case, the extra term is zero. This convenient property comes from the symmetry of the kernel which appears in the U–statistics that make up our test statistic.

[6] Note that while Box and Jenkins (1976, p. 290–291) show that the standard errors for fitted autocorrelations on estimated residuals must be adjusted for estimation error even when the model is correctly specified, they also show that only a minor adjustment for the number of estimated parameters is required for the Q statistic.

statistic easy to interpret as a measure of the presence of "potential pockets of predictability" over the whole space.

Third, the software packages of Dechert and LeBaron are easy to use and are fast enough that the distribution of our statistic can be bootstrapped under the null when the sample size is small enough to render the asymptotics suspect.[7] Theiler and Eubank (1993) have recommended bootstrapping the distribution of our statistic directly under the null rather than on estimated residuals because it avoids the corrupting influence of structure clouding caused by pre-filtering. The speed of our software makes this feasible.

Fourth, it is pointed out by Bollerslev, Engle, and Nelson (1994), and Brock and Potter (1993) that provided the mean term is small, models in the ARCH type class may be tested for specification error with the use of our statistic as a residual diagnostic on a transformation of the model and the invariance result still holds. Since the ARCH class is very popular in financial applications[8] and since the mean term is usually small because of efficient markets, our statistic is useful as a diagnostic test of ARCH models.

Fifth, our statistic can be used to test for (a notion of) stochastic linearity. Define a stochastic process to be IID (MDS)–linear if it has a Wold type moving average representation with IID (MDS) innovations, not just uncorrelated innovations.[9] One can estimate the moving average representation and test IID–linearity by testing the estimated residuals for IID. Brock and Potter (1993) contains a discussion of this type of testing procedure for IID–linearity.

Sixth, the asymptotic distribution theory of the BDS statistic does not require higher moments to exist. This is important in financial economics because of the problems that thick tailed distributions can cause for many test statistics.

Many of the applications of our statistic are in the style of the Box–Pierce Q statistic. Like the Q statistic our test does not give the investigator much information on the cause of the rejection of the null. While there are versions of our statistic such as Wu, Savit, and Brock (1993) where one shuts off the power against dependence at different lags in order to get information on which lag the rejection is occurring, this is still limited information on the cause of rejection.[10] If the null model is rejected by a diagnostic test such as ours then the investigator can turn to other methods (possibly more costly) to find the cause of the rejection.

---

[7]LeBaron's software (C–source code) can compute our statistic for m between 2 and 10, and for samples of length 250 in approximately 5 seconds. For samples of length 2000 it takes approximately 25 to 35 seconds. This speed (achieved on the Macintosh II cx 68030 running at 16Mhz) is now fast enough that the small sample distribution of our test may be easily bootstrapped under the null hypothesis. Dechert's software (DOS executable file) calculates the BDS statistic for a sample length of 250 in approximately 0.3 seconds, length 1000 in 5 seconds and length 2000 in 20 seconds. (These speeds are for a 486 chip @ 33 MHz.) Hence one can deal with smaller sample sizes by bootstrapping.

[8]Bollerslev, Chou, Jayaraman and Kroner (1990) survey many of the applications of ARCH models in finance.

[9]The assumption of IID (MDS) innovations is key, else one would be trying to refute the existence of a Wold representation which always exists under suitable regularity conditions.

[10]The generalized BDS statistic in Dechert (1994) can also be used to identify the lags at which rejection occurs. Software to implement this test is bundled with the BDS software at `gopher.econ.wisc.edu`.

Evidence gleaned from testing GARCH type models using our test, led Brock, Lakonishok, LeBaron (1992) to develop a bootstrap based test which bootstraps (under the null hypothesis that the GARCH type model under scrutiny is the correct data generating process), test statistics gleaned off of what traders are actually observed to be doing. The idea is to treat traders as though they are constructing tests of optimal power against alternatives they believe to exist which are possibly too subtle for econometricians to see. The idea is that the traders are more motivated to detect alternatives to the null (a parameterized version of the Efficient Market Hypothesis) than a scientific observer.

This work relates naturally to the extensive work on specification testing in econometrics surveyed by Pagan and Hall (1983) and White (1987). For example, some of that work tests the properness of the specification (1.1), by testing the "orthogonality restrictions" imposed by the usual nonlinear "regression" requirement that the conditional expectation of the model innovations be zero. As de Jong (1992) shows, the Bierens (1990) test may be adapted to a time series setting to deliver a consistent conditional moment test of these orthogonality restrictions. In fact we think the Bierens and de Jong method may be a particularly promising method of testing MDS Linearity with conditional moment tests as discussed by Brock and Potter (1993).

Such a test does not require that the $\{u_t\}$ be IID, but it may be more difficult to compute than ours. Also more parametric structure is typically required to operationalize, for example, some of the likelihood based tests in the Pagan and Hall and the White surveys.[11]

We require little parametric structure but require that the $\{u_t\}$ be IID.[12] Furthermore under certain conditions all of our test statistics have the same asymptotic distribution for estimated standardized residuals as for true standardized residuals provided that the estimated model is correctly specified. Brock (1987) located sufficient conditions and indicated a proof for the first order asymptotic properties of the test developed in Brock, Dechert and Scheinkman (1987) to be the same whether or not the statistic is evaluated at the estimated residuals or the true ones. We go beyond the Brock paper here because we treat a much broader class of models. We locate sufficient conditions here for the first order asymptotic distribution of the test statistic to be the same for the estimated residuals $\{u_{t,n}\}$ as it is for the actual residuals $\{u_t\}$ where

$$u_{t,n} = g(y_t, x_t, b_n) \tag{1.4}$$

provided the model (1.1) correctly describes the data.

This test has advantages and disadvantages relative to the ones surveyed by Pagan/Hall and White. If one has strong beliefs about the class of alternative hypotheses to the model (1.1), then a test in the class surveyed by White can be constructed which has better power against those alternatives. But if one does not have strong beliefs about the class of

---

[11] For likelihood based specification tests also see Newey (1985) and Tauchen (1985).

[12] In fact the proofs are all done under the weaker assumption that the $\{u_t\}$ are weakly dependent. The only place that the IID assumption is used is in calculating the variance of the statistic. With the recent advances in bootstrapping weakly dependent time series data (c.f., LePage and Billard (1992)) the distribution of the BDS statistic can be numerically approximated without using the variance.

alternative hypotheses to the model, then this test which has good power against a broad class of alternatives should be useful. Thus we view these methods as complementary, rather than competitive, to the methods in Pagan/Hall and White.

In Section 5 we present Monte Carlo estimates of the empirical size of the non parametric test, and estimates of the power of the test of the IID null hypothesis against various alternatives. The other tests are a test for nonlinearity due to Tsay (1986), the bispectral test for nonlinearity due to Hinich (1982) and Subba Rao and Gabr (1980), and the test for ARCH due to Engle (1982). As an example of the performance of our test, the Engle test asymptotically has maximal power against ARCH alternatives yet our test comes within 80% of the power of the Engle test for a class of ARCH alternatives. Sometimes the power equals that of the Engle test even though the test in this paper is not optimal against ARCH. Since the sample sizes are fairly large we believe that we are close to the Neyman–Pearson maximal power against this class of ARCH alternatives.

The paper is organized as follows. In section 2 we develop the basic test of the IID null hypothesis, and in section 3 we show how to adapt and use the test to test the goodness of fit of a null model. In this way we can make a test for nonlinearity out of it as did Brock and Sayers (1988) and Scheinkman and LeBaron (1989a,b). These authors applied our test to the residuals of estimated linear models in order to test for extra structure. We show how recent authors such as Hsieh (1987), LeBaron (1987,1988), Schwert (1989), and Gallant, Hsieh and Tauchen (1990) have applied the test as a diagnostic for adequacy of fit for GARCH models to exchange rate and stock returns data. Some of the applications to economic models are discussed in section 4. In section 5 we briefly discuss the extensive Monte Carlo work on size and power that was done by Hsieh and LeBaron (1988a,b,c) which is contained in the book by Brock, Hsieh, LeBaron (1991). We then present the results of Monte Carlo studies on the power of the test against several alternative tests. The performance is quite good.

2   A Test for Independence

2.1   Definitions and Notation

Let $\{u_t\}$ be a strictly stationary stochastic process of real random variables with distribution function F. We shall call $(u_t, u_{t+1} \ldots, u_{t+m-1})$ an m–history, and denote it by u_t^m. Its distribution function is denoted by F_m. When the $\{u_t\}$ are independent, then $F_m(x_1, \ldots, x_m) = \prod_{k=1}^{m} F(x_k)$. Let $\mathcal{G}_i^j$ be σ–$\{u_i, u_{i+1}, \ldots, u_j\}$ for $1 \leq i < j \leq \infty$. The stochastic process $\{u_t\}$ is absolutely regular (see e.g., Denker and Keller (1983)) if

$$\beta_k = \sup_{n \geq 1} \left\{ \mathsf{E} \left[\sup \left\{ |\mathsf{P}(A \mid \mathcal{G}_1^n) - \mathsf{P}(A)| \ \middle| \ A \in \mathcal{G}_{n+k}^{\infty} \right\} \right] \right\} \tag{2.1}$$

converges to zero.

For $x \in \mathbf{R}^m$ we shall use the max norm, $\|x\| = \max_{1 \leq k \leq m} \{|x_k|\}$. When it is important to emphasize the dimension of the underlying space we shall use the notation $\| \cdot \|_m$ for this norm. The characteristic function of the set A is χ_A, and for the special case that

$A = [0, \epsilon)$ we denote its characteristic function by χ_ϵ. If $\Phi : \mathbf{R}^m \to \mathbf{R}^1$ is differentiable, then $(D\Phi)_x \in \mathbf{R}^m$ is the vector of partial derivatives of Φ evaluated at $x \in \mathbf{R}^m$. The directional derivative of Φ at x in the direction v is

$$(D\Phi)_x \cdot v = \lim_{\epsilon \to 0} \frac{\Phi(x + \epsilon v) - \Phi(x)}{\epsilon}.$$

2.2   The Correlation Integral

In Grassberger and Procaccia (1983) the correlation integral was introduced as a method of measuring the fractal dimension of deterministic data. It is a measure of the frequency with which temporal patterns are repeated in the data. The correlation integral at embedding dimension m is given by

$$C_{m,n}(\epsilon) = \frac{1}{\binom{n}{2}} \sum_{1 \le s < t \le n} \sum \chi_\epsilon(\|u_s^m - u_t^m\|), \tag{2.2}$$

and

$$C_m(\epsilon) = \lim_{n \to \infty} C_{m,n}(\epsilon). \tag{2.3}$$

If the data is generated by a strictly stationary stochastic process which is absolutely regular, then this limit exists. In this case the limit in equation (2.3) is

$$C_m(\epsilon) = \iint \chi_\epsilon(\|u - v\|) dF_m(u) dF_m(v). \tag{2.4}$$

(In equation (2.4) there are in fact $2m$ variables of integration. The convention that we use is that we will write one integral sign for each *vector* of variables in the integrand.) When the process is independent, and since $\chi_\epsilon(\|u - v\|) = \prod_{i=1}^{m} \chi_\epsilon(|u_i - v_i|)$, equation (2.4) implies that

$$C_m(\epsilon) = C_1(\epsilon)^m \tag{2.5}$$

2.3   Asymptotic Distribution of the Correlation Integral

$C_{m,n}$ in equation (2.2) is a generalized U-statistic (c.f., Serfling (1980, Chapter 5), and Denker and Keller (1983)) with symmetric kernel $\chi_\epsilon(\|x - y\|)$. Define

$$
\begin{aligned}
K(\epsilon) &= \int \left(\int \chi_\epsilon(|u - v|) dF(u) \right)^2 dF(v) \\
&= \int [F(u + \epsilon) - F(u - \epsilon)]^2 dF(u)
\end{aligned} \tag{2.6}
$$

and let $C(\epsilon) = C_1(\epsilon)$. When the process $\{u_t\}$ is independent, then the process $\{u_t^m\}$ is absolutely regular with $\beta_k = 0$ for $k > m$, and by Theorem 1 of Denker and Keller (1983) for generalized U–statistics we get

Theorem 2.1 *Let $\{u_t\}$ be IID. If $K(\epsilon) > C(\epsilon)^2$*

$$\sqrt{n} \frac{C_{m,n}(\epsilon) - C_1(\epsilon)^m}{\sigma_m(\epsilon)} \tag{2.7}$$

converges in distribution to N(0,1), where

$$\frac{1}{4}\sigma_m^2 = K^m - C^{2m} + 2\sum_{i=1}^{m-1}[K^{m-i}C^{2i} - C^{2m}]. \tag{2.8}$$

(The dependence of the terms in equation (2.8) on ϵ has been suppressed for notational clarity.)

Proof: All proofs are in section 7.

For $m = 1$, this theorem was proved independently by Denker and Keller (1986). In equation (2.8), if $K(\epsilon) = C(\epsilon)^2$ then $\sigma_m(\epsilon) = 0$, and we have a degenerate case. A discussion about the asymptotic properties of U–statistics for degenerate cases can be found in Serfling (1980, Chapter 5). Dechert (1989) showed that $K(\epsilon) = C(\epsilon)^2$ if and only if $F(u+\epsilon)-F(u-\epsilon)$ is constant for all u.[13] This is the case for a uniform distribution on a circle.[14] It is also the case for $\epsilon = 0$ as well as when the closure of the support of the random variables is compact and ϵ is large enough. For all other cases, $K(\epsilon) > C^2(\epsilon)$.

Notice that C_1, C and K appear in the statistic (2.7), and so when the distribution function is unknown, the statistic cannot be calculated. This leads to the study of the properties of

$$T_{m,n}(\epsilon) = C_{m,n}(\epsilon) - C_{1,n}(\epsilon)^m. \tag{2.9}$$

It turns out that $\lambda_1 C_{1,n}(\epsilon) + \lambda_2 C_{m,n}(\epsilon)$ is a generalized U–statistic with kernel $\lambda_1\chi_\epsilon(|u_1 - v_1|) + \lambda_2\chi_\epsilon(\|u - v\|)$, for $u, v \in \mathbf{R}^m$ and for all λ_1, λ_2. Again, by Denker and Keller (1983, Theorem 1) this linear combination is asymptotically normal, and so $C_{1,n}$ and $C_{m,n}$ are jointly asymptotically normal. Now, by the delta method (cf., Pollard (1984), Appendix A) the asymptotic variance of $\sqrt{n}T_{m,n}$ is the same as the asymptotic variance of the linear combination,

$$\sqrt{n}\left[C_{m,n}(\epsilon) - C_1(\epsilon)^m - mC_1(\epsilon)^{m-1}\left(C_{1,n}(\epsilon) - C_1(\epsilon)\right)\right].^{15} \tag{2.10}$$

These are the key steps of the proof of:

Theorem 2.2 *Let $\{u_t\}$ be IID. If $K(\epsilon) > C(\epsilon)^2$ then for $m \geq 2$*

$$\sqrt{n}\frac{T_{m,n}(\epsilon)}{V_m}$$

---

[13] Note that $K(\epsilon) - C(\epsilon)^2 = \mathrm{Var}\left(\mathsf{E}\left[\chi^\epsilon(u_t - u_s \mid u_t]\right)\right.$

[14] That is, u_t is the angle of the position of a point on the unit circle, with the standard convention that $u_t + 2\pi = u_t$. See Theiler (1988) for a numerical study of this case.

[15] The statistic is of the form $T(w, z) = w - z^m$. The first order Taylor series approximation around the means, $\bar{w}$ and $\bar{z}$, is:

$$T(w, z) - T(\bar{w}, \bar{z}) = w - \bar{w} - m\bar{z}^{m-1}(z - \bar{z}).$$

converges in distribution to $N(0,1)$. The asymptotic variance is given by

$$
\begin{aligned}
\frac{1}{4}V_m^2 \;=\;& m(m-2)C^{2m-2}(K-C^2)+K^m-C^{2m} \\
& +2\sum_{j=1}^{m-1}\left[C^{2j}\left(K^{m-j}-C^{2m-2j}\right)-mC^{2m-2}\left(K-C^2\right)\right].
\end{aligned}
\tag{2.11}
$$

(The dependence of the terms in equation (2.11) on ϵ has been suppressed for notational clarity.)

The constants C and K in equation (2.11) for the variance of $T_{m,n}$ can be consistently estimated by

$$
C_n(\epsilon) \;=\; \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\chi_\epsilon(|u_s-u_t|)
\tag{2.12}
$$

$$
K_n(\epsilon) \;=\; \frac{1}{n^3}\sum_{r=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\chi_\epsilon(|u_r-u_s|)\chi_\epsilon(|u_s-u_t|).
\tag{2.13}
$$

Let $V_{m,n}^2(\epsilon)$ denote the value of the variance in equation (2.11) evaluated with $C_n(\epsilon)$ and $K_n(\epsilon)$ in place of $C(\epsilon)$ and $K(\epsilon)$. $C_n(\epsilon)$ is a V–statistic[16] (Serfling (1980, Chapter 5) and K_n can be symmetrized so that it too is a V–statistic. C_n and K_n converge almost surely to C and K. Thus $V_{m,n}$ converges almost surely to V_m, and by Slutsky's Theorem and Theorem 2.2, the BDS statistic discussed in Brock, Dechert and Scheinkman (1987)

$$
W_{m,n}(\epsilon) = \sqrt{n}\frac{T_{m,n}(\epsilon)}{V_{m,n}(\epsilon)}
\tag{2.14}
$$

converges in distribution to $N(0,1)$.

Equation (2.14) is a distribution free statistic. Thus it has the advantage that no distributional assumptions need to be made in using it as a test statistic for IID random variables. On the other hand, it has the disadvantage that it will not be the most powerful statistic to use in testing a parametric hypothesis against a parametric alternative. Nevertheless, as our Monte Carlo studies in section 5 show, it does have good power against a variety of alternatives that are common in economics.

3 A Diagnostic Test

3.1 Data Generating Processes

Consider a data generating process of the following type:

$$
y_t = f(y_{t-1}, b, u_t)
\tag{3.1}
$$

---

[16]Estimating the variance in equation (2.11) involves terms in $K_n-C_n^2$, and powers of these terms. In order to insure that the estimated variance is positive, it is necessary that $K_n-C_n^2>0$. This is the case for the statistics in equations (2.12) and (2.13). If U-statistics are used for the estimators of C and K in equation (2.11), it is numerically possible for the estimated variance to be negative.

where b is an unknown vector of parameters, and $\{u_t\}$ is IID, $\mathsf{E}u_t^2 < \infty$, and for each t, u_t is independent of y_{t-1}. If equation (3.1) can be solved for u_t so that

$$u_t = g(Y_t, b) \tag{3.2}$$

where $Y_t = (y_t, y_{t-1})$, then when b_n is an estimate of b based on n observations of the process (3.1), the residuals are

$$u_{t,n} = g(Y_t, b_n). \tag{3.3}$$

If $b_n \xrightarrow{\mathcal{P}} b$, then $u_{t,n} \xrightarrow{\mathcal{P}} u_t$ and so for large n the sequence $\{u_{t,n}\}$ is "approximately" IID. A non parametric test for this can be based on the statistic (2.14), evaluated at the estimated residuals, $\{u_{t,n}\}$. In order for Theorem 2.2 to hold, we must locate sufficient conditions so that the variance of the estimated parameters does not affect the asymptotic variance of the statistic (2.14). Throughout the remainder of this paper, we shall treat b as a scalar parameter. There is no loss in generality, as all of the results go through for the multi-parameter case. A multivariate extension of the results reported here, including the invariance result, is contained in Baek and Brock (1992). Also, the model in equation (3.1) can be extended to include more lagged dependent variables, so that $Y_t = (y_t, y_{t-1}, \ldots)$ for as many lags as are present in the model.

The basic assumptions on the data generating process that we make are that the function g in equation (3.2) satisfies:

$$g_b(Y_t, b) \text{ is } \mathcal{G}_1^{t-1}\text{--measurable}, \tag{3.4}$$

$$\sup_t \sup_b \mathsf{E}[|g_b(Y_t, b)|] < \infty \tag{3.5}$$

and

$$\sup_t \sup_b \mathsf{E}[|g_{bb}(Y_t, b)|] < \infty. \tag{3.6}$$

The later two assumptions will hold if for example, $g \in C^2$ and there is a compact set K with $(Y_t, b) \in K$ for all t with probability one. These assumptions are used in the approximation

$$u_{t,n} - u_t \quad = \quad g(Y_t, b_n) - g(Y_t, b) \tag{3.7}$$

$$\approx \quad g_b(Y_t, b)(b_n - b) + \frac{1}{2}g_{bb}(Y_t, b)(b_n - b)^2 \tag{3.8}$$

in order to get $\sqrt{n}(b_n - b)^2 n^{-1} \sum_{t=1}^n g_{bb}(Y_t, b) \xrightarrow{\mathcal{P}} 0$. We shall also require an orthogonality condition in Theorem 3.2 which happens to be satisfied for the special case of a nonlinear autoregression,

$$y_t = f(y_{t-1}, b) + u_t \tag{3.9}$$

where the partial derivatives, f_b and f_{bb} satisfy the same conditions as g_b and g_{bb} in equations (3.5) and (3.6). In this case,

$$u_{t,n} - u_t \quad = \quad f(y_{t-1}, b_n) - f(y_{t-1}, b)$$

$$\approx \quad f_b(y_{t-1}, b)(b_n - b) + \frac{1}{2}f_{bb}(y_{t-1}, b)(b_n - b)^2. \tag{3.10}$$

We will make use of the fact that the coefficients of $b_n - b$ in equation (3.10) are $\mathcal{G}_1^{t-1}$--measurable.

3.2   The Basic Theorem

The main theorem of this section which extends the work of Brock and Dechert (1989) can be applied to the data generation models of section 3.1 to show that the presence of estimated parameters does not change the asymptotic variance of the statistic in equation (2.9). The following theorem is for a kernel which is a function of u_t^m. For any function, $\phi : \mathbf{R} \to \mathbf{R}$, and for $u, v \in \mathbf{R}^m$ define

$$\Phi(u - v) = \prod_{i=1}^{m} \phi(u_i - v_i).$$

Theorem 3.1 *Let $\{u_t\}$ be IID with a symmetric distribution function[17] F, and let $\{u_{t,n}\}$, $\{\xi_t\}$ and $\{\psi_n\}$ satisfy the following:*

  (A1) $u_{t,n} - u_t = \psi_n \xi_t + o_p(\psi_n)$,

  (A2) there exists Z with $\mathsf{E}|Z| < \infty$ such that $\sqrt{n}\psi_n \overset{\mathcal{D}}{\to} Z$,

  (A3) ξ_t is $\mathcal{G}_1^{t-1}$ measurable,

  (A4) $\displaystyle\sup_n \frac{1}{n}\sum_{t=1}^{n} \mathsf{E}|\xi_t|^2 < \infty$,

Let ϕ be an even C^2 function which also satisfies

  (A5) $0 \le \phi \le 1$,

  (A6) $\displaystyle\sup_u |\phi'(u)| < \infty$,

  (A7) $\displaystyle\sup_u |\phi''(u)| < \infty$.

Let

  (A8) $U_n = \dfrac{1}{\binom{n}{2}} \displaystyle\sum_{1 \le s < t \le n} \Phi(u_t^m - u_s^m)$

  (A9) $\widehat{U}_n = \dfrac{1}{\binom{n}{2}} \displaystyle\sum_{1 \le s < t \le n} \Phi(u_{t,n}^m - u_{s,n}^m)$,

and define $\Delta_n = \widehat{U}_n - U_n$. Then

$$\sqrt{n}\Delta_n \overset{\mathcal{P}}{\to} 0.$$

Notice that in this theorem, assumption (A1) has $u_{t,n} - u_t - \psi_n \xi_t$ as $o_p(\psi_n)$ *uniformly* in t. Yet in the applications we have in mind, as for example in equation (3.9) this is not the case. It is here that the uniform boundedness of g_{bb} enters. By Taylor's Theorem, we can find a b_n' such that $|b_n' - b| \le |b_n - b|$ and

$$u_{t,n} - u_t = g_b(Y_t, b)(b_n - b) + \frac{1}{2} g_{bb}(Y_t, b_n')(b_n - b)^2.$$

---

[17]F is symmetric if $\forall x, F(-x) + F(x) = 1$.

The dependence of the second order term on t is through g_{bb} which is uniformly bounded by assumption, and therefore it is $O_p(\psi_n^2)$, uniformly in t. Thus, we will be able to use this theorem as a basis for our results on the types of models which we have described above.

In the proof of Theorem 3.1, assumption (A3) is used to show that

$$\mathsf{E}\left[\xi_t \int \phi'(u_t - v)dF(v)\right] = \mathsf{E}[\xi_t]\mathsf{E}\left[\int \phi'(u_t - v)dF(v)\right] \tag{3.11}$$

$$= \mathsf{E}[\xi_t] \iint \phi'(u - v)dF(u)dF(v) \tag{3.12}$$

$$= 0, \tag{3.13}$$

where equation (3.11) follows from the assumption that ξ_t is $\mathcal{G}_1^{t-1}$–measurable and is therefore independent of u_t, and equation (3.13) is proved in Lemma 7.1. More generally, assumption (A3) can be replaced by

$$\mathsf{E}\left[\xi_t \int \phi'(u_t - v)dF(v)\right] = 0,$$

which is a one dimensional ($m = 1$) version of the orthogonality condition which we referred to in the introduction.

Theorem 3.1 is for smooth kernels, ϕ, while the statistic $C_{m,n}$ in equation (2.2) uses a non-smooth kernel, χ_ϵ. The following approximation result shows that Theorem 3.1 holds for smooth kernels arbitrarily close to χ_ϵ. Let F be continuous, $0 < \delta < \epsilon$ and ϕ_δ be an even C^2 function which also satisfies:

$$\phi_\delta(x) = \left\{ \begin{array}{ll} 0 & |x| \geq \epsilon \\ 1 & |x| \leq \epsilon - \delta \end{array} \right. \tag{3.14}$$

as well as

$$0 \leq \phi_\delta(x) \leq 1.$$

Let $A_\delta = \{(u, v) \mid \epsilon - \delta < |u - v| < \epsilon\}$. Then for any $0 < \delta < \epsilon$ and $p \geq 1$

$$\iint |\phi_\delta(u - v) - \chi_\epsilon(|u - v|)|^p dF(u)dF(v) \leq \iint_{A_\delta} dF(u)dF(v)$$

$$= \mathsf{P}(\epsilon - \delta < |U - V| < \epsilon)$$

$$\xrightarrow{\delta \to 0} 0.$$

Therefore, $\phi_\delta \xrightarrow{\mathcal{L}_p} \chi_\epsilon$, and the kernel for the BDS statistic can be approximated arbitrarily well in probability by ϕ_δ. As for the estimate of the variance using ϕ_δ, define

$$C_\delta = \iint \phi_\delta(u - v)dF(u)dF(v)$$

and

$$K_\delta = \int \left(\int \phi_\delta(u - v)dF(v) \right)^2 dF(u).$$

Then

$$
\begin{aligned}
|C_\delta - C| &= \left| \iint [\phi_\delta(u-v) - \chi_\epsilon(|u-v|)] dF(u) dF(v) \right| \\
&\leq \iint \left| \phi_\delta(u-v) - \chi_\epsilon(|u-v|) \right| dF(u) dF(v) \qquad (3.15) \\
&\leq \iint_{A_\delta} dF(u) dF(v)
\end{aligned}
$$

and therefore $\lim_{\delta \to 0} C_\delta = C$. Similarly, $\lim_{\delta \to 0} K_\delta = K$. Since equation (2.11) for the variance involves terms in C and K, it can also be approximated arbitrarily closely by using the kernel ϕ_δ.

The symmetry assumption on F is used in the proof of Theorem 3.1 to show that

$$
\iint \phi'(u-v) dF(u) dF(v) = 0.
$$

For a general kernel, ϕ, the symmetry assumption is sufficient in order to draw this conclusion. However, for the approximating functions ϕ_δ to the kernel χ_ϵ we need only assume that F'' exists and is integrable. For this case,

$$
\iint \phi'(u-v) f(u) f(v) du dv = \iint \phi(u-v) f(u) f'(v) du dv
$$

and

$$
\begin{aligned}
\lim_{\delta \to 0} \iint \phi_\delta(u-v) f(u) f'(v) du dv &= \iint \chi_\epsilon(|u-v|) f(u) f'(v) du dv \\
&= \iint_{u-\epsilon}^{u+\epsilon} f(u) f'(v) du dv \\
&= \int [f(u+\epsilon) - f(u-\epsilon)] f(u) du \\
&= 0.
\end{aligned}
$$

Therefore, for the statistic (2.9) it is sufficient to assume that the distribution function, F, of the u_t be twice differentiable and that F'' is integrable for those cases when it is not symmetric.

A similar approximation result as in the remarks after Theorem 3.1 can be applied to the statistic (2.14). Let $\widehat{T}_{m,n}$ denote the statistic (2.9) evaluated at the residuals of (3.3), and let $\widehat{V}_{m,n}^2$ denote the variance where the terms (2.12) and (2.13) are also evaluated at these residuals. We would like to have a result that shows that

$$
\sqrt{n} \frac{\widehat{T}_{m,n}}{\widehat{V}_{m,n}}
$$

is asymptotically $N(0,1)$. For a smooth approximation to the kernel χ_ϵ we can get such a result. Let

$$
\Phi_\delta(u-v) = \prod_{i=1}^m \phi_\delta(u_i - v_i).
$$

be based on the kernel in equation (3.14), and let $\widehat{T}_{\delta,m,n}$ and $\widehat{V}_{\delta,m,n}^2$ be the statistic and variance based on the smooth kernel, Φ_δ. Finally, let $V_{\delta,m}^2$ be the variance of the statistic in equation (2.10), also calculated using the smooth kernel.

Theorem 3.2 *Let $\{u_t\}$ be IID. Assume that the data generating process (3.1) has a representation (3.2) which satisfies the conditions in equations (3.5) and (3.6) and let $\xi_t = g_b(Y_t, b)$. If the orthogonality condition*

$$\mathsf{E}\left[\int (D\Phi_\delta)_{(u_t^m - u)} \cdot \xi_t^m \, dF_m(u)\right] = 0 \tag{3.16}$$

holds, then

$$\sqrt{n}\frac{\widehat{T}_{\delta,m,n}}{\widehat{V}_{\delta,m,n}} \tag{3.17}$$

converges in distribution to $N(0,1)$. (Recall that $(D\Phi_\delta)_{(u_t^m - u)} \cdot \xi_t^m$ is the directional derivative of Φ_δ at $u_{t,n} - u$ in the direction ξ_t^m.)

It is worth pointing out that the statistic $\sqrt{n}\widehat{T}_{\delta,m,n}$ is asymptotically normal with variance $V_{\delta,m}$. By equation (3.15) this variance is continuous in δ and converges to V_m as δ goes to 0.

Remark: We have *not* been able to show that $\sqrt{n}\widehat{T}_{m,n}/\widehat{V}_{m,n}$ is asymptotically normal. It appears difficult to give a formal proof.[18] The results of Hsieh and LeBaron (1988a,b,c) show that the distribution of the statistic $\sqrt{n}\widehat{T}_{m,n}/\widehat{V}_{m,n}$ appears to be roughly normal with unit variance on true residuals as well as on estimated residuals for a class of parametric models driven by IID errors with sample sizes around 500. Our own Monte Carlo work in Section 5 shows the approximation to be fairly good for three classes of models for sample sizes of 250 and 500.

For certain models the technique of Randles (1982, Theorem 2.8) will work. Suppose that the model can be put into the form of equation (1.2) and that b_0 is the true value of the parameter vector. Denote the parameterized process by $\{u_t(b)\}$. If there is a neighborhood, $N(b_0)$, and a constant $A > 0$ so that for each $b \in N(b_0)$, and for all $d > 0$

$$\mathsf{E}\left[\sup_{|b'-b|<d} |\chi_\epsilon(u_1(b) - u_2(b)) - \chi_\epsilon(u_1(b') - u_2(b'))|\right] \le Ad \tag{3.18}$$

then our Theorem (3.2) holds.[19] The condition in equation (3.18) is not an easy one to check in general. Here is a simple example of the technique.

Example: Let the model be $y_{t+1} = f(y_t) + \sigma_0 u_{t+1}$, where $\{u_t\}$ is IID with $\mathsf{E}u_t = 0$, $\mathsf{E}u_t^2 = 1$, and has a bounded density function. A $\sqrt{n}$–consistent estimator for σ_0 is

$$\sigma_n = \sqrt{\frac{1}{n}\sum_{t=0}^{n-1}[y_{t+1} - f(y_t)]^2}$$

---

[18]Recently, two different proofs of this result have been given. One by de Lima (1992a) and the other by Lai, Wang and Wiorkowski (1994).

[19]It is worth pointing out that there are two conditions that must be met to apply Randles's Theorem. However, the condition in his Lemma (2.6) trivially holds in our case, and so by the conclusion of that Lemma, equation (3.18) above is all that needs to hold.

In our notation the residuals are

$$u_{t,n} = \frac{y_{t+1} - f(y_t)}{\sigma_n}$$

which forms the sequence of observations that is used in calculating $\widehat{T}_{m,n}(\epsilon)$. Following Randles's notation,

$$
\begin{aligned}
\mu(\sigma) &= \mathsf{E}_{\sigma_0} \chi_\epsilon (u_1(\sigma) - u_2(\sigma)) \\[2mm]
&= \mathsf{E}_{\sigma_0} \chi_\epsilon \left(\frac{\sigma_0}{\sigma} [u_1(\sigma_0) - u_2(\sigma_0)] \right) \\[2mm]
&= \mathsf{E}_{\sigma_0} \chi_{\frac{\epsilon\sigma}{\sigma_0}} (u_1(\sigma_0) - u_2(\sigma_0)) \\[2mm]
&= C \left(\frac{\epsilon\sigma}{\sigma_0} \right) \\[2mm]
&= \int \left[F\left(x + \frac{\epsilon\sigma}{\sigma_0}\right) - F\left(x - \frac{\epsilon\sigma}{\sigma_0}\right) \right] dF(x).
\end{aligned}
$$

Since we are assuming that the $\{u_t(\sigma_0)\}$ have a density function, f,

$$\mu'(\sigma) = \frac{2\epsilon}{\sigma_0} \int f\left(x + \frac{\epsilon\sigma}{\sigma_0}\right) f(x)dx$$

i.e., $\mu(\sigma)$ is a differentiable function.

Now to check the condition in equation (3.18). Notice that

$$|\chi_\epsilon(u_1(\sigma) - u_2(\sigma)) - \chi_\epsilon(u_1(\sigma') - u_2(\sigma'))| =$$

$$\left| \chi_{\frac{\epsilon\sigma}{\sigma_0}}(u_1(\sigma_0) - u_2(\sigma_0)) - \chi_{\frac{\epsilon\sigma'}{\sigma_0}}(u_1(\sigma_0) - u_2(\sigma_0)) \right|$$

and the supremum over $|\sigma - \sigma'| < d$ occurs at $\sigma' = \sigma \pm d$. (It is for small d that equation (3.18) must be checked.) The expected value of the supremum is

$$\mathsf{P}\left(\frac{\epsilon(\sigma - d)}{\sigma_0} < |u_1(\sigma_0) - u_2(\sigma_0)| < \frac{\epsilon\sigma}{\sigma_0} \right) +$$

$$\mathsf{P}\left(\frac{\epsilon\sigma}{\sigma_0} < |u_1(\sigma_0) - u_2(\sigma_0)| < \frac{\epsilon(\sigma + d)}{\sigma_0} \right).$$

The density function for $U_1 - U_2$ is the convolution

$$f_{U_1 - U_2}(u) = \int_{-\infty}^{\infty} f(u + x)f(x)dx$$

and the sum of the probabilities above is

$$\int_{\frac{\epsilon(\sigma - d)}{\sigma_0}}^{\frac{\epsilon(\sigma + d)}{\sigma_0}} \int_{-\infty}^{\infty} f(u + x)f(x)dxdu + \int_{-\frac{\epsilon(\sigma + d)}{\sigma_0}}^{\frac{-\epsilon(\sigma - d)}{\sigma_0}} \int_{-\infty}^{\infty} f(u + x)f(x)dxdu$$

which is bounded by

$$\frac{4M\epsilon}{\sigma_0}d$$

where M is an upper bound for the density function of u_t. This is the condition for Randles's Theorem (2.8) to hold. By our Theorem (2.1) we know that $T_{m,n}$ is asymptotically normal and by Randles' Theorem we get that

$$\sqrt{n}\left[\widehat{T}_{m,n} - T_{m,n}\right] \xrightarrow{\mathsf{P}} 0.$$

Therefore the limiting distribution of $\sqrt{n}\widehat{T}_{m,n}$ is the same as the limiting distribution of $T_{m,n}$.

By a similar computation as in equations $(3.11) - (3.13)$, the condition in equation (3.16) is satisfied whenever ξ_{t+1} is $\mathcal{G}_1^t$–measurable. This is the case, for example, when the data generating process is of the form (3.9), where the variance of $\{u_t\}$ is known. A separate computation shows that the condition (3.16) also holds when the variance of $\{u_t\}$ must be estimated. This is relevant in applications where one typically normalizes the data by the estimated standard deviation.

The conclusion of this theorem is that for data generating processes that satisfy the hypotheses of the theorem there is no correction term needed in the asymptotic variance of the BDS statistic (2.14) when it is evaluated at the estimated residuals, $\{u_{t,n}\}$. The hypothesis (3.16) is satisfied for any model of the form (1.3) where $\sigma(x_t, b)$ is constant. This includes all autoregressions with IID errors. It does not include heteroscedastic errors, even when the heteroscedasticity can be estimated as in ARCH models, except in the important case where the mean is zero as discussed by Brock and Potter (1993) and Bollerslev, Engle and Nelson (1994).

There have been two main extensions to the results reported here. First, de Lima (1992a) has proved the invariance theorem for the indicator kernel for a much larger class of statistics which includes our statistic by extending Randles's work as indicated above. De Lima has rigorously proved that the invariance result holds for indicator kernels under regularity conditions comparable to those assumed above for smooth kernels. So, under our conditions, the invariance theorem is valid for the indicator kernel for models (such as the one with additive IID errors) where the parameters can be estimated $\sqrt{n}$–consistently.

Second, de Lima (1992b) has compared the moment requirements of many diagnostic and specification type tests to the moment requirements of ours. Ours essentially requires no moments at all because of the boundedness of the indicator kernel. Of course this boundedness may choke off the ability of the test to detect certain types of dependence because part of the space is "cut off." However, the moment requirements of many other tests are at the level of finite fourth moments or even finite higher moments.

Work of Loretan and Phillips (1994) has shown that the evidence is against the existence of moments of order four (or higher) in financial applications. It is not obvious one can get around this problem. De Lima (1992b) has shown that moment condition failure is important in financial applications because it effects the size of the test. This work has important implications in finance because of the problems that heavy tails (moment condition failure) cause for many popular tests.

4 Applications

The test for independence in the previous section can be used on several problems in applied econometrics. It can be used as a test for nonlinearity as well as a test of model misspecification. As a specification test it can be used in the following way. First, fit a parametric model of the form

$$y_t = \mu(x_t, a) + \sigma(x_t, b)u_t \tag{4.1}$$

(where a and b are parameters and $\{u_t\}$ is an IID stochastic process) with $\sqrt{n}$–consistent estimates of the parameters. If the orthogonality condition holds, then Theorem 3.2 locates sufficient conditions for the asymptotic distribution of the BDS statistic (2.14) to be the same whether it is evaluated at the estimated residuals

$$u_{t,n} = \frac{y_t - f(x_t, a_n)}{h(x_t, b_n)} \tag{4.2}$$

or at the true residuals $\{u_t\}$. In general, a correction term based on (3.16) must be added to the asymptotic variance of the statistic (2.14). Since most of the estimation procedures such as maximum likelihood are $\sqrt{n}$–consistent for most problems, one can use our methods as a test for the adequacy of models which can be transformed into the form (4.1). Many models can be transformed into this form, including most of the applications to finance of ARCH and GARCH models that are currently used.

The general procedure can be easily used to test for nonlinearity, i.e., non IID–linearity in the sense of Brock and Potter (1993) discussed above, of a univariate series such as detrended U. S. industrial production.[20] This was done by Brock and Sayers (1988) by taking first differences of the logarithms of monthly industrial production and fitting autoregressions to these first differences. The residuals of the fitted autoregressions were then tested for IID. The IID hypothesis was rejected when tested on residuals of the best fitting autoregressions. This is consistent with non IID–linearity. It is also consistent with ARCH effects and nonstationarity.

This procedure was also applied by Brock and Sayers (1988) to U.S. real GNP post World War II (WWII) but the estimated residuals failed to reject IID. However, Stock (1988) and Potter (1990) find evidence of nonlinearity in U.S. real GNP using other methods. Scheinkman and LeBaron (1988a) studied U.S. real GNP over a much longer time span and failed to reject IID for the estimated residuals after standardizing them by estimated pre WWII variance and estimated post WWII variance. The models in both of these applications are special cases of (4.1).

The BDS statistic has been used to test the residuals of GARCH type models for independence. If the null model is indeed GARCH then the standardized residuals of the fitted GARCH model should be independent.

---

[20]In the case where one tests the null hypothesis of an auto regression of *known* finite order driven by IID errors, the correction term (3.16) is zero. We have not extended Theorem 3.2 to the more realistic case where the order is unknown and, hence must be *estimated*.

In an interesting study, Hsieh (1987) applied this procedure to exchange rate data and rejected the GARCH model for three out of five currencies. Since Theorem 3.2 indicates the presence of a correction term to the GARCH residuals as in equation (4.2), Hsieh (1987) bootstrapped the distribution of (3.17) under the null hypothesis. Hsieh (1991) applied our test investigate the goodness-of-fit of a broad class of parametric models of conditional volatility by using it to test the IID null hypothesis of standardized residuals.

Scheinkman and LeBaron (1989a) applied the test in this paper to the CRSP value weighted portfolio daily and weekly returns series, and found that even after extracting the residuals of a linear auto–regression the IID hypothesis was rejected. LeBaron (1988) has applied our test to the residuals of GARCH models fitted to weekly returns on the value weighted CRSP stock returns index. He finds that the GARCH residuals reject IID for the period July 1962-March 1974 but not for the period April 1974-December 1985. As did Hsieh, LeBaron bootstrapped the distribution of (3.17) under the null hypothesis.

Finally our test has been used by Gallant, Hsieh, and Tauchen (1988) to test standardized residuals for predictability of higher order conditional moments than means and variances. It has been used by Schwert (1989) to test the adequacy of a broad class of models to predict stock market volatility as a function of past information such as macroeconomic variables and past returns.

5 Size and Power Experiments

This section explores the size and power of the BDS statistic. The power is tested against some nonlinear alternatives, and the BDS test is compared with some other tests for non-linear structure. The power tests are run using simulated critical values for all tests making this a fair small sample comparison.

Table I presents size simulations for an IID normal null. In this table $W_{m,n}(\epsilon)$ is the normalized BDS statistics from equation (2.14). The parameter ϵ is set to 1/2 the sample standard deviation. Table I represents the results of 5000 replications for sample sizes of 250 and 500. For results on other distributions, epsilons, and sample sizes see Brock, Hsieh, LeBaron(1991).

Rather than using asymptotic distributions to get the critical values for power comparisons, critical values will be taken from small sample simulations for all tests used. Three comparison tests will be performed. The first is the test of Tsay (1986). This tests looks for significance in the regression of cross terms ($x_{t-i}x_{t-j}$ on x_t). This test is performed in four steps:

1. Regress x_t on the lagged vector $v_t = (1, x_{t-1}, \ldots, x_{t-M})$ and save the residuals as u_t. $\{x_t\}$ is a time series of length n.

2. Build the vector z_t as the value of the lower half (on or below the diagonal) of the matrix $(x_{t-1}, \ldots, x_{t-M})^T (x_{t-1}, \ldots, x_{t-M})$ stacked in a vector. For $M = 2, z_t = (x_{t-1}^2, x_{t-1}x_{t-2}, x_{t-2}^2)$. Regress z_t on w_t saving the residual vector y_t.

3. Regress u_t on y_t saving the resulting residuals as e_t.

4. Calculate:
$$F = \frac{(\Sigma y_t u_t)(\Sigma y_t' y_t)^{-1}(\Sigma y_t' u_t)(n - M - m - 1)}{(\Sigma \epsilon_t^2)(m)}$$

where,
$$m = M(M+1)/2.$$

Tsay shows that if x_t follows an autoregressive process of order M then F is asymptotically distributed $F(m, T - M(M+3)/2 - 1)$. For this paper $M = 2$ and $M = 3$ are used.

TABLE I

Size of $W_{m,n}(\epsilon)$ statistic

5000 Replications of samples of i.i.d. Normal

$\epsilon = 1/2$(Standard Deviation)

n = 250	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$	N(0,1)
% < −2.33	3.46	5.12	8.52	1.00
% < −1.96	7.26	9.14	13.04	2.50
% < −1.64	11.56	13.74	17.98	5.00
% > 1.64	11.16	12.78	16.44	5.00
% > 1.96	7.62	9.74	12.36	2.50
% > 2.33	4.96	6.54	9.20	1.00
n = 500	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$	N(0,1)
% < −2.33	1.92	2.48	4.20	1.00
% < −1.96	4.28	5.34	8.02	2.50
% < −1.64	8.10	9.60	12.52	5.00
% > 1.64	8.84	9.82	11.74	5.00
% > 1.96	5.52	6.92	8.76	2.50
% > 2.33	2.90	4.02	5.94	1.00

The second test used is a test proposed by Engle (1982) for ARCH, computed as follows. Take the residuals of a fitted linear model u_t and regress u_t^2 on p lags plus a constant. Engle shows that the statistic nR^2, formed by multiplying the sample size by the R^2 of the regression of the squares is asymptotically $\chi^2(p)$. For this paper we will use $p = 1$ and $p = 2$.

The power of these tests will be compared on 5 different nonlinear models. These models were chosen both for their common use, and the important fact that they all have no linear structure. This eliminates the problem of removing linear structure by taking residuals of a fitted linear model.

The first two are the ARCH and GARCH models of Engle (1982), and Bollerslev (1986) respectively. The GARCH model can be represented in the following form:

$$x_t \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i x_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

The ARCH model is the above system for $p = 0$. We will use an ARCH model where $q = 1$, $\alpha_0 = 1$, and $\alpha_1 = 0.5$. The GARCH model used has $q = p = 1$, $\alpha_0 = 1$, $\alpha_1 = .1$, and $\beta_1 = 0.8$. The third model used is the nonlinear moving average (NLMA). It has the following form.

$$x_t = 0.5\epsilon_{t-1}\epsilon_{t-2} + \epsilon_t$$

The ϵ_t terms are IID normal. A closely related model that will be tested is the extended nonlinear moving average (ENLMA). It has the following form.

$$x_t = 0.8\epsilon_{t-1}\sum_{j=2}^{20}(0.8)^{j-2}\epsilon_{t-j} + \epsilon_t$$

The last model tested will be the threshold autoregressive (TAR).

$$x_t = \begin{cases} -0.5x_{t-1} + \epsilon_t & \text{if} \quad x_{t-1} \leq 1 \\ 0.4x_{t-1} + \epsilon_t & \text{otherwise} \end{cases}$$

Table II displays the power for the different tests against each of these models. The numbers in table II represent frequency of rejection at the 5% confidence level. Critical values for all tests are determined from 5000 simulations of each sample size for an IID normal. T_2 and T_3 are the Tsay statistics for $m = 2$ and $m = 3$. E_1 and E_2 are the Engle statistics for $p = 1$ and $p = 2$. The other three are the BDS tests. Although the BDS statistic is out performed in many instances, it appears to have a broad range of power against several alternatives. It should be noted that the Monte Carlo work of Brock, Hsieh, LeBaron (1991) have shown that the BDS statistic performs better with large data sets (> 500).

In practical applications one rarely finds series that exhibit zero autocorrelation as the above models do. In most cases the BDS test will be used as a residual diagnostic. Tables III, IV and V check size for residuals of fitted models. Table III reports the results for residuals from fitted AR(1), with an autoregressive coefficient of 0.5. Comparing table III to table I shows very little change in size when fitted residuals are used. In table IV the experiment is repeated for an MA(1) with moving average term equal to 0.5. Estimation is done using maximum likelihood with the starting value set to the true value. The size results in this table also do not differ much from table I. The last group of residuals is from the NLMA model simulated above, and the results are given in Table V.

TABLE II

Power Comparisons

1000 Replications

$\epsilon = 1/2$(Standard Deviation)

n = 250							
Alternative	T_2	T_3	E_1	E_2	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$
ARCH	0.40	0.44	0.98	0.98	0.95	0.89	0.75
GARCH	0.15	0.18	0.42	0.49	0.22	0.22	0.21
NLMA	0.99	0.99	0.59	0.72	0.28	0.35	0.31
ENLMA	0.53	0.64	0.87	0.89	0.83	0.89	0.84
TAR	1.00	0.99	0.70	0.63	0.68	0.53	0.37
n = 250							
Alternative	T_2	T_3	E_1	E_2	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$
ARCH	0.45	0.52	1.00	1.00	1.00	1.00	0.99
GARCH	0.15	0.20	0.64	0.73	0.42	0.52	0.53
NLMA	1.00	1.00	0.80	0.93	0.56	0.76	0.71
ENLMA	0.70	0.84	0.98	0.99	1.00	1.00	1.00
TAR	1.00	1.00	0.94	0.90	0.95	0.89	0.76

TABLE III

Residual Size of $W_{m,n}(\epsilon)$ statistic

1000 Replications of AR(1) Residuals

$\epsilon = 1/2$(Standard Deviation)

n = 250	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$	N(0,1)
$\% < -2.33$	4.20	5.80	8.80	1.00
$\% < -1.96$	7.70	10.70	14.00	2.50
$\% < -1.64$	12.70	15.80	19.20	5.00
$\% > \ \ 1.64$	9.10	12.00	15.10	5.00
$\% > \ \ 1.96$	5.80	8.30	12.30	2.50
$\% > \ \ 2.33$	3.00	5.60	9.40	1.00
n = 500	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$	N(0,1)
$\% < -2.33$	1.30	2.10	3.80	1.00
$\% < -1.96$	5.00	4.80	7.50	2.50
$\% < -1.64$	8.40	8.80	12.10	5.00
$\% > \ \ 1.64$	6.90	8.30	10.80	5.00
$\% > \ \ 1.96$	4.40	4.60	7.50	2.50
$\% > \ \ 2.33$	2.10	2.40	4.80	1.00

TABLE IV

Residual Size of $W_{m,n}(\epsilon)$ statistic

1000 Replications of MA(1) Residuals

$\epsilon = 1/2$(Standard Deviation)

n = 250	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$	N(0,1)
% < −2.33	3.60	5.40	8.50	1.00
% < −1.96	6.30	10.20	14.80	2.50
% < −1.64	11.10	15.90	20.20	5.00
% > 1.64	9.10	11.30	15.20	5.00
% > 1.96	5.90	8.70	12.80	2.50
% > 2.33	4.30	6.40	9.60	1.00
n = 500	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$	N(0,1)
% < −2.33	2.20	2.30	3.90	1.00
% < −1.96	5.40	4.80	7.40	2.50
% < −1.64	8.70	9.90	12.90	5.00
% > 1.64	6.50	7.80	11.20	5.00
% > 1.96	4.40	4.60	7.40	2.50
% > 2.33	2.40	2.80	4.80	1.00

TABLE V

Residual Size of $W_{m,n}(\epsilon)$ statistic

1000 Replications of NLMA Residuals

$\epsilon = 1/2$(Standard Deviation)

n = 250	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$	N(0,1)
% < −2.33	4.40	7.40	10.70	1.00
% < −1.96	7.90	12.40	15.30	2.50
% < −1.64	13.10	18.50	21.30	5.00
% > 1.64	9.00	9.40	12.00	5.00
% > 1.96	5.60	6.60	10.00	2.50
% > 2.33	3.90	4.50	8.30	1.00
n = 500	$W_{2,n}(\epsilon)$	$W_{3,n}(\epsilon)$	$W_{4,n}(\epsilon)$	N(0,1)
% < −2.33	2.10	2.60	4.10	1.00
% < −1.96	4.50	5.10	8.70	2.50
% < −1.64	8.50	10.40	14.10	5.00
% > 1.64	7.30	7.10	10.80	5.00
% > 1.96	4.50	4.90	8.00	2.50
% > 2.33	2.60	2.70	5.00	1.00

Here, we are not looking at residuals from a linear model, but residuals from the fitted true nonlinear model. Even though the fitted model is much more complicated the residuals still appear close to IID for the test. This suggests that the this test is not just a test of nonlinearity, but might be a good test of model specification.

6 Summary and Conclusions

This paper has presented a new method of testing for serial dependence in a strictly stationary time series. The method has its roots in chaos theory. This method has been extensively evaluated by Monte Carlo studies, some of which appear in this paper. For sample sizes of 500 or more the test has quite good size performance and good power against a range of alternatives. For models driven by additive IID errors with finite variances where the parameters of the model can be estimated at least $\sqrt{n}$–consistently, the first order asymptotics, under the null hypothesis, are the same on estimated errors as on the true errors. This property makes the test useful as a general diagnostic test for neglected nonlinearity (or neglected nonstationarity) and model specification.

In discussing the advantages and disadvantages of the testing procedure that has been presented in this paper we point out four issues.

First, our test is not consistent against *all* departures from IID. This lack of consistency means that our statistic cannot be used as a consistent model specification test as the de Jong (1992) extension of Bierens (1990). However, Dechert (1989, 1994) has shown that a variant of our test is consistent within the class of Gaussian processes. Hence, this result suggests that our test may be approximately consistent, although Dechert (1989) has provided alternatives against which our test is not consistent.[21]

Second, there is the issue of how to choose the parameters ϵ and m. The approach of Hsieh and LeBaron (1988a,b,c) was to choose ϵ to optimize the size and power performance on their experimental design. They recommend choosing ϵ between 0.5 and 1.5 standard deviations of the data. One approach to taking ϵ to zero with increasing sample size is contained in Dechert (1989, 1994). Another is to choose ϵ to maximize power against some particular alternative.[22] Another approach to choosing ϵ is in Wolff (1995). The choice of m depends upon which lag the investigator wishes to test for dependence. De Lima (1992a) develops a class of statistical tests of IID which "integrate" over m and ϵ but still preserve the invariance property.

Third, there are many tests of independence including tests based on the difference between the joint distribution and the product of the marginals, factorization of the characteristic function, properties of Kullback–Liebler entropy, and the like. Many of these are consistent. This may lead some to argue that one of these tests should be used instead of ours.

Fourth, for uses of the statistic as a diagnostic test for models which cannot be transformed into models with additive IID errors, a correction term must be added to the asymptotic distribution. Baek and Brock (1992) show how to calculate the correction term for

---

[21] There are departures against which tests like the Box–Pierce Q Statistic are not consistent either. Yet, such tests are popular in practice because they are easy to use and do give some useful information. The ease of use, coupled with easily available software, and the invariance property make our test useful.

[22] As mentioned in the introduction, when testing a specific hypothesis and alternative, deriving an optimal test is clearly preferred. However, if the null hypothesis is IID in the class of all non–parametric distributions (which is composite) vs. a specific alternative, then our test with choosing ϵ to maximize power might be a sound one.

multivariate cases which include the single variable case treated here. Furthermore, a more conventional proof of the invariance result is given there. The approach developed here states the regularity conditions at a level of abstraction that includes, for example, $\sqrt{n}$–consistency of the estimation of model parameters as a special case.

In view of the nice theoretical properties of the diagnostic test, together with the technical results presented in section 5 on the Monte Carlo results and software to implement the test, we believe that the statistic developed here is a useful diagnostic supplement for time series analysis.

7 Proofs

Proof of Theorem 2.1: This follows directly from Denker and Keller (1983, Theorem 1c) by treating $\{u_t^m\}$ as a stochastic process. Since u_s^m and u_t^m are independent for $|s - t| \geq m$, the coefficients β_k in (2.1) are zero for this process when $k \geq m$. Thus, for $\{u_t\}$ IID, $\{u_t^m\}$ is an absolutely regular stationary stochastic process. Also, the condition $\sum_k \beta_k^{\frac{\delta}{2+\delta}} < \infty$ is trivially satisfied. Finally, since $|\chi_\epsilon| \leq 1$, it follows that $\mathsf{E}|\chi_\epsilon(u_t^m - u_s^m)|^{2+\delta} \leq 1$. Thus the conditions for Theorem 1c of Denker and Keller are met. The formula for the variance comes from the following: for $u \in \mathbf{R}^m$, let

$$
\begin{align}
h_1(u) &= \int \chi_\epsilon(\|u - v\|) dF_m(v) - C^m \tag{7.1} \\
&= \prod_{i=1}^m [F(u_i + \epsilon) - F(u_i - \epsilon)] - C^m. \tag{7.2}
\end{align}
$$

Since $\int [F(u + \epsilon) - F(u - \epsilon)] \, dF(u) = C$, therefore $\mathsf{E}[h_1(u_t^m)] = 0$. The formula for the asymptotic variance from Denker and Keller (1983) is given by:

$$
\frac{1}{4}\sigma_m^2 = \mathsf{E}\left[h_1(u_1^m)^2\right] + 2\sum_{t>1} \mathsf{E}\left[h_1(u_1^m)h_1(u_t^m)\right]. \tag{7.3}
$$

Now,

$$
\begin{align}
\mathsf{E}\left[h_1(u_1^m)^2\right] &= \int \prod_{i=1}^m [F(u_i + \epsilon) - F(u_i - \epsilon)]^2 \, dF_m(u) - C^{2m} \\
&= \prod_{i=1}^m \int [F(u_i + \epsilon) - F(u_i - \epsilon)]^2 \, dF(u_i) - C^{2m} \\
&= K^m - C^{2m}
\end{align}
$$

and for $t \leq m$,

$$
\mathsf{E}\left[h_1(u_1^m)h_1(u_t^m)\right] = \int \left\{ \prod_{i=1}^{t-1} [F(u_i + \epsilon) - F(u_i - \epsilon)] \times \right.
$$
$$
\prod_{i=t}^{m-1} [F(u_i + \epsilon) - F(u_i - \epsilon)]^2 \times
$$

$$\left. \prod_{i=m}^{t+m-1} [F(u_i + \epsilon) - F(u_i - \epsilon)] \right\} dF_{t+m-1}(u)$$
$$- C^{2m}$$
$$= \quad C^{t-1} K^{m-t+1} C^{t-1} - C^{2m}$$
$$= \quad K^{m-(t-1)} C^{2(t-1)} - C^{2m}$$

For $t > m$, u_1^m and u_t^m are independent and so $\mathsf{E}[h_1(u_1^m)h_1(u_t^m)] = 0$. Therefore,

$$\tfrac{1}{4}\sigma_m^2 \quad = \quad K^m - C^{2m} + 2\sum_{t=2}^{m}\left[K^{m-(t-1)}C^{2(t-1)} - C^{2m}\right]$$

$$= \quad K^m - C^{2m} + 2\sum_{t=1}^{m-1} C^{2t}\left[K^{m-t} - C^{2(m-t)}\right]$$

<div align="right">QED</div>

Proof of Theorem 2.2: This proof follows the same lines as Theorem 2.1. The formula for the variance can be derived as follows: for any λ, $C_{m,n}(\epsilon) + \lambda C_{1,n}(\epsilon)$ is a U–statistic with kernel $\chi_\epsilon(|u_t^m - u_s^m|) + \lambda \chi_\epsilon(|u_1 - u_s|)$. For $u \in \mathbf{R}^m$ let

$$h_1(u) \quad = \quad \int [\chi_\epsilon(|u - v|) + \lambda\chi_\epsilon(|u_1 - v_1|)]dF_m(u)$$
$$-C(\epsilon)^m - \lambda C(\epsilon)$$
$$= \quad \prod_{i=1}^{m}[F(u_i + \epsilon) - F(u_i - \epsilon)] + \lambda[F(u_1 + \epsilon) - F(u_1 - \epsilon)]$$
$$-C(\epsilon)^m - \lambda C(\epsilon)$$

For ease of notation, let $A_i = F(u_i + \epsilon) - F(u_i - \epsilon)$. Then

$$\mathsf{E}\left[h_1(u_1^m)^2\right] \quad = \quad \int\left[\prod_{i=1}^{m} A_i + \lambda F_1\right]^2 dF_m(u) - [C^m + \lambda C]^2$$
$$= \quad \prod_{i=1}^{m}\int A_i^2 dF(u_i) + 2\lambda\int A_1\prod_{i=1}^{m}[A_i dF(u_i)]$$
$$+\lambda^2\int A_1^2 dF(u_1) - [C^m + \lambda C]^2$$
$$= \quad K^m - C^{2m} + 2\lambda C^{m-1}[K - C^2] + \lambda^2[K - C^2].$$

Making the substitution, $\lambda = -mC^{m-1}$ we get,

$$\mathsf{E}\left[h_1(u_1^m)^2\right] = K^m - C^{2m} \ - 2mC^{2m-2}\left(K - C^2\right) + m^2 C^{2m-2}\left(K - C^2\right).$$

Similarly, for $1 \leq t \leq m$,

$$\mathsf{E}\left[h_1(u_1^m)h_1(u_t^m)\right] \quad = \quad \mathsf{E}\left[\prod_{i=1}^{m} A_i A_{i+t-1} + \lambda\left(A_1\prod_{i=1}^{m} A_{i+t-1} + A_t\prod_{i=1}^{m} A_i\right)\right]$$

$$+\lambda^2 A_1 A_t] - (C^m + \lambda C)^2$$
$$= C^{t-1} K^{m-t+1} C^{t-1} + \lambda \left(C^{m+1} + K C^{m-1} \right)$$
$$+ \lambda^2 C^2 - C^{2m} - 2\lambda C^{m+1} - \lambda^2 C^2.$$

Making the substitution, $\lambda = -mC^{m-1}$ we get,

$$\mathsf{E}\left[h_1(u_1^m) h_1(u_t^m)\right] = K^{m-(t-1)} C^{2(t-1)} - mC^{2m}$$
$$-mKC^{2m-2} - C^{2m} + 2mC^{2m}$$
$$= C^{2(t-1)} \left[K^{m-(t-1)} - C^{2m-2(t-1)} \right]$$
$$-mC^{2m-2} \left(K - C^2 \right).$$

By substituting these terms into equation (7.3) we get the formula for the variance in equation (2.11)   QED

The proof of Theorem 3.1 uses the following lemmas. Define the function ϕ_1 by

$$\phi_1(u) = \int \phi(u-v) dF(v),$$

and $\phi_{\delta,1}$ by

$$\phi_{\delta,1}(u) = \int \phi_\delta(u-v) dF(v).$$

Lemma 7.1 *If F is symmetric then*

$$\int \phi_1'(u) dF(u) = \iint \phi'(u-v) dF(v) dF(u) = 0$$

Proof:

$$\iint \phi'(u-v) dF(v) dF(u) =$$
$$\int_{-\infty}^{\infty} \int_{-\infty}^{u} \phi'(u-v) dF(v) dF(u) \qquad (7.4)$$
$$+ \int_{-\infty}^{\infty} \int_{u}^{\infty} \phi'(u-v) dF(v) dF(u)$$

Symmetry of F implies that

$$\int_{a}^{b} g(u) dF(u) = \int_{-b}^{-a} g(-u) dF(u)$$

and applying this to the second integral in equation (7.4):

$$\int_{-\infty}^{\infty} \int_{u}^{\infty} \phi'(u-v) dF(v) dF(u)$$
$$= \int_{-\infty}^{\infty} \int_{-u}^{\infty} \phi'(-u-v) dF(v) dF(u)$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{u} \phi'(-u+v) dF(v) dF(u)$$
$$= - \int_{-\infty}^{\infty} \int_{-\infty}^{u} \phi'(u-v) dF(v) dF(u)$$

where the last equality follows from the fact that ϕ' is an odd function, $\phi'(-x) = -\phi'(x)$. Thus we have that

$$\iint \phi'(u-v)dF(u)dF(v)$$
$$= \int_{-\infty}^{\infty}\int_{u}^{\infty} \phi'(u-v)dF(v)dF(u)$$
$$- \int_{-\infty}^{\infty}\int_{u}^{\infty} \phi'(u-v)dF(v)dF(u)$$
$$= 0.$$

<div align="right">QED</div>

Lemma 7.2 *Under the hypotheses of Theorem 3.1*

$$n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n}[\phi'(u_t - u_s) - \phi_1'(u_t)]\xi_t \tag{7.5}$$

converges to zero in probability.

Proof: Consider the L_1 norm of equation (7.5):

$$\mathsf{E}\left| n^{-1}\sum_{t=1}^{n}\left[n^{-1}\sum_{s=1}^{n} \phi'(u_t - u_s) - \phi_1'(u_t)\right]\xi_t \right|$$
$$\leq n^{-1}\sum_{t=1}^{n}\mathsf{E}\left| n^{-1}\sum_{s=1}^{n}\phi'(u_t - u_s) - \phi_1'(u_t)\right| |\xi_t|$$
$$\leq \left(n^{-1}\sum_{t=1}^{n}\mathsf{E}\left[n^{-1}\sum_{s=1}^{n}\phi'(u_t - u_s) - \phi_1'(u_t)\right]^2 \right)^{\frac{1}{2}}$$
$$\cdot \left(n^{-1}\sum_{t=1}^{n}\mathsf{E}[\xi_t]^2 \right)^{\frac{1}{2}}$$

where the latter inequality follows from Hölder's inequality applied to the product measure of $n^{-1}\sum_{t=1}^{n}\mathsf{E}[\cdot]$. By assumption, $\sup_n n^{-1}\sum_{t=1}^{n}\mathsf{E}[\xi_t^2] < \infty$, so we need only show that the first term converges to zero:

$$n^{-1}\sum_{t=1}^{n}\mathsf{E}\left[n^{-1}\sum_{s=1}^{n}\phi'(u_t - u_s) - \phi_1'(u_t)\right]\left[n^{-1}\sum_{r=1}^{n}\phi'(u_t - u_r) - \phi_1'(u_t)\right]$$
$$= n^{-1}\sum_{t=1}^{n}\mathsf{E}\left[n^{-2}\sum_{r=1}^{n}\sum_{s=1}^{n}\phi'(u_t - u_s)\phi'(u_t - u_r)\right.$$
$$\left. -2\phi_1'(u_t)n^{-1}\sum_{s=1}^{n}\phi'(u_t - u_s) + \phi_1'(u_t)^2\right]. \tag{7.6}$$

The expected value of these terms can be evaluated with the help of Lemma 7.1. The terms

in the summation in equation (7.6) become:

$$
\mathsf{E}[\phi'(u_t - u_r)\phi'(u_t - u_s)] \;=\; \begin{cases} \phi'(0)^2 & r = s = t \\ \phi'(0)\mathsf{E}[\phi'(u_t - u_s)] = 0 & s \neq r = t \\ \phi'(0)\mathsf{E}[\phi'(u_t - u_r)] = 0 & r \neq s = t \\ \iint \phi'(u-v)^2 dF(u)dF(v) & r = s \neq t \\ \int \phi'_1(u)^2 dF(u) & r \neq s \neq t \end{cases} \tag{7.7}
$$

and

$$
\mathsf{E}[\phi'_1(u_t)\phi'(u_t - u_s)] \;=\; \begin{cases} \phi'(0)\mathsf{E}[\phi'_1(u_t)] = 0 & s = t \\ \int \phi'_1(u)^2 dF(u) & s \neq t \end{cases} \tag{7.8}
$$

and

$$
\mathsf{E}[\phi'_1(u_t)^2] = \int \phi'_1(u)^2 dF(u). \tag{7.9}
$$

For simplicity, let $\alpha = \phi'(0)^2$, $\beta = \iint \phi'(u-v)^2 dF(v)dF(u)$ and $\gamma = \int \phi'_1(u)^2 dF(u)$. Then,

$$
\frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\left[\frac{1}{n}\sum_{s=1}^{n} \phi'(u_t - u_s) - \phi'_1(u_t) \right]^2
$$

$$
= \frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{n^2}\{\alpha + (n-1)\beta + (n-1)(n-2)\gamma\} - \frac{2}{n}(n-1)\gamma + \gamma \right]
$$

$$
= \frac{1}{n^2}\alpha + \frac{n-1}{n^2}\beta + \left(\frac{n^2 - 3n + 2}{n^2} - 2\frac{n-1}{n} + 1 \right)\gamma
$$

$$
= \frac{1}{n^2}\alpha + \frac{n-1}{n^2}\beta + \frac{2-n}{n^2}\gamma \;\longrightarrow\; 0 \;\;\text{as}\;\; n \to \infty. \tag{7.10}
$$

Therefore equation (7.5) converges to zero in $\mathcal{L}_1$ norm, and hence in probability.

$$\text{QED}$$

Lemma 7.3 *Under the hypotheses of Theorem 3.1*

$$
n^{-1}\sum_{t=1}^{n}\phi'_1(u_t)\xi_t \tag{7.11}
$$

converges to zero in probability.

Proof: Consider the $\mathcal{L}_2$ norm of equation (7.11):

$$
\mathsf{E}\left[\left(\frac{1}{n}\sum_{t=1}^{n}\phi'_1(u_t)\xi_t \right)^2 \right] = \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\mathsf{E}[\phi'_1(u_t)\phi'_1(u_s)\xi_t\xi_s]. \tag{7.12}
$$

For $s = t$,

$$
\begin{aligned} \mathsf{E}[\phi'_1(u_t)^2\xi_t^2] &= \mathsf{E}\left[\xi_t^2 \mathsf{E}^{\mathcal{G}_1^{t-1}}[\phi'_1(u_t)^2] \right] \\ &= \mathsf{E}[\xi_t^2] \int \phi'_1(u)^2 dF(u) \end{aligned}
$$

where the first equality follows from the fact that ξ_t is $\mathcal{G}_1^{t-1}$ measurable and the second equality follows from the fact that u_t is independent of $\mathcal{G}_1^{t-1}$. For $s < t$, u_s, ξ_s and ξ_t are $\mathcal{G}_1^{t-1}$–measurable, and so

$$
\begin{aligned} \mathsf{E}[\phi'_1(u_t)\phi'_1(u_s)\xi_t\xi_s] &= \mathsf{E}\left[\xi_s\phi'_1(u_s)\xi_t \mathsf{E}^{\mathcal{G}_1^{t-1}}[\phi'_1(u_t)] \right] \\ &= 0 \end{aligned}
$$

and similarly for $s > t$, $\mathsf{E}\left[\phi'(u_t)\phi'(u_s)\xi_t\xi_s\right] = \mathsf{E}\left[\xi_t\phi_1'(u_t)\xi_s\,\mathsf{E}^{\mathcal{G}_1^{s-1}}[\phi_1'(u_s)]\right] = 0$. Thus,

$$\mathsf{E}\left[\left(\frac{1}{n}\sum_{t=1}^n \phi_1'(u_t)\xi_t\right)^2\right] = \frac{1}{n^2}\sum_{t=1}^n \mathsf{E}[\xi_t^2]\int \phi_1'(u)^2 dF(u)$$

$$= \left[\frac{1}{n}\sum_{t=1}^n \mathsf{E}[\xi_t^2]\right]\left[\frac{1}{n}\int \phi_1'(u)^2 dF(u)\right]$$

which converges to zero. Since convergence in $\mathcal{L}_2$ norm implies convergence in probability,

$$\frac{1}{n}\sum_{t=1}^n \phi_1'(u_t)\xi_t \xrightarrow{\mathcal{P}} 0.$$

Proof of Theorem 3.1: We prove this first for $m = 1$. From the definition of Δ_n,

$$\Delta_n = \frac{1}{\binom{n}{2}}\sum\sum_{1\le s<t\le n}[\phi(u_{t,n}-u_{s,n})-\phi(u_t-u_s)]$$

Let $\tilde{\Delta}_n = (1-n^{-1})\Delta_n$ so that

$$\tilde{\Delta}_n = \frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n[\phi(u_{t,n}-u_{s,n})-\phi(u_t-u_s)]$$

$$= \frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n[\phi(u_{t,n}-u_{s,n})-\phi(u_t-u_s)]$$

$$\stackrel{L}{=} \frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n \phi'(u_t-u_s)\left[(u_{t,n}-u_t)-(u_{s,n}-u_s)\right] \tag{7.13}$$

$$= \frac{2}{n^2}\sum_{s=1}^n\sum_{t=1}^n \phi'(u_t-u_s)(u_{t,n}-u_t) \tag{7.14}$$

$$= \psi_n\frac{2}{n^2}\sum_{s=1}^n\sum_{t=1}^n \phi'(u_t-u_s)\xi_t$$

$$+o_p(\psi_n)n^{-2}\sum_{s=1}^n\sum_{t=1}^n \phi'(u_t-u_s) \tag{7.15}$$

where $\stackrel{L}{=}$ means that only the first order terms have been retained, and the third equality follows from the even property of ϕ. Let $\|\phi'\|_\infty = \sup_u |\phi'(u)|$ which is finite by hypothesis, and consider the coefficient of $o_p(\psi_n)$ in the second term of equation (7.15):

$$\left|\frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n \phi'(u_t-u_s)\right| \le \frac{1}{n^2}\sum_{s=1}^n\sum_{t=1}^n |\phi'(u_t-u_s)|$$

$$\le \|\phi'\|_\infty.$$

Therefore the second term in equation (7.15) is $o_p(\psi_n)$. Continuing from equation (7.15),

$$\tilde{\Delta}_n = \psi_n\frac{2}{n^2}\sum_{s=1}^n\sum_{t=1}^n \phi'(u_t-u_s)\xi_t + o_p(\psi_n)$$

$$= \psi_n \left\{ \frac{2}{n} \sum_{t=1}^n \frac{1}{n} \sum_{s=1}^n [\phi'(u_t - u_s) - \phi_1'(u_t)]\xi_t \right.$$

$$\left. + \frac{2}{n} \sum_{t=1}^n \phi_1'(u_t)\xi_t \right\} + o_p(\psi_n)$$

Thus (up to the linear terms) $\sqrt{n}\tilde{\Delta}_n$ (and hence $\sqrt{n}\Delta_n$) is of the form $X_n W_n + o_p(\sqrt{n}\psi_n)$ where $X_n \overset{\mathcal{D}}{\to} X$ and $o_p(\sqrt{n}\psi_n) \overset{P}{\to} 0$. Now by Slutsky's Theorem, if $W_n \overset{P}{\to} 0$ then $X_n W_n \overset{\mathcal{D}}{\to} 0$ and since 0 is a constant, $X_n W_n \overset{P}{\to} 0$.[23] Therefore, to show that $\sqrt{n}\Delta_n \overset{P}{\to} 0$, it is sufficient to show that dropping the higher order terms in equation (7.13) is justified and that

$$n^{-1} \sum_{t=1}^n n^{-1} \sum_{s=1}^n [\phi'(u_t - u_s) - \phi_1'(u_t)]\xi_t + n^{-1} \sum_{t=1}^n \phi_1'(u_t)\xi_t \qquad (7.16)$$

converges to zero in probability.

Consider one of the quadratic terms in the exact form of the remainder that have been dropped in equation (7.13):

$$\frac{1}{2n^2} \left| \sum_{s=1}^n \sum_{t=1}^n \phi''(\tilde{u}_{t,n} - \tilde{u}_{s,n})(u_{t,n} - u_t)^2 \right|$$

$$\leq \frac{1}{2n^2} \|\phi''\|_\infty \sum_{s=1}^n \sum_{t=1}^n (u_{t,n} - u_t)^2$$

$$= \|\phi''\|_\infty \left[\frac{1}{2n} \psi_n^2 \sum_{t=1}^n \xi_t^2 + o_p(\psi_n^2) \right]$$

where $|\tilde{u}_{t,n} - u_t| < |u_{t,n} - u_t|$. By assumption, $n^{-1} \sum_{t=1}^n \xi_t^2$ is stochastically bounded, $\sqrt{n}\psi_n \overset{\mathcal{D}}{\to} Z$, $\psi_n \overset{P}{\to} 0$ and thus $o_p(\sqrt{n}\psi_n^2) \overset{P}{\to} 0$. Therefore

$$\sqrt{n} \frac{1}{2n^2} \left| \sum_{s=1}^n \sum_{t=1}^n \phi''(\tilde{u}_{t,n} - \tilde{u}_{s,n})(u_{t,n} - u_t) \right| \overset{P}{\to} 0,$$

and so the higher order terms in equation (7.13) can be neglected. The above argument also applies to the cross product terms, $(u_{t,n} - u_t)(u_{s,n} - u_s)$, since

$$\frac{1}{2n^2} \left| \sum_{s=1}^n \sum_{t=1}^n \phi''(\tilde{u}_{t,n} - \tilde{u}_{s,n})(u_{s,n} - u_s)(u_{t,n} - u_t) \right|$$

$$\leq \|\phi''\|_\infty \left[\frac{1}{2n^2} \psi_n^2 \sum_{s=1}^n \sum_{t=1}^n |\xi_s \xi_t| + o_p(\psi_n^2) \right]$$

$$\leq \|\phi''\|_\infty \left[\frac{1}{2n} \psi_n^2 \sum_{t=1}^n \xi_t^2 + o_p(\psi_n^2) \right].$$

---

[23]If $Z_n \overset{\mathcal{D}}{\to} 0$ then for all $\epsilon > 0$, $\lim_{n\to\infty} F_n(\epsilon) = 1$ and $\lim_{n\to\infty} F_n(-\epsilon) = 0$, where F_n is the distribution function of Z_n. Since $\mathcal{P}(|Z_n| < \epsilon) = F_n(\epsilon) - F_n(-\epsilon)$, it follows that $\lim_{n\to\infty} \mathcal{P}(|Z_n| < \epsilon) = 1$ for all ϵ.

By Lemmas 7.1 — 7.3 each of the two terms in equation (7.16) converges to zero in probability, and therefore $\sqrt{n}\Delta_n \overset{P}{\to} 0$. The case that $m > 1$ is dealt with as part of the proof of Theorem 3.2 below.

$$\text{QED}$$

Lemma 7.4 *Let $\widehat{C}_n$ and $\widehat{K}_n$ be given by equations (2.12) and (2.13) evaluated at the calculated residuals, $\{u_{tn}\}$; let V_m^2 be given by (2.11) and let $\widehat{V}_{m,n}^2$ be the same formula evaluated at $\widehat{C}_n$ and $\widehat{K}_n$. Then $\widehat{V}_{m,n}^2$ converges to V_m^2 in probability.*

Proof: By Theorem 3.1 and the approximation result in section 3.2 we have already shown that $\widehat{C}_n$ converges to C in probability. The same steps can be followed for $\widehat{K}_n$. The crucial step is equation (7.13) in the proof of Theorem 3.1, which in the estimation of K is as follows: let

$$\Delta_n = \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \left[\phi(u_{r,n} - u_{s,n})\phi(u_{s,n} - u_{t,n}) - \phi(u_r - u_s)\phi(u_s - u_t) \right].$$

Then

$$\begin{aligned}
\Delta_n \overset{L}{=} \quad & \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'(u_r - u_s)\phi(u_s - u_t)(u_{r,n} - u_r) \\
& + \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi(u_r - u_s)\phi'(u_s - u_t)(u_{s,n} - u_s) \\
& - \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'(u_r - u_s)\phi(u_s - u_t)(u_{s,n} - u_s) \\
& - \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi(u_r - u_s)\phi'(u_s - u_t)(u_{t,n} - u_t).
\end{aligned}$$

By relabeling the indices in the second, third and fourth terms, we get:

$$\begin{aligned}
\Delta_n \overset{L}{=} \quad & \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'(u_r - u_s)\phi(u_s - u_t)(u_{r,n} - u_r) \\
& + \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi(u_t - u_r)\phi'(u_r - u_s)(u_{r,n} - u_r) \\
& - \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'(u_s - u_r)\phi(u_r - u_t)(u_{r,n} - u_r) \\
& - \frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi(u_t - u_s)\phi'(u_s - u_r)(u_{r,n} - u_r) \\
= \quad & 2\frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'(u_r - u_s)[\phi(u_s - u_t) + \phi(u_r - u_t)](u_{r,n} - u_r) \\
= \quad & 2\frac{1}{n^3} \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'(u_r - u_s)[\phi(u_s - u_t) + \phi(u_r - u_t)]\xi_r \\
& \qquad\qquad\qquad\qquad\qquad + o_p(\psi_n) \qquad\qquad (7.17)
\end{aligned}$$

where the second equality follows from the evenness of ϕ. Since

$$|\phi'(u_r - u_s)[\phi(u_s - u_t) + \phi(u_r - u_t)]| \leq 2\|\phi'\|_\infty$$

the first term in the right hand side of equation (7.17) is bounded by

$$4\|\phi'\|_\infty |\psi_n| \frac{1}{n} \sum_{r=1}^n |\xi_r|,$$

which converges to zero in probability.

<div align="right">QED</div>

For $u \in \mathbf{R}^m$ define

$$\Phi_1(u) = \int \Phi(u - v) dF^m(u)$$

and

$$\Phi_{\delta,1}(u) = \int \Phi_\delta(u - v) dF^m(u)$$

Proof of Theorem 3.2: The steps of the proof for the case that $m > 1$ are the same as in the proof of Theorem 3.1 for $m = 1$. In the proof that follows, we will point out the differences in the vector case from the scalar case.

The summation that corresponds to equation (7.5) is

$$n^{-2} \sum_{t=1}^n \sum_{s=1}^n [(D\Phi_\delta)_{(u_t^m - u_s^m)} - (D\Phi_{\delta,1})_{u_t^m}] \cdot \xi_t^m, \tag{7.18}$$

(recall that $(D\Phi_\delta)_u \cdot v$ is the directional derivative of Φ_δ in the direction v) and the summation that corresponds to equation (7.6) is,

$$\frac{1}{n} \sum_{t=1}^n \mathsf{E} \left[\frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n (D\Phi_\delta)_{(u_t^m - u_s^m)} \cdot (D\Phi_\delta)_{(u_t^m - u_r^m)} \right.$$
$$\left. - \frac{2}{n} \sum_{s=1}^n (D\Phi_{\delta,1})_{u_t^m} \cdot (D\Phi_\delta)_{(u_t^m - u_s^m)} + \left| (D\Phi_{\delta,1})_{u_t^m} \right|^2 \right]. \tag{7.19}$$

Here, $|\cdot|$ is the Euclidean norm. When r, s and t are such that there are no overlapping components of u_t^m, u_r^m and u_s^m (i.e., when $|r - s| \geq m$, $|r - t| \geq m$ and $|s - t| \geq m$) then

$$\mathsf{E} \left[(D\Phi_\delta)_{(u_t^m - u_s^m)} \cdot (D\Phi_\delta)_{(u_t^m - u_r^m)} \right] = \mathsf{E} \left| (D\Phi_{\delta,1})_{u_t^m} \right|^2.$$

For each t, there are $(n - m)(n - 2m)$ such terms in the double summation on r and s. The remaining $m(3n - 2m)$ terms are all uniformly bounded. When we divide by n^2 these will go to zero, and so for the ease of notation let α' be the value of the bound for these terms. Similarly, for $|t - s| \geq m$

$$\mathsf{E} \left[(D\Phi_{\delta,1})_{u_t^m} \cdot (D\Phi_\delta)_{(u_t^m - u_s^m)} \right] = \mathsf{E} \left| (D\Phi_{\delta,1})_{u_t^m} \right|^2.$$

For each t, there $n - m$ such terms in the second summation on s in equation (7.19). Let α be the maximum of α' and the value of the bound of the remaining m terms. Let $\gamma = \mathsf{E} \left| (D\Phi_{\delta,1})_{u_t^m} \right|^2$, which by stationarity is constant for all t. Then an upper bound for the value of equation (7.19) is

$$\frac{1}{n^2} \left[m(3n - 2m)\alpha + (n - m)(n - 2m)\gamma \right] - \frac{2}{n} \left[m\alpha + (n - m)\gamma \right] + \gamma$$

$$= \frac{m(n - 2m)}{n^2}\alpha + \left(1 - 3\frac{m}{n} + 2\frac{m^2}{n^2} - 2 + 2\frac{m}{n} + 1 \right)\gamma$$

$$= \frac{m(n - 2m)}{n^2}\alpha + \left(2\frac{m^2}{n^2} - \frac{m}{n} \right)\gamma$$

which converges to zero.

Corresponding to equation (7.12) we have:

$$\frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \mathsf{E} \left[\left((D\Phi_{\delta,1})_{u_t^m} \cdot \xi_t^m \right) \left((D\Phi_{\delta,1})_{u_s^m} \cdot \xi_s^m \right) \right]. \tag{7.20}$$

Now for $|s - t| \geq m$,

$$\mathsf{E} \left[\left((D\Phi_{\delta,1})_{u_t^m} \cdot \xi_t^m \right) \left((D\Phi_{\delta,1})_{u_s^m} \cdot \xi_s^m \right) \right] = \mathsf{E} \left[(D\Phi_{\delta,1})_{u_t^m} \cdot \xi_t^m \right] \mathsf{E} \left[(D\Phi_{\delta,1})_{u_s^m} \cdot \xi_s^m \right]$$

by independence. Let us show that this is zero for $m = 3$. The general case follows a similar argument. Now

$$
\begin{aligned}
(D\Phi_{\delta,1})_{u_t^3} \cdot \xi_t^3 \quad = \quad & \phi'_{\delta,1}(u_t)\phi_{\delta,1}(u_{t+1})\phi_{\delta,1}(u_{t+2})\xi_t \\
& + \phi_{\delta,1}(u_t)\phi'_{\delta,1}(u_{t+1})\phi_{\delta,1}(u_{t+2})\xi_{t+1} \\
& + \phi_{\delta,1}(u_t)\phi_{\delta,1}(u_{t+1})\phi'_{\delta,1}(u_{t+2})\xi_{t+2}
\end{aligned}
$$

and so

$$
\begin{aligned}
\mathsf{E}[(D\Phi_{\delta,1})_{u_t^3} \cdot \xi_t^3] \quad = \quad & \mathsf{E}[\phi'_{\delta,1}(u_t)\xi_t]\mathsf{E}[\phi_{\delta,1}(u_{t+1})]\mathsf{E}[\phi_{\delta,1}(u_{t+2})] \\
& + \mathsf{E}[\phi_{\delta,1}(u_t)\xi_{t+1}]\mathsf{E}[\phi'_{\delta,1}(u_{t+1})]\mathsf{E}[\phi_{\delta,1}(u_{t+2})] \\
& + \mathsf{E}[\phi_{\delta,1}(u_t)\phi_{\delta,1}(u_{t+1})\xi_{t+2}]\mathsf{E}[\phi'_{\delta,1}(u_{t+2})] \\
= \quad & 0, \tag{7.21}
\end{aligned}
$$

where the first equality follows by independence, and the second equality by Lemma 7.1.

In going from equation (7.14) to (7.15) and the analogous equations of r $m > 1$, it is necessary that the difference $u_{t,n} - u_t - \psi_n \xi_t$ be $o_p(\psi_n)$ *uniformly* in t. Assumption (3.6) is sufficient. Using the exact Taylor series expansion,

$$u_{t,n} - u_t = g_b(Y_t, b)(b_n - b) + \frac{1}{2}g_{bb}(Y_t, b'_n)(b_n - b)^2$$

where $|b'_n - b| \leq |b_n - b|$, the missing quadratic terms in equation (7.16) are

$$\frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'_\delta(u_s - u_t) g_{bb}(Y_t, b'_n)(b_n - b)^2.$$

Now, by hypothesis, $\sup_u |\phi'_\delta(u)| < \infty$ and $\sup_{t,b} \mathsf{E}|g_{bb}(Y_t, b)| < \infty$. Therefore,

$$\frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'_\delta(u_s - u_t) g_{bb}(Y_t, b'_n)$$

is stochastically bounded. By assumption $\sqrt{n}(b_n - b)$, converges in distribution to a random variable which is integrable, and so $(b_n - b) \xrightarrow{\mathcal{P}} 0$ and

$$\sqrt{n}(b_n - b)^2 \frac{1}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \phi'_\delta(u_s - u_t) g_{bb}(Y_t, b'_n) \xrightarrow{\mathcal{P}} 0.$$

As to condition (3.16) in the theorem, it is satisfied if $g_b(Y_t, b)$ is $\mathcal{G}_1^{t-1}$-measurable. This follows from the fact that

$$\mathsf{E}\left[\int (D\Phi_\delta)_{(u_t^m - v)} \cdot \xi_t^m \, dF_m(v)\right] = \mathsf{E}\left[(D\Phi_{\delta,1})_{u_t^m} \cdot \xi_t^m\right] = 0$$

as in equation (7.21).

<div align="right">QED</div>

ACKNOWLEDGMENTS

References

ABHYANKAR, A., L. COPELAND, AND W. WONG (1995): "Nonlinear Dynamics in Real–Time Equity Market Indices: Evidence form the U.K.," Economic Journal, forthcoming.

BAEK, E. G., AND W. A. BROCK (1992): "A Nonparametric Test for Independence of a Multivariate Time Series," Statistica Sinica, 2, 137–156.

BARNETT, W., R. GALLANT, AND M. HINICH (1993): "Detection of Nonlinearity and Chaos: Application to Money Stock," Proceedings of the American Statistical Association, Business and Economics Section.

BARNETT, W., R. GALLANT, M. HINICH, J. JUNGEILGES, D. KAPLAN, AND M. JENSEN (1994): "A Single–Blind Controlled Competition Between Tests for Nonlinearity and Chaos," Department of Economics, Washington University (St. Louis), University of North Carolina (Chapel Hill), University of Texas (Austin), University of Osnabrück, McGill University, Washington University (St. Louis).

BARNETT, W. A., AND M. HINICH (1993): "Has Chaos Been Discovered With Economic Data?," in Nonlinear Dynamics and Evolutionary Economics, ed. by R. H. Day, AND P. Chen, pp. 254–265. Oxford University Press, Oxford.

BIERENS, H. (1990): "A Consistent Conditional Moment Test of Functional Form," Econometrica, 58, 1443–1458.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," Journal of Econometrics, 31, 307–327.

BOLLERSLEV, T., R. Y. CHOU, N. JAYARAMAN, AND K. F. KRONER (1990): "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," Journal of Econometrics, 52(1), 5–60.

BOLLERSLEV, T., R. F. ENGLE, AND D. B. NELSON (1993): "ARCH Models," Department of Finance Working Paper No. 154, Kellog School of Management, Northwestern University.

——— (1994): "ARCH Models," in Handbook of Econometrics, vol. IV, chap. 49, pp. 2959–3038. North Holland, Amsterdam.

BROCK, W. A. (1987): "Notes On Nuisance Parameter Problems in BDS Type Tests for IID," Department of Economics, University of Wisconsin at Madison, revised 1989.

BROCK, W. A., AND W. D. DECHERT (1989): "A General Class of Specification Tests: The Scalar Case," in Business and Economics Statistics Section of the Proceedings of the American Statistical Society, pp. 70–79.

BROCK, W. A., W. D. DECHERT, AND J. A. SCHEINKMAN (1987): "A Test for Independence Based On the Correlation Dimension," Department of Economics, University of Wisconsin at Madison, University of Houston, and University of Chicago.

BROCK, W. A., D. HSIEH, AND B. LEBARON (1991): Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence. MIT Press, Cambridge, Massachusetts.

BROCK, W. A., J. LAKONISHOK, AND B. LEBARON (1992): "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns," Journal of Finance, XLVII, 1731–1764.

BROCK, W. A., AND S. POTTER (1993): "Nonlinear Time Series and Macroeconomics," in *Handbook of Statistics*, ed. by G. S. Maddala, C. R. Rao, AND H. D. Vinod, vol. II, pp. 195–229. North Holland, Amsterdam.

BROCK, W. A., AND C. SAYERS (1988): "Is The Business Cycle Characterized by Deterministic Chaos?," *Journal of Monetary Economics*, 22, 71–90.

CASTI, J. (1992): *Reality Rules: Picturing the World in Mathematics*. John Wiley and Sons, New York.

CHAVAS, J., AND M. HOLT (1991): "On Nonlinear Dynamics: The Case of the Pork Cycle," *American Journal of Agricultural Economics*, 73, 819–828.

DE JONG, R. M. (1992): "The Bierens Test Under Data Dependence," Department of Economics, Southern Methodist University.

DE LIMA, P. (1992a): "A Test for IID Based upon the BDS Statistic," Department of Economics, Johns Hopkins University.

——— (1992b): "Testing Nonlinearities in Economic Time Series Under Moment Condition Failure," Department of Economics, Johns Hopkins University.

DECHERT, W. D. (1989): "An Application of Chaos Theory to Stochastic and Deterministic Observations," Department of Economics, University of Houston.

——— (1994): "The Correlation Integral and the Independence of Gaussian and Related Processes," SSRI Working Paper 9412, University of Wisconsin–Madison.

DENKER, M., AND G. KELLER (1983): "On U-Statistics and von Mises Statistics for Weakly Dependent Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64, 505–522.

——— (1986): "Rigorous Statistical Proceedures for Data from Dynamical Systems," *Journal of Statistical Physics*, 44, 67–93.

ENGLE, R. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987–1007.

FRENCH, K., G. SCHWERT, AND R. STAMBAUGH (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

GALLANT, A., D. HSIEH, AND G. TAUCHEN (1990): "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974-83," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics. Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, ed. by W. A. Barnett, J. Powell, AND G. Tauchen. Cambridge University Press.

GRASSBERGER, P., AND I. PROCACCIA (1983): "Measuring the Strangeness of Strange Attractors," *Physica*, 9D, 189.

GRAUWE, P. D., H. DEWACHTER, AND M. EMBRECHTS (1993): *Exchange Rate Theory*. Blackwell, Oxford.

HIEMSTRA, C., AND J. JONES (1994): "Testing for Linear and Nonlinear Granger Causality," *Journal of Finance*, 49, 1639–1664.

HINICH, M. (1982): "Testing for Gaussianity and Linearity of a Stationary Time Series," *Journal of Time Series Analysis*, 3(3), 169–176.

HSIEH, D. (1987): "Testing for Nonlinear Dependence in Foreign Exchange Rates," *Journal of Business*, 62, 339–368.

————— (1990): "Implications of Nonlinear Dynamics for Financial Risk Management," Fuqua School of Business, Duke University.

————— (1991): "Chaos and Nonlinear Dynamics: Applications to Financial Markets," *Journal of Finance*, 46, 1839–1878.

HSIEH, D., AND B. LEBARON (1988a): "Small Sample Properties of the BDS Statistic, I," in W. A. Brock, D. Hsieh and B. LeBaron, editors, *Nonlinear Dynamics, Chaos, and Instability*, MIT Press, Cambridge, Massachusetts, 1991.

————— (1988b): "Small Sample Properties of the BDS Statistic, II," in W. A. Brock, D. Hsieh and B. LeBaron, editors, *Nonlinear Dynamics, Chaos, and Instability*, MIT Press, Cambridge, Massachusetts, 1991.

————— (1988c): "Small Sample Properties of the BDS Statistic, III," In W. A. Brock and D. Hsieh and B. LeBaron, editors, *Nonlinear Dynamics, Chaos, and Instability*, MIT Press, Cambridge, Massachusetts, 1991.

JADITZ, T. (1994): "Spurious Non-Linear Granger Causality," Bureau of Labor Statistics, Division of Price and Index Numbers, Washington, D.C.

JOHNSON, D., AND R. MCCLELLAND (1994): "A General Dependence Test and Applications," Bureau of Labor Statistics, Division of Price and Index Number Research, Washington, D.C.

KRAGER, H. (1992): "Modelling Cyclical Asymmetry in a Production Series Using Threshold Autoregressive Models," in *Output and Employment Fluctuations*, ed. by K. Zimmermann, vol. 58, #3–4 of *Special Issue, Recherches Economiques de Louvain*, pp. 473–486. Institut de Recherches Economiques, Universite Catholique de Louvain.

KULLBACK, S., AND R. A. LEIBLER (1961): "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79–86.

LAI, D., X. WANG, AND J. WIORKOWSKI (1994): "Local Asymptotic Normality for Multivariate Nonlinear AR Processes," University of Texas (Houston), Gifford Fang Associates and Unversity of Texas (Dallas).

LEBARON, B. (1987): "Nonlinear Puzzles in Stock Returns," in W. A. Brock, D. Hsieh and B. LeBaron, editors, *Nonlinear Dynamics, Chaos, and Instability*, MIT Press, Cambridge, Massachusetts, 1991.

————— (1988): "Stock Return Nonlinearities: Some Initial Tests and Findings," Ph.D. thesis, University of Chicago.

LEE, T.-H., H. WHITE, AND C. W. J. GRANGER (1993): "Testing for Neglected Nonlinearity in Time Series, A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics*, 56, 269–290.

LEPAGE, R., AND L. BILLARD (eds.) (1992): *Exploring the Limits of Bootstrap*. John WIley & Sons, New York.

LIU, T. (1990): "Time Series of Synchronization and Chaos," Ph.D. thesis, Univeristy of California at San Diego.

LORETAN, M., AND P. PHILLIPS (1994): "Testing the Covariance Stationarity of Heavy-Tailed Time Series: An Overview of the Theory with Applications to Several Financial Datasets," *Journal of Empirical Finance*, 1, 211–248.

MCCLELLAND, R., AND D. JOHNSON (1993): "A General Nonparametric Dependence Test for Independence and Applications," Bureau of Labor Statistics, Division of Price and Index Number Research, Washington, D.C.

NEWEY, W. K. (1985): "Maximum Likelihood Specification Testing and Instrumental Score Tests," *Econometrica*.

PAGAN, A., AND A. HALL (1983): "Diagnostic Tests as Residual Analysis," *Econometric Reviews*, 2, 159–218.

PESARAN, M. H., AND S. M. POTTER (eds.) (1993): *Nonlinear Dynamics, Chaos and Econometrics*. John Wiley & Sons, New York.

PETERS, E. (1994): *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*. John Wiley and Sons, New York.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. Springer-Verlag, New York.

POTTER, S. (1995): "Nonlinear Models of Economic Fluctuations," in *Macroeconomics: Developments, Tensions, and Prospects*, ed. by K. D. Hoover. Kluwer Academic, Boston.

POTTER, S. M. (1990): "Nonlinear Time Series and Economic Fluctuations," Ph.D. thesis, University of Wisconsin.

RANDLES, R. H. (1982): "On the Asymptotic Normality of Statistics with Estimated Parameters," *The Annals of Statistics*, 10, 462–474.

SCHEINKMAN, J. A., AND B. LEBARON (1989b): "Nonlinear Dynamics and GNP Data," in *Economic Complexity: Chaos, Sunspots, Bubbles, and Nonlinearity*, ed. by W. A. Barnett, J. Geweke, AND K. Shell. Cambridge University Press.

——— (1989a): "Nonlinear Dynamics and Stock Returns," *Journal of Business*, 62, 311–337.

SCHWERT, G. (1989): "Business Cycles, Financial Crises and Stock Volatility," *Carneige Rochester Conference Series on Public Policy*, 31, 83–126.

SERFLING, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

STOCK, J. (1988): "Estimating Continuous Time Processes Subject to Time Deformation: An Application to Postwar U.S. GNP," *Journal of American Statistical Association*, 83, 77–85.

SUBBA RAO, T., AND M. GABR (1980): "A Test for Linearity of Stationary Time Series," *Journal of Time Series Analysis*, 1(1), 145–158.

TAUCHEN, G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443.

TERASVIRTA, T., D. TJOSTHEIM, AND C. W. J. GRANGER (1994): "Aspects of Modelling Nonlinear Time Series," in *Handbook of Econometrics*, vol. IV, chap. 48, pp. 2917–2957. North Holland, Amsterdam.

THEILER, J. (1988): "Statistical Precision of Dimension Estimators," *Physics Review*, 41A(6), 3038–3051.

THEILER, J., AND S. EUBANK (1994): "Don't Bleach Chaotic Data," *Chaos*, 3, 771–782.

THEILER, J., S. EUBANK, A. LONGTIN, B. GALDRIKIAN, AND J. D. FARMER (1992): "Testing for Nonlinearity in Time Series: The Method of Surrogate Data," *Physica*, 58D, 77–94.

TSAY, R. (1986): "Nonlinearity Tests for Time Series," *Biometrica*, 73(2), 461–466.

VAGA, T. (1994): *Profiting from Chaos: Using Chaos Theory for Market Timing, Stock Selection, and Option Valuation*. McGraw-Hill, New York.

WHITE, H. (1987): "Specification Testing in Dynamic Models," in *Advances in Econometrics*, ed. by T. Bewley. Cambridge University Press, New York.

WOLFF, R. C. L. (1995): "A Poisson Distribution for the BDS Test Statistic for Independence in a Time Series," in *Chaos and Forecasting*, ed. by H. Tong, chap. 5, pp. 109–127. World Scientific, London.

WU, K., R. SAVIT, AND W. A. BROCK (1993): "Statistical Tests for Deterministic Effects in Broad Band Time Series," *Physica*, 69D, 172–188.