Statistica Sinica 2(1992), 137-156

A NONPARAMETRIC TEST FOR INDEPENDENCE OF A MULTIVARIATE TIME SERIES

Ehung G. Baek and William A. Brock

Korea Development Institute/Iowa State University and University of Wisconsin

Abstract: This paper develops a general nonparametric test for the null hypothesis that the vector of time series under scrutiny is temporally and cross sectionally independent. This test can be used to test the adequacy of a fitted model. We can diagnostically test a vector autoregressive model fitted to given data. This procedure is legitimate because the first order asymptotic distribution of the test statistic is robust with respect to the estimated residual vector.

Key words and phrases: Chaos, BDS test, Denker-Keller projection, U-statistic, V-statistic, vector autoregressive model.

1. Introduction

Chaos theory has recently attracted a lot of attention in economics. Discussions with a survey bent appear in Brock (1986), Brock and Sayers (1988), and Baumol and Benhabib (1989). The present paper develops statistical tests that are capable of determining whether the innovations of a conventional multivariate time series model such as a vector autoregressive (VAR) model are a deterministic chaos which is short term forecastable, a nonlinear stochastic process which is partially forecastable or a stochastic process which is not forecastable.

For motivation consider the following chaotic map, called the tent map, which generates the same autocorrelation functions (ACF) as second order white noise

$$x_{t+1} = -2|x_t - 0.5| + 1. \tag{1}$$

More interesting examples of chaotic processes can be generated by letting $e_t = x_t - 0.5$ where x_t is generated by the tent map. Note that, μ-almost surely, $0.5 = \lim_{T\to\infty}(1/T)\sum_{t=1}^{T} x_t = \int_0^1 x\mu(dx)$ where μ is the invariant measure (which is Lebesgue on $[0,1]$) over $[0,1]$ for $\{x_t\}$ generated by the tent map. Then use the sequence $\{e_t\}$, called tent noise, as innovations in the ARMA(p,q) process

$$\Phi(L)y_t = \Psi(L)e_t \tag{2}$$

where L is the backward operator.

A statistician using Box-Jenkins methods will be hard put to detect that (2) is not a stochastic process because the ACF for $\{y_t\}$ will be the same as if $\{e_t\}$ is a true uncorrelated stochastic process such as Independently and Identically Distributed (IID) rather than deterministic chaos with a white ACF.

This problem motivates the following procedure. Consider the family of statistical models

$$y_t = F(\mathbf{Y}_{t-1}; \boldsymbol{\theta}) + e_t \tag{3}$$

where $\mathbf{Y}_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$, $\boldsymbol{\theta}$ is a vector of parameters to be estimated, and $\{e_t\}$ is IID. Moreover F is a known and given function. Assume that $\boldsymbol{\theta}$ can be estimated $\sqrt{T}$ consistently. Denote this estimator by $\boldsymbol{\theta}^*$ for a sample of size T. Then estimated innovations are denoted by e_t^* and satisfy

$$e_t^* = y_t - F(\mathbf{Y}_{t-1}; \boldsymbol{\theta}^*). \tag{4}$$

We now come to the main problem addressed in this paper:

Problem. How can one tell from $\{e_t^*\}$ whether the true innovations are IID or possess hidden structure that is potentially forecastable at least in the short term as in a deterministic chaos?

We attack this problem by use of a new statistical test that compares a measure of the degree of spatial correlation present in the stochastic process $\{e_t\}$ to the same measure computed on an IID counterpart when the process $\{e_t\}$ is "embedded" in m-dimensional space by constructing "m-futures" $e_t^m = (e_t, e_{t+1}, \dots, e_{t+m-1})$. The intuitive idea is to look at the process $\{e_t^m\}$ and measure how well it fills m-space relative to a comparison IID process that has all the same unconditional moments as the original process. In order to explain a special case, let e_t be scalar valued. This comparison is performed by looking at the following measure of spatial correlation, called the *correlation integral*,

$$\begin{aligned} C(m, \epsilon, T) &= \#\{(t,s)\ 1 \le t, s \le T \mid \|e_t^m - e_s^m\| < \epsilon\}/(T-m+1)^2 \\ &= \sum_t \sum_s 1(e_t^m, e_s^m; \epsilon)/(T-m+1)^2 \\ &= \sum_t \sum_s \Big[\prod_{i=0}^{m-1} 1(e_{t+i}, e_{s+i}; \epsilon) \Big] \Big/ (T-m+1)^2 \end{aligned} \tag{5}$$

where $\|\cdot\|$ is the sup norm, and the indicator function $1(a, b; \epsilon)$ takes on the value 1 for the event $\|a - b\| < \epsilon$ and takes on the value 0 otherwise. If $\{e_t\}$ were IID it was shown in Brock and Dechert (1988a) that

$$C(m,\epsilon,T) \to [\lim C(1,\epsilon,T)]^m \quad \text{a.s. as } T \to \infty. \tag{6}$$

This result suggests looking at the following statistic, called BDS (Brock-Dechert-Scheinkman (1988)) or W statistic,

$$W(m,\epsilon,T) = T^{1/2}(C(m,\epsilon,T) - C(1,\epsilon,T)^m) \tag{7}$$

which converges in distribution to a normal distribution with zero mean and constant variance under the null hypothesis of IID. Furthermore, it was shown by Brock, Dechert, Scheinkman and LeBaron (Brock, et al. hereafter) (1988) that the first order asymptotics of W are the same for $\{e_t^*\}$ as for $\{e_t\}$. This property makes W a useful test of whether the form of the nonlinear or linear model that you estimate is correct. This type of test is sensitive to deterministic chaos or other nonlinear dependence between variables (Hsieh and LeBaron (1988a,b), Hsieh (1989)).

Unfortunately the Brock, et al. (1988) paper treated only the scalar case. In economics, we often face multivariate dynamic time series models rather than univariate models, for instance VAR or multivariate Autoregressive Conditional Heteroscedasticity (ARCH) model. Therefore we need to develop the vector version of the BDS test; and this is the aim of the current paper.

2. Notations and Assumptions

The BDS type tests may be extended to vectors of time series in the following way. Let $\{u_{i,t}\}$, $i = 1,2,\ldots,N$; $t = 1,2,\ldots$ be strictly stationary. Let the null hypothesis be $H_0 : \{u_{i,t}\}$ is independent across all i, t; $i = 1,\ldots,N$; $t = 1,2,\ldots$.

We develop the V-statistic form here, because the U-statistic form is similar but involves more notation. Since we are going to use Denker and Keller's (1983) projection method for general V-statistics and the delta method (Serfling (1980), p.118), the notation is designed to suggest this. For simplification, endpoint problems are ignored, and we define the following model and notation.

(i) (Model) The past of an N-vector stochastic process $\{\boldsymbol{y}_s\}_{s=0}^{\infty}$ to time t is written as $\boldsymbol{Y}_t = (\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots)$; $\{\boldsymbol{y}_t\}$ are generated by, $\boldsymbol{y}_t = F(\boldsymbol{Y}_{t-1}, \boldsymbol{\theta}_1) + \boldsymbol{e}_t$ where $\boldsymbol{\theta}_1$ is a k_1 dimension vector of the data generating function (DGF) F which is twice continuously differentiable (C^2 hereafter). Moreover $E[\boldsymbol{e}_t|\boldsymbol{Y}_{t-1}] = 0$ and $E[\boldsymbol{e}_t\boldsymbol{e}_t'|\boldsymbol{Y}_{t-1}] = \boldsymbol{H}_t$, $\boldsymbol{H}_t = H(\boldsymbol{Y}_{t-1}, \boldsymbol{\theta}_2)$, and $\boldsymbol{\theta}_2$ is a vector of dimension k_2 in the covariance matrix $\boldsymbol{H}_t$. If $\boldsymbol{u}_t = \boldsymbol{H}_t^{-1/2}\boldsymbol{e}_t$, then $E[\boldsymbol{u}_t|\boldsymbol{Y}_{t-1}] = 0$ and $E[\boldsymbol{u}_t\boldsymbol{u}_t'|\boldsymbol{Y}_{t-1}] = \boldsymbol{I}$ where $\boldsymbol{I}$ is an $N \times N$ identity matrix. Here $\boldsymbol{H}_t^{-1/2}$ denotes the inverse of the positive definite symmetric matrix $\boldsymbol{H}_t^{1/2}$ whose elements are continuous functions of $\boldsymbol{H}_t$ (White (1984), p.65). Put $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, and for later use, write $\boldsymbol{u}_t$ as $G(\boldsymbol{Y}_t, \boldsymbol{\theta}) = \boldsymbol{H}_t^{-1/2}(\boldsymbol{y}_t - F(\boldsymbol{Y}_{t-1}, \boldsymbol{\theta}))$. Then $\boldsymbol{u}_t = G(\boldsymbol{Y}_t, \boldsymbol{\theta})$ is the standardized actual innovation vector and $\boldsymbol{u}_t^* = G(\boldsymbol{Y}_t, \boldsymbol{\theta}^*)$ is the standardized estimated innovation vector from solving the data generating process for $\boldsymbol{e}_t$. For practical use we assume that G is a function of finite past of the process to avoid technicalities caused by an infinite past.

(ii) (m_i-Futures) Let m_i-futures of ith element of an N-vector $\boldsymbol{u}$ be denoted by $\boldsymbol{u}_{i,t}^{m_i} = (u_{i,t}, u_{i,t+1}, \cdots, u_{i,t+m_i-1})$, and the collection of all these futures for all i at time t be written as $\boldsymbol{u}_t^{\boldsymbol{m}} = (u_{1,t}^{m_1}, u_{2,t}^{m_2}, \cdots, u_{N,t}^{m_N})$. We use boldface $\boldsymbol{m}$ to denote a vector.

(iii) (Symmetric Kernel) $h(u, v; \epsilon) : R^2 \to R$ is a symmetric kernel with parameter ϵ and $h \in C^2$. Define the projection of a kernel as $h_1(u;\epsilon) = E[h(u,v;\epsilon)|u]$. Generally a symmetric kernel and the projection of it in higher dimensional space is defined in a similar way. If $h(u,v;\epsilon) = 1(u,v;\epsilon)$, note that $h_1(u;\epsilon) = F(u+\epsilon) - F(u-\epsilon)$ where $F(u) = \text{Prob}(U \le u)$. Also note that $Eh_1'(u;\epsilon) = \int[F'(u+\epsilon) - F'(u-\epsilon)]dF(u) = 0$ where $'$ attached to a function denotes the first derivative. This fact will be used to motivate assumption (II) below.

(iv) (Correlation Integral) The correlation integral in the vector case is defined similar to (5). That is

$$C(\boldsymbol{m}, \epsilon, T) = (1/T^2) \sum_{t=1}^{T} \sum_{s=1}^{T} h(\boldsymbol{u}_t^{\boldsymbol{m}}, \boldsymbol{u}_s^{\boldsymbol{m}}; \epsilon), \quad (8)$$

where h is a symmetric kernel defined on $R^{\Sigma m_i} \times R^{\Sigma m_i}$. If the kernel h is the indicator function, (8) becomes

$$(1/T^2) \sum_{t=1}^{T} \sum_{s=1}^{T} \left[\prod_{i=1}^{N} \prod_{j=0}^{m_i-1} h(u_{i,t+j}, u_{i,s+j}; \epsilon) \right]. \quad (9)$$

The approximation of the indicator kernel to a smooth kernel such as h is discussed in Theorem 5 and Section 5 in detail. Furthermore, the following three notations are used in the vector case of the correlation integral,

$$C(i, \epsilon, T) = (1/T^2) \sum_t \sum_s [h(u_{i,t}, u_{i,s}; \epsilon)], \quad (10)$$

$$C_i = Eh(u_{i,t}, u_{i,s}; \epsilon), \quad (11)$$

$$C_{\boldsymbol{m}} = \prod_{i=1}^{N} \{E[h(u_{i,t}, u_{i,s}; \epsilon)]^{m_i}\} = \prod_{i=1}^{N} \{(C_i)^{m_i}\}. \quad (12)$$

For simplicity $C(\boldsymbol{m},\epsilon,T)$, $C(i,\epsilon,T)$, $h(a,b;\epsilon)$ and $h_1(a;\epsilon)$ will be written as $C(\boldsymbol{m})$, $C(i)$, $h(a,b)$ and $h_1(a)$ for fixed ϵ, i.

(v) (W Statistic) $$W(C(\boldsymbol{m}),C(i)) = C(\boldsymbol{m}) - \prod_{i=1}^{N}[C(i)]^{m_i}. \tag{13}$$

W is a measure of difference between a stochastic process that is non-IID across t or non-independent across i and a stochastic process that is IID across t and independent across i. It is computed from the $\{u_{i,t}\}$ process. In the population W will be zero if $\{u_{i,t}\}$ is IID across time t and independent across i.

(vi) (Kernel $\mathcal{K}$) Let $\boldsymbol{x}_t = \boldsymbol{u}_t^{\boldsymbol{m}}$, and $\boldsymbol{x}_s = \boldsymbol{u}_s^{\boldsymbol{m}}$ where $\boldsymbol{u}_t^{\boldsymbol{m}}$ is defined in (ii). Define a kernel

$$\begin{aligned}\mathcal{K}(\boldsymbol{x}_t,\boldsymbol{x}_s) &= \prod_{i=1}^{N}\prod_{j=0}^{m_i-1} h\big(u_{i,t+j},u_{i,s+j}\big) - C_{\boldsymbol{m}} \\ &\quad - \sum_{i=1}^{N}\Big[m_i(C_i)^{m_i-1}\prod_{j\neq i}^{N}(C_j)^{m_j}(h(u_{i,t},u_{i,s}) - C_i)\Big]. \end{aligned}\tag{14}$$

(vii) (Gradient) Let $\nabla_{\boldsymbol{u}}\mathcal{K}_{t,s}$ be the gradient of dimension $2(m_1+\cdots+m_N)\times 1$ of the kernel $\mathcal{K}$ at $\boldsymbol{u}$. That is, $\nabla_{\boldsymbol{u}}\mathcal{K}_{t,s} = [\partial\mathcal{K}/\partial u_{1,t},\ldots,\partial\mathcal{K}/\partial u_{1,t+m_1-1},\partial\mathcal{K}/\partial u_{2,t}, \ldots, \partial\mathcal{K}/\partial u_{2,t+m_2-1}, \ldots, \partial\mathcal{K}/\partial u_{N,t},\ldots,\partial\mathcal{K}/\partial u_{N,t+m_N-1}, \partial\mathcal{K}/\partial u_{1,s}, \ldots, \partial\mathcal{K}/\partial u_{1,s+m_1-1}, \partial\mathcal{K}/\partial u_{2,s}, \ldots, \partial\mathcal{K}/\partial u_{2,s+m_2-1}, \ldots \partial\mathcal{K}/\partial u_{N,s},\ldots, \partial\mathcal{K}/\partial u_{N,s+m_N-1}]$. Define $u_{i,t+j} = G_i(\boldsymbol{Y}_{t+j},\boldsymbol{\theta})$, $u_{i,s+j} = G_i(\boldsymbol{Y}_{s+j},\boldsymbol{\theta})$ for $i=1,\ldots,N$, $j=0,\ldots,m_i-1$ for given i, and let θ_i be the ith element of a vector $\boldsymbol{\theta}$. Denote $D_{\theta_i}G_{t,s} = [\partial G_1(\boldsymbol{Y}_t,\boldsymbol{\theta})/\partial\theta_i, \ldots, \partial G_1(\boldsymbol{Y}_{t+m_1-1},\boldsymbol{\theta})/\partial\theta_i, \ldots, \partial G_N(\boldsymbol{Y}_t,\boldsymbol{\theta})/\partial\theta_i, \ldots, \partial G_N(\boldsymbol{Y}_{t+m_N-1},\boldsymbol{\theta})/\partial\theta_i, \partial G_1(\boldsymbol{Y}_s,\boldsymbol{\theta})/\partial\theta_i, \ldots, \partial G_1(\boldsymbol{Y}_{s+m_1-1},\boldsymbol{\theta})/\partial\theta_i, \ldots, \partial G_N(\boldsymbol{Y}_s,\boldsymbol{\theta})/\partial\theta_i, \ldots, \partial G_N(\boldsymbol{Y}_{s+m_N-1},\boldsymbol{\theta})/\partial\theta_i]$. Therefore $D_{\theta_i}G_{t,s}$ is a vector of dimension $2(m_1+\cdots+m_N)\times 1$, and it has the same dimension as $\nabla_{\boldsymbol{u}}\mathcal{K}_{t,s}$. Denote $\nabla_{\boldsymbol{u}}\mathcal{K}_{t,s}$ and $D_{\theta_i}G_{t,s}$ by $\nabla_{\boldsymbol{u}}\mathcal{K}$ and $D_{\theta_i}G$ without the subscripts t and s for simplicity.

(viii) (Function J) The kernel $\mathcal{K}$ is written as a function J of all observations and parameter vectors based on the inverse function $J(\boldsymbol{Y}_{t+\|\boldsymbol{m}\|-1},\boldsymbol{Y}_{s+\|\boldsymbol{m}\|-1};\boldsymbol{\theta}) = \mathcal{K}[G_1(\boldsymbol{Y}_t,\boldsymbol{\theta}),\ldots,G_1(\boldsymbol{Y}_{t+m_1-1},\boldsymbol{\theta}),\ldots,G_N(\boldsymbol{Y}_t,\boldsymbol{\theta}), \ldots, G_N(\boldsymbol{Y}_{t+m_N-1},\boldsymbol{\theta}),G_1(\boldsymbol{Y}_s,\boldsymbol{\theta}), \ldots,G_1(\boldsymbol{Y}_{s+m_1-1},\boldsymbol{\theta}),\ldots,G_N(\boldsymbol{Y}_s,\boldsymbol{\theta}),\ldots,G_N(\boldsymbol{Y}_{s+m_N-1},\boldsymbol{\theta})]$ where $\|\boldsymbol{m}\| = \max(m_1, m_2,\ldots,m_N)$.

(ix) (Function $\tilde{K}$) Let $\tilde{K}(\boldsymbol{u}_i,\boldsymbol{u}_j,\boldsymbol{u}_k) = (1/3!)\sum_\sigma h(\boldsymbol{u}_i,\boldsymbol{u}_j)h(\boldsymbol{u}_j,\boldsymbol{u}_k)$ where the summation is over all possible permutations σ of indices (i,j,k).

(x) (Function H, L) Let $H(\mathbf{Y}_t, \mathbf{Y}_s, \mathbf{Y}_r; \boldsymbol{\theta}, i) = \tilde{K}(G_i(\mathbf{Y}_t, \boldsymbol{\theta}), G_i(\mathbf{Y}_s, \boldsymbol{\theta}), G_i(\mathbf{Y}_r, \boldsymbol{\theta}))$ and $L(\mathbf{Y}_t, \mathbf{Y}_s; \boldsymbol{\theta}, i) = h(G_i(\mathbf{Y}_t, \boldsymbol{\theta}), G_i(\mathbf{Y}_s, \boldsymbol{\theta}))$. Also, $\partial H / \partial \theta_j$, $\partial^2 L / (\partial \theta_j \partial \theta_k)$, $\partial^2 J / (\partial \theta_j \partial \theta_k)$ will be abbreviated as H_j, $L_{j,k}$, and $J_{j,k}$ to clarify kernel functions.

We develop the test method and investigate its properties through the next five theorems, which are based on the following assumptions:

Assumptions

(I) (Uniform Mixing Condition) The DGF generates a stochastic vector process $\{\mathbf{y}_t\}$ that satisfies a uniform mixing assumption of Denker and Keller (1983, Theorem 1, p.507).

(II) (Moment Condition) $\mu_{\theta_i} = E[\nabla_{\mathbf{u}} \mathcal{K} \cdot D_{\theta_i} G] = 0$ for all θ_i, (15)
where $\cdot$ is the scalar product.

Remark. The condition $Eh_1'(u) = 0$, in the univariate case, implies that (II) holds in many applications. If h is the indicator kernel, $1(u_t, u_s; \epsilon)$ then $Eh_1'(u_t) = 0$. Even though the indicator function is not differentiable Theorem 5 below shows that we can approximate $1(u_t, u_s; \epsilon)$ with C^2 kernels.

(III) (Asymptotic Normality) $T^{1/2}(\theta_i^* - \theta_i) \xrightarrow{\text{d}} N(0, V)$ for every θ_i. (16)

(IV) (Compactness of Parameter Space) There is a compact set Ω such that the range of $\boldsymbol{\theta}^*$ is contained in Ω for all T.

(V) (Bounded Moment Condition) All kernels are non-degenerate, i.e. the variance of each projection is positive. As in Denker and Keller (1983, p.507) all kernels, h, of V-statistics appearing below have bounded "$2 + d$" moments, $\sup E\{|h|^{2+d}\} < \infty$ for some $d > 0$. Here "sup" denotes supremum of the expectation over all permutations of temporal arguments.

(VI) (Smoothness) All kernels appearing below are at least C^2. Even if the non-differentiable indicator function is used as a kernel, Theorem 5 shows that an approximating sequence of C^2 kernels always exists such that Eh_1' is approximately zero and the asymptotic variance of the test statistic in Theorems 1 and 6 is continuous in the approximating sequence.

(VII) (Continuity) For X equal to $H_j, L_{j,k}, J_{j,k}$ in (viii) and (x), $\sup\{E|X(\cdot, \cdot; \boldsymbol{\theta})|\}$ is continuous in $\boldsymbol{\theta}$. Recall the sup is taken over all temporal permutations as in Denker and Keller (1983, p.507).

3. A Test for Independence of a Vector Time Series

A test for independence of a vector of time series is developed in this section. All proofs of theorems are put in the Appendix. Under the null hypothesis of temporal and cross sectional independence we have

Theorem 1. *Assume* $\{u_{i,t}\}$ *is* IID *across* t *and independent across* i. *Assumption* (V) *holds for the kernels in the* $C(m)$, W *statistic and in the* $\tilde{K}$ *statistic in* (21) *below. Then*

$$T^{1/2}[W(C(m),C(i))] \xrightarrow{\mathrm{d}} N(0,V_m) \quad as\ T\to\infty. \tag{17}$$

And

$$\begin{aligned} V_m = 4\Big[&\prod_{i=1}^{N}(K_i)^{m_i} - \prod_{i=1}^{N}(C_i)^{2m_i} \\ &+ \sum_{i=1}^{N} m_i(m_i-2)(C_i)^{2(m_i-1)}(K_i-(C_i)^2)\prod_{j\neq i}^{N}(C_j)^{2m_j} \\ &+ 2\sum_{p=2}^{\|m\|}\Big\{\prod_{i=1}^{N} K_i^{\max(m_i-p+1,0)}(C_i)^{2(p-1)} - \prod_{i=1}^{N}(C_i)^{2m_i} \\ &\quad - \sum_{i=1}^{N} m_i(K_i-(C_i)^2)(C_i)^{2(m_i-1)}\prod_{j\neq i}^{N}(C_j)^{2m_j}\Big\}\Big] \end{aligned} \tag{18}$$

where

$$C_i = Eh(u_{i,t+j},u_{i,s+j}) \quad and$$
$$K_i = \begin{cases} E[h(u_{i,r+j},u_{i,s+j})h(u_{i,s+j},u_{i,t+j})] & when\ m_i-p+1>0 \\ (C_i)^2 & otherwise. \end{cases}$$

The variance V_m is consistently estimated by using the following quantities:

$$C(i) = (1/T^2)\sum_t\sum_s h(u_{i,t},u_{i,s}) \quad \text{for } C_i \tag{19}$$

$$K(i) = (1/T^3)\sum_t\sum_s\sum_r[h(u_{i,t},u_{i,s})h(u_{i,s},u_{i,r})] \tag{20}$$

$$\tilde{K}(i) = (1/T^3)\sum_t\sum_s\sum_r \tilde{k}(u_{i,t},u_{i,s},u_{i,r}) \quad \text{for } K_i \text{ where} \tag{21}$$

$$\tilde{k}(u_{i,t},u_{i,s},u_{i,r}) = (1/3!)\sum_\sigma h(u_{i,t},u_{i,s})h(u_{i,s},u_{i,r}). \tag{22}$$

A consistent estimator of V_m is obtained by replacing K_i by $\tilde{K}(i)$ and C_i by $C(i)$. The nonsymmetric kernel in (20) can be symmetrized without loss of generality via (22). Assumption (V) on the kernel $\tilde{k}$ and application of Denker and Keller (1983, Theorem 1) implies the estimator $\tilde{K}(i)$ converges in probability to K_i. Application of convergence of $C(i)$ and $\tilde{K}(i)$ proves that the estimator of the variance converges to V_m in probability. Hence, if one computes $T^{1/2}(V_m)^{-1/2}W$ as the consistent estimator of V_m we have a statistic that converges in distribution to $N(0,1)$ asymptotically under the null hypothesis.

For practical use, under (II), we show that the first order asymptotics of the W statistic in (17) are the same for $\{u^*_{i,t}\}$ and for $\{u_{i,t}\}$ in the next theorem. Superscript * denotes estimated values.

Theorem 2. *Assume the same conditions of Theorem* 1. *Assume* (I), (II) *on* J_{θ_i}, (III) *and* (V) *on the kernels* $J_{\theta_i}(Y_{t+\|m\|-1}, Y_{s+\|m\|-1}; \theta)$, $L_{\theta_i}(Y_t, Y_s; \theta)$ *and* (IV) *and* (VII) *with* X *equal to* $J_{\theta_i\theta_j}(= J_{i,j})$. *Then*

$$T^{1/2}[W(C^*(m), C^*(1)) - W(C(m), C(1))] \xrightarrow{\text{P}} 0 \quad \text{as } T \to \infty. \tag{23}$$

To complete the proof of Theorem 2 the second order terms of (A5), (A6) in Appendix are disposed of in the following Lemma.

Lemma 3. *Consider the second order terms* M_1, M_2 *in* (A5) *and* (A6). *Assume the kernel* $h(u,v)$ *is bounded between* 0 *and* $B < \infty$, *and assume* (IV), (V), (VI) *and* (VII). *Then* $T^{1/2}M_i \xrightarrow{\text{P}} 0$ *as* $T \to \infty$ *for* $i = 1$ *and* 2.

While Theorem 2 shows that the first order asymptotic distribution of W is invariant to evaluation at u^*_t or u_t we still need an estimator of the variance V_m for the statistic to be of practical use.

Theorem 4. *The estimator* V^*_m, *say, of the asymptotic variance* (18) *of Theorem* 1 *evaluated at all estimated innovations* $u^*_{i,t}$ *is a consistent estimator for* V_m.

Applications of the W statistic developed in this paper use the nonsmooth kernel function like the indicator function. Since this kernel is not smooth, it does not satisfy the twice continuously differentiable assumption posed in (i). However the formula (18) for the variance, V_m, in Theorem 1, is continuous in its arguments C_i and K_i. The next theorem based on Brock and Dechert (1988b), shows that approximation of the nonsmooth kernel by a smooth kernel can be accomplished so that V_m is continuous in the approximation.

Theorem 5. *For every* $\delta > 0$ *we can find a kernel* $h(u,v)$ *that satisfies* (V), (VI) *and* (VII) *such that* (i) *the absolute value of the difference between the variance formula* (18) *computed at the indicator kernel and the variance formula*

(18) *computed at the smooth kernel h is less than* $f(\delta)$, (ii) $|Eh_1'| < g(\delta)$ *where* $f(\delta)$ *and* $g(\delta)$ *converge to* 0 *as* δ *goes to* 0. *If* $F(\cdot)$ *is symmetric about zero, the approximating sequence of smooth kernels can be chosen such that* $Eh_1'(u) = 0$ *for each term of sequence.*

Remark. Note, however, that Theorem 5 does not allow one to assert (23) for the indicator kernel itself. We have not been able to show (23) holds for the indicator kernel. Hsieh and LeBaron (1988a,b) have shown by Monte Carlo work that (23) approximately holds for the BDS test for a class of estimated models with additive errors. Since the BDS test is a special case of the test in the current paper we suspect (23) will hold approximately for our test for a useful class of estimated models with additive errors.

4. An Application of the Multivariate Test of Independence to VAR

We show here how the multivariate test of independence can be practically applied to autoregressive models. To make our test operational we need to show the following desirable property. That is, if you estimate the correct null autoregressive model to your vector of time series then the W statistic in (17) evaluated at the estimated standardized residuals has the same asymptotic distribution as the W statistic evaluated at the true standardized innovations. Call this "the invariance property". This is true provided your estimation procedure is $\sqrt{T}$ consistent.

We will make assumptions that suffice for the above property to hold in a broad class of cases. Denote the true covariance matrix by $\boldsymbol{\Gamma}$ and the estimated covariance matrix by $\boldsymbol{\Gamma}^*$. Suppose

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \tag{24}$$

with

$$E[\boldsymbol{u}_t|\boldsymbol{y}_{t-s},\ \text{all } s > 0] = 0 \quad \text{and } E[\boldsymbol{u}_t\boldsymbol{u}_t'|\boldsymbol{y}_{t-s},\ \text{all } s > 0] = \boldsymbol{\Gamma}, \tag{25}$$

and

$\{\boldsymbol{u}_t\}$ a stationary $N \times 1$ vector IID stochastic process
with finite fourth momemts. (26)

If (24) is the true data generating process and we use the right model to fit the given data, the standardized estimated VAR residuals are asymptotically IID across time and independent across variables. The standardized VAR residuals are given by

$$\boldsymbol{v}_t^* = (\boldsymbol{\Gamma}^*)^{-1/2}(\boldsymbol{I} - \boldsymbol{A}^*(L))\boldsymbol{y}_t \tag{27}$$

where A^* is estimate of A. In (27) $A^*(L)y_t$ is equal to A^*y_{t-1}. Since $\{v_t^*\}$ is a sequence of N-vectors we can use them as arguments for the W test.

It is easy to show that the invariance property holds for the univariate AR(1) model with known variance. We show that the invariance property holds in the more complicated bivariate VAR(1) model. Higher order VAR models can be handled in a similar way.

Consider the following bivariate VAR(1) model,

$$y_{1,t} = a_0 + a_1 y_{1,t-1} + a_2 y_{2,t-2} + u_{1,t}$$
$$y_{2,t} = b_0 + b_1 y_{1,t-1} + b_2 y_{2,t-2} + u_{2,t},$$

where

$$\begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \sim N(\mathbf{0}_{2\times 1}, \boldsymbol{\Gamma}) \text{ and } \boldsymbol{\Gamma} = \begin{bmatrix} \sigma_1 & \sigma_3 \\ \sigma_3 & \sigma_2 \end{bmatrix} \text{ is positive definite.}$$

Let us use the column stacking operation vec(·) to represent the vector $\boldsymbol{\theta}$. Let

$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} = (\boldsymbol{\Gamma})^{-1/2}(\boldsymbol{y}_t - \boldsymbol{A}(L)\boldsymbol{y}_t) = G(\boldsymbol{Y}_t, \boldsymbol{\theta}) \quad \text{where } \boldsymbol{\theta} = \text{vec}(\boldsymbol{A}, \boldsymbol{\Gamma}) \text{ and}$$
$$\begin{bmatrix} v_{1,t}^* \\ v_{2,t}^* \end{bmatrix} = (\boldsymbol{\Gamma}^*)^{-1/2}(\boldsymbol{y}_t - \boldsymbol{A}^*(L)\boldsymbol{y}_t) = G(\boldsymbol{Y}_t, \boldsymbol{\theta}^*) \quad \text{where } \boldsymbol{\theta}^* = \text{vec}(\boldsymbol{A}^*, \boldsymbol{\Gamma}^*).$$

Now we are ready to establish the following theorem. The key assumption in Theorem 6 is $Eh_1'(u) = 0$ which Theorem 5 has shown to be innocuous for practical purposes.

Theorem 6. *Assume $Eh_1'(u) = 0$. If (24), (25) and (26) are true, and $\boldsymbol{\Gamma}$ is symmetric and positive definite then the independence test W has the same first order asymptotic distribution whether it is evaluated at $\{v_t^*\}$ or $\{v_t\}$.*

5. Conclusion

This paper has developed a test of independence of a vector of time series. This test statistic has the same first order asymptotics when evaluated at estimated residuals of a null parametric model as when evaluated at the true IID residuals, provided that the null model is correct. This property is not shared by diagnostics such as the autocorrelation function (Box and Jenkins (1976, p.291)). Special cases investigated by Brock, et al. (1988), Baek (1988), Hsieh and LeBaron (1988a,b) and Hsieh (1989) indicate that the vector generalization of the BDS proposed here should have good power properties against broad classes of alternatives. Hence we believe that our diagnostic test has promise of being worked into a test of wide usefulness to applied econometricians.

We did not discuss small sample properties of our test statistic; however, there is some related work pertaining to this. Baek (1988) studied the special case $N = 2$, $m_1 = 1$, $m_2 = 1$ of the W test. He did preliminary Monte Carlos on size and power against several contemporaneously dependent alternatives. The power of Baek's test was compared to two contemporaneous independence tests: (i) Kendall's tau; and, (ii) Blum, Kiefer, Rosenblatt's Cramer Von Mises test (BKR) which is based upon the difference of the joint distribution and the product of the marginals. In the bivariate model, power was calculated for three alternatives: (i) v_t a piecewise linear transformation of u_t; (ii) v_t a sine function of u_t; (iii) contemporaneous version of the ARCH model (Engle (1982)), i.e. the variance of v_t contemporaneously depends upon u_t but its mean does not. The test appears promising. The power is much better than Kendall's test and compares favorably with BKR for some of the alternatives while beating BKR for others. Baek's test beats BKR for alternatives that have many wiggles which confuse the other tests into concluding that the two series are independent. See Table 1.2 in Baek (1988) for contemporaneous independence. The BDS test (7) is also a special case, $N = 1$, of the test proposed in this paper. It is a test of IID for a univariate series. Encouraging Monte Carlo results on performance are reported in Hsieh and LeBaron (1989) and Hsieh (1989, Table 8). Since the performance of the BDS test was quite good for large sample size ($T \geq 500$) and small values of m we hazard the guess that the vector version of BDS propounded here will exhibit similar good performance provided m is small enough and T is large enough.

There is still much to do however. First, we have been deliberately vague about the class of alternatives against which the null hypothesis is being tested. Once a class of alternatives has been chosen the kernel vector can be chosen to maximize some criterion, such as power against this class. Following up on this line of thinking it is natural to try to characterize our test by finding the alternatives against which it has least and most power respectively for fixed T, ϵ, m. We have not done this.

Second, we have not developed a theory of the optimal choice of ϵ, m for a given sample size T. This requires a more precise commitment to a set of alternatives, possibly a simple alternative, before this problem can be stated precisely.

Third, nothing has been said about the choice of kernels $h(\cdot,\cdot)$. The first basic proposition on the limit distribution holds for any vector of smooth kernels. Baek and Brock (1988) showed Theorem 6, i.e. the moment condition, (II), holds for any vector of kernels such that $Eh_1' = 0$. This includes the indicator kernel used in Baek (1988) and Brock, et al. (1988).

Once one has concentrated on a set of alternatives to test against, a possible criterion for choice of kernel vector would be to maximize some useful version of power against the given set of alternatives. This is yet another research problem that is beyond the scope of this paper.

In conclusion, we hope enough has been said in this paper to convince the reader that the diagnostic test for temporal dependence proposed here is worthy of serious attention by the profession.

Acknowledgements

Brock thanks the Guggenheim Foundation, The National Science Foundation (Grant # SEC-8420872), the Wisconsin Alumni Research Foundation, and the Vilas Trust for essential financial support. We thank an associate editor and two referees for very helpful comments. None of the above are responsible for errors or shortcomings in this paper.

Appendix

(1) Proof of Theorem 1

This follows from use of the delta method (Serfling (1980), p.118) and Denker and Keller (1983, Theorem 1, DK hereafter). DK's uniform mixing conditions are trivially satisfied for the stochastic process $\{u_{i,t}^{m_i}\}$ since it is m-dependent. We now use the DK projection method to reduce the $C(m)$ and $C(i)$ statistics to a simpler form.

$$C(m) - C_m = (2/T)\sum_t \Big[\prod_i \prod_j h_1(u_{i,t+j}) - C_m\Big] + R_1, \tag{A1}$$

$$C(i) - C_i = (2/T)\sum_t [h_1(u_{i,t}) - C_i] + R_2 \tag{A2}$$

where both remainder terms multiplied by $T^{1/2}$ go to zero in probability. This representation is nice because it drastically simplifies the central limit theory. Now observe that the statistic $W(C(m), C(i))$ is a smooth function g of the statistics $C(m)$ and $C(i)$, $i = 1, \dots, N$, i.e.

$$W(C(m), C(i)) = g(C(m), C(1), \dots, C(N)) = C(m) - \prod_i C(i)^{m_i}.$$

Expand this function $g(\cdot)$ in a Taylor series about the vector of means of the statistics $C(m)$, $C(i)$ and use (A1), (A2) to obtain the following representation:

$$W = (2/T)\sum_t \Big[\prod_i \prod_j h_1(u_{i,t+j}) - C_m - \sum \Phi_i\{h_1(u_{i,t}) - C_i\}\Big] + R$$

where $\Phi_i = m_i(C_i)^{m_i - 1}\prod_{j \neq i}(C_j)^{m_j}$. We know that $T^{1/2}R \to 0$ in probability as T goes to infinity. Apply the function g, evaluated at all arguments u_t^m, which is equal to

$$2\Big[\prod_i \prod_j h_1(u_{i,t+j}) - \prod_i (C_i)^{m_i} - \sum_i \Phi_i\{h_1(u_{i,t}) - C_i\}\Big]. \tag{A3}$$

Then $W = (1/T)\sum_t g(u_t^{\boldsymbol{m}}) + R$. Therefore

$$T^{1/2}W \xrightarrow{\text{d}} N(0, V_{\boldsymbol{m}}) \quad \text{where}$$

$$V_{\boldsymbol{m}} = E\Big[\{g(u_t^{\boldsymbol{m}})\}^2 + 2\sum_{p=2}^{\|\boldsymbol{m}\|} g(u_t^{\boldsymbol{m}})g(u_p^{\boldsymbol{m}})\Big]. \tag{A4}$$

Now we show that all the second order terms go to zero in probability. $W(C(\boldsymbol{m}), C(i)) = g(C(\boldsymbol{m}), C(1), \dots, C(N)) = C(\boldsymbol{m}) - \prod_i C(i)^{m_i}$.

$$\begin{aligned} W = {} & [C(\boldsymbol{m}) - C_{\boldsymbol{m}}] - \sum_i m_i C_i^{m_i - 1} \prod_{j \neq i} C_j^{m_j} (C(i) - C_i) \\ & - (1/2)\sum_i m_i(m_i - 1)\tilde{C}_i^{m_i - 2} \prod_{j \neq i} C_j^{m_j} (C(i) - C_i)^2 \\ & - (1/2)\sum_i \sum_{j \neq i} m_i \tilde{C}_i^{m_i - 1} m_j \tilde{C}_j^{m_j - 1} \prod_{k \neq i,j} \tilde{C}_k^{m_k} (C(i) - C_i)(C(j) - C_j) \end{aligned}$$

where $\tilde{C}_i$, $\tilde{C}_j$ and $\tilde{C}_k$ are intermediate points between $C(i)$ and C_i, $C(j)$ and C_j, and $C(k)$ and C_k. Then the second order terms are

$$\begin{aligned} & (1/2)T^{1/2}\Big[\sum m_i(m_i - 1)\tilde{C}_i^{m_i - 2} \prod_{j \neq i} C_j^{m_j} \{(2/T)\sum_t (h_1(u_{i,t}) - C_i) + R_{C_i}\}^2 \\ & + \sum_i \sum_{j \neq i} m_i \tilde{C}_i^{m_i - 1} m_j \tilde{C}_j^{m_j - 1} \prod_{k \neq i,j} \tilde{C}_k^{m_k} \{(2/T)\sum_t (h_1(u_{i,t}) - C_i) + R_{C_i}\} \cdot \\ & \{(2/T)\sum_t (h_1(u_{j,t}) - C_j) + R_{C_j}\}\Big]. \end{aligned}$$

Lemma 3 implies that the second order terms go to zero in probability. This proof enables us to derive the variance formula for the case of the indicator kernel function. The variance formula of (18) follows straightforwardly from (A4).

(2) Proof of Theorem 2

Write an exact Taylor expansion for each W by using notation $\mathcal{K}$ in (vii).

$$\begin{aligned} & W(C^*(\boldsymbol{m}), C^*(i)) \\ & = (1/T^2)\sum_t \sum_s \mathcal{K}^* - (1/2)\sum_i m_i(m_i - 1)(\tilde{C}_i^*)^{m_i - 2}\Big(\prod_{j \neq i} \tilde{C}_j^*\Big)[C^*(i) - C_i^*]^2 \end{aligned}$$

$$-(1/2)\sum_i\sum_{j\neq i} m_i \tilde{C}_i^{*\,m_i-1} m_j \tilde{C}_j^{*\,m_j-1}\Big(\prod_{k\neq i,j}\tilde{C}_k^{*\,m_k}\Big)[C^*(i)-C_i^*][C^*(j)-C_j^*] \quad \text{(A5)}$$

$$\begin{aligned} &W(C(\boldsymbol{m}),C(i)) \\ &=(1/T^2)\sum_t\sum_s\mathcal{K}-(1/2)\sum_i m_i(m_i-1)\tilde{C}_i^{m_i-2}\Big(\prod_{j\neq i}\tilde{C}_j^{m_j}\Big)[C(i)-C_i]^2 \\ &-(1/2)\sum_i\sum_{i\neq j} m_i \tilde{C}_i^{m_i-1} m_j \tilde{C}_j^{m_j-1}\Big(\prod_{k\neq i,j}\tilde{C}_k^{m_k}\Big)[C(i)-C_i][C(j)-C_j] \qquad \text{(A6)}\end{aligned}$$

where $\tilde{C}_i^*$, $\tilde{C}_i$ denote evaluation at an intermediate point so that the expansion is exact. The second order term will be disposed of in Lemma 3 where condition (V) on L' will be used. We must show

$$A=T^{1/2}\Big[\sum_t\sum_s(\mathcal{K}^*-\mathcal{K})/T^2\Big]\xrightarrow{\text{P}}0 \quad \text{as } T\to\infty. \qquad \text{(A7)}$$

Insert the formula (A8)

$$u_{i,t}=G_i(\boldsymbol{Y}_t,\boldsymbol{\theta}), \quad u_{i,t}^*=G_i(\boldsymbol{Y}_t,\boldsymbol{\theta}^*) \qquad \text{(A8)}$$

into (A7), expand in a Taylor series about $\boldsymbol{\theta}$ with exact second order remainder to obtain (A9) below. Noting $J_{\theta_i}=[\nabla_{\boldsymbol{u}}\mathcal{K}\cdot D_{\theta_i}G]$, we have

$$\begin{aligned} &T^{1/2}\Big[\sum_t\sum_s\{\mathcal{K}^*(G_1(\boldsymbol{Y}_t,\boldsymbol{\theta}^*),\dots,G_1(\boldsymbol{Y}_{t+m_1-1},\boldsymbol{\theta}^*),\dots,G_N(\boldsymbol{Y}_t,\boldsymbol{\theta}^*),\dots, \\ &\quad G_N(\boldsymbol{Y}_{t+m_N-1},\boldsymbol{\theta}^*),G_1(\boldsymbol{Y}_s,\boldsymbol{\theta}^*),\dots,G_1(\boldsymbol{Y}_{s+m_1-1},\boldsymbol{\theta}^*),\dots,G_N(\boldsymbol{Y}_s,\boldsymbol{\theta}^*),\dots, \\ &\quad G_N(\boldsymbol{Y}_{s+m_N-1},\boldsymbol{\theta}^*))-\mathcal{K}(G_1(\boldsymbol{Y}_t,\boldsymbol{\theta}),\dots,G_1(\boldsymbol{Y}_{t+m_1-1},\boldsymbol{\theta}),\dots,G_N(\boldsymbol{Y}_t,\boldsymbol{\theta}),\dots, \\ &\quad G_N(\boldsymbol{Y}_{t+m_N-1},\boldsymbol{\theta}),G_1(\boldsymbol{Y}_s,\boldsymbol{\theta}),\dots,G_1(\boldsymbol{Y}_{s+m_1-1},\boldsymbol{\theta}),\dots,G_N(\boldsymbol{Y}_s,\boldsymbol{\theta}),\dots, \\ &\quad G_N(\boldsymbol{Y}_{s+m_N-1},\boldsymbol{\theta}))\}/T^2\Big] \\ &=T^{1/2}\Big[\sum_t\sum_s\big(J(\boldsymbol{Y}_{t+\|\boldsymbol{m}\|-1},\boldsymbol{Y}_{s+\|\boldsymbol{m}\|-1};\boldsymbol{\theta}^*)-J(\boldsymbol{Y}_{t+\|\boldsymbol{m}\|-1},\boldsymbol{Y}_{s+\|\boldsymbol{m}\|-1};\boldsymbol{\theta})\big)/T^2\Big] \\ &=T^{1/2}\Big[\sum_t\sum_s\{J(\boldsymbol{Y}_{t+\|\boldsymbol{m}\|-1},\boldsymbol{Y}_{s+\|\boldsymbol{m}\|-1};\boldsymbol{\theta})-J(\boldsymbol{Y}_{t+\|\boldsymbol{m}\|-1},\boldsymbol{Y}_{s+\|\boldsymbol{m}\|-1};\boldsymbol{\theta}) \\ &\quad+\sum_i J_{\theta_i}(\boldsymbol{Y}_{t+\|\boldsymbol{m}\|-1},\boldsymbol{Y}_{s+\|\boldsymbol{m}\|-1};\boldsymbol{\theta})(\theta_i^*-\theta_i) \\ &\quad+(1/2)\sum_i\sum_j J_{i,j}(\boldsymbol{Y}_{t+\|\boldsymbol{m}\|-1},\boldsymbol{Y}_{s+\|\boldsymbol{m}\|-1};\boldsymbol{\theta})(\theta_i^*-\theta_i)(\theta_j^*-\theta_j)\}/T^2\Big]. \qquad \text{(A9)}\end{aligned}$$

Note $\sum_t \sum_s J_{\theta_i}/T^2 \xrightarrow{\text{P}} \mu$ where by (II),

$$\mu_{\theta_i} = EJ_{\theta_i} = E(\nabla_{\boldsymbol{u}}\mathcal{K} \cdot D_{\theta_i}G) = 0. \tag{A10}$$

Let Λ_T be $T^{1/2}(\theta_i^* - \theta_i)$. Now Λ_T is $O_p(1)$ by (III), therefore it is sufficient to show

$$\sum_t \sum_s [\nabla_{\boldsymbol{u}}\mathcal{K} \cdot D_{\theta_i}G]/T^2 \xrightarrow{\text{P}} \mu_{\theta_i} = 0$$

and the second order terms in (A9) converge in probability to zero. Observe $\nabla_{\boldsymbol{u}}\mathcal{K} \cdot D_{\theta_i}G = J_{\theta_i}$ is the derivative of a symmetric kernel $J(\cdot,\cdot;\boldsymbol{\theta})$ with respect to θ_i, hence $\sum_t \sum_s J_{\theta_i}/T^2$ is a V-statistic. Under the mixing condition (I) and the nondegeneracy, and bounded second moment condition (V) on the kernel J_{θ_i}, Denker and Keller Theorem 1 asserts

$$\sum_t \sum_s J_{\theta_i}/T^2 \xrightarrow{\text{P}} EJ_{\theta_i} \quad \text{as } T \to \infty. \tag{A11}$$

But $EJ_{\theta_i} = \mu_{\theta_i} = 0$ by (II). The second order terms in (A9) converge in probability to zero. Since Λ_T converges in distribution to a random variable it is enough to show for each element i, j

$$\sum_t \sum_s J_{i,j}(\cdot,\cdot;\tilde{\boldsymbol{\theta}}^*)/T^{5/2} \xrightarrow{\text{P}} 0 \quad \text{as } T \to \infty \tag{A12}$$

where $\tilde{\boldsymbol{\theta}}^*$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$. To show (A12) it is sufficient to show convergence in L_1. Thus it is sufficient to show there is an upper bound $B(< \infty)$ such that for all elements i, j

$$\sup\{E|J_{i,j}(\cdot,\cdot;\boldsymbol{\theta})|, \text{ all nonnegative } t \text{ and } s\} < B \tag{A13}$$

where the sup is taken over $\boldsymbol{\theta}$ in some compact set K. But (A9) follows from (IV) which states that the values of $\boldsymbol{\theta}^*$ lie in a compact set K which is independent of T, and A6 which states $\sup\{E|J_{i,j}(\cdot,\cdot;\boldsymbol{\theta})|$, all nonnegative t and $s\}$ is continuous in $\boldsymbol{\theta}$.

(3) Proof of Lemma 3

Since m_i is not less than 2 the terms involving $\tilde{C}_i^*$, $\tilde{C}_i$ are bounded above and below. Hence upon division by $T^{1/2}$ and using Serfling (1980, p.19) it is sufficient to show

$$T^{1/2}(C^*(i) - C_i) = O_p(1), \tag{A14}$$

$$T^{1/2}(\dot{C}(i) - C_i) = O_p(1). \tag{A15}$$

Convergence of the second term (A15) follows from the same type of argument used in Theorem 1. The first term (A14) requires attention. The Taylor series expansion gives

$$\begin{aligned} C^*(i) &= \sum_t \sum_s h(u^*_{i,t}, u^*_{i,s})/T^2 = \sum_t \sum_s h\big(G_i(Y_t, \theta^*), G_i(Y_s, \theta^*)\big)/T^2 \\ &= \sum_t \sum_s L(Y_t, Y_s; \theta^*, i)/T^2 \\ &= C(i) + \sum_t \sum_s \Big[\sum_j L_{\theta_j}(Y_t, Y_s; \theta^*, i)(\theta^*_j - \theta_j)\Big]/T^2 \\ &\quad + (1/2)\sum_t \sum_s \Big[\sum_i \sum_j L_{i,j}(Y_t, Y_s; \tilde{\theta}^*, i)(\theta^*_i - \theta_i)(\theta^*_j - \theta_j)\Big]/T^2 \\ &= C(i) + S + U \end{aligned} \tag{A16}$$

where S is the second term and U is the third term of (A16). Here $\tilde{\theta}^*$ is between θ and θ^* and $T^{1/2}(\theta^*_j - \theta_j) \xrightarrow{d} N(0, V_\theta)$ as $T \to \infty$ for each j. So $T^{1/2}S = O_p(1)$. Since (V) allows application of Denker and Keller Theorem 1 to $\sum_t \sum_s L'/T^2$, it follows that $\sum_t \sum_s L_{\theta_i}/T^2 \xrightarrow{P} EL_{\theta_i}$. Furthermore $T^{1/2}U \xrightarrow{P} 0$ by using (IV), (VI) and (VII) with X equal to $L_{i,j}$. Therefore $T^{1/2}(C^*(i) - C_i) = O_p(1)$ and $T^{1/2}(C(i) - C_i) = O_p(1)$. It follows that M_1 and M_2 converge to zero in probability, since

$$T^{1/2}(1/2)\Big[m_i(m_i - 1)\tilde{C}_i^{m_i-2} \prod_{j\neq i} C_j^{m_j}(C^*(i) - C_i)^2\Big] \tag{A17}$$

$= O_p(1)/T^{1/2} \to 0$ as $T \to \infty$, and

$$T^{1/2}(1/2)\Big[m_i\tilde{C}_i^{m_i-1} m_j \tilde{C}_j^{m_j-1} \prod_{k\neq i,j} (\tilde{C}_j)^{m_k}(C^*(i) - C_i)(C^*(j) - C_j)\Big] \tag{A18}$$

$= O_p(1)/T^{1/2} \to 0$ as $T \to \infty$.

(4) Proof of Theorem 4

Noting $V_{\boldsymbol{m}} = V(C_i, K_i)$, we must show

$$V^*_{\boldsymbol{m}} \xrightarrow{P} V_{\boldsymbol{m}} \quad \text{as } T \to \infty \quad \text{where } V^*_{\boldsymbol{m}} = V(C^*(i), \tilde{K}^*(i)). \tag{A19}$$

The estimate $\tilde{K}^*(i)$ can be written

$$\tilde{K}^*(i) = \sum_t \sum_s \sum_r H(Y_t, Y_s, Y_r; \theta^*, i)/T^3. \tag{A20}$$

Expand (A20) in an exact first order Taylor series about $\boldsymbol{\theta}$. One gets $\tilde{K}(i)$ plus first order terms in $(\theta_i^* - \theta_i)$ for $i = 1, \ldots, N(k_1 + k_2)$. $\tilde{K}(i)$ is a consistent estimator of K_i almost surely. We have $T^{1/2}(\theta_i^* - \theta_i) = O_p(1)$ as $T \to \infty$ for each i. By (VII) there is a bound $B < \infty$ such that

$$\sup E\{|H_j(\mathbf{Y}_r, \mathbf{Y}_s, \mathbf{Y}_t; \boldsymbol{\theta}, i)|, \text{ all nonnegative } t,\, s,\, r\} < B < \infty. \qquad \text{(A21)}$$

Here we may show as we did for the second order terms in Theorem 2 that the first order terms go to zero in distribution. Therefore V_m^* converges to V_m in probability since $\tilde{K}^*(i) \xrightarrow{\text{P}} K_i$ and $C^*(i) \xrightarrow{\text{P}} C_i$.

(5) Proof of Theorem 5

(i) We will sketch the main idea of the proof for simple univariate case. Let $\epsilon > 0$ be given and for $\delta < \epsilon$ let h_δ be an even C^1 function such that

$$h_\delta(u,v) = \begin{cases} 1 & \text{when } |u - v| \le \epsilon - \delta \\ 0 & \text{when } |u - v| \ge \epsilon \end{cases}$$

as well as $0 \le h_\delta(u,v) \le 1$, $|\partial h_\delta(u,v)/\partial u| \le 2\delta^{-1}$. Let $A = \{(u,v) | \epsilon - \delta < |u - v| < \epsilon\}$. Then for any h_δ,

$$\iint \left[h_\delta(u,v) - 1(u,v;\epsilon)\right]^p dF(u)dF(v) \le \iint_A dF(u)dF(v) \to 0 \quad \text{as} \quad \delta \to 0.$$

Therefore a smooth kernel $h_\delta(u,v)$ approximates the indicator function $1(u,v;\epsilon)$ in L_p norm. Define

$$C_\delta = \iint h_\delta(u,v)dF(u)dF(v) \text{ and } K_\delta = \iiint h_\delta(u,v)h_\delta(v,w)dF(u)dF(v)dF(w).$$

Then we can show that $|C_\delta - C|$ and $|K_\delta - K|$ both depend on the value δ, and $\lim_{\delta \to 0} C_\delta = C$ and $\lim_{\delta \to 0} K_\delta = K$. Since the variance formula corresponding to the multivariate formula (18) is only a function of C and K, it can also be approximated by using a smooth kernel which satisfies the above condition.

(ii) $$\lim_{\delta \to 0} \int \left[d\Big(\int h_\delta(u,v)dF(v)\Big)/du\right] dF(u) = \lim_{\delta \to 0} \iint h_\delta(u,v)f'(u)f(v)dudv$$
$$= \iint 1(u,v;\epsilon)f'(u)f(v)dudv = \int [f(v+\epsilon) - f(v-\epsilon)]f(v)dv = 0.$$

Therefore $Eh'_{\delta 1}(u)$ depends on the value δ, but it converges to 0 as δ goes to 0. Finally

$$Eh'_{\delta 1}(u) = \int \left[\partial\Big(\int h_\delta(u,v)dF(v)\Big)/\partial u\right] dF(u) = \iint (\partial h_\delta(u,v)/\partial u)dF(v)dF(u)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{u} (\partial h_\delta(u,v)/\partial u)dF(v)dF(u) + \int_{-\infty}^{\infty}\int_{u}^{\infty} (\partial h_\delta(u,v)/\partial u)dF(v)dF(u). \qquad \text{(A22)}$$

Symmetry of $F(\cdot)$ implies that $\int_a^b g(u)dF(u) = \int_{-b}^{-a} g(-u)dF(u)$. Therefore

$$\int_{-\infty}^{\infty}\int_{u}^{\infty}(\partial h_\delta(u,v)/\partial u)dF(v)dF(u) = -\int_{-\infty}^{\infty}\int_{-\infty}^{u}(\partial h_\delta(u,v)/\partial u)dF(v)dF(u).$$

Dropping the subscript δ, we obtain the above theorem.

(6) Proof of Theorem 6

If we replace all $u_{i,t}$ variables with $v_{i,t}$ in the proof of Theorem 2, it is sufficient to show that the moment condition, (II), holds. The idea of the proof is along the same lines as the previous proofs. We will represent $\boldsymbol{v}_t^*$ by $\boldsymbol{Y}_t$ including current y and past values of y, and $\boldsymbol{\Gamma}^*$, and $\boldsymbol{A}^*$. The next step is a Taylor expansion of the W statistic about a true parameter. The first order terms will disappear by the moment condition, (II), and the second and higher order terms also converge to zero in the same way as the proof of Theorem 2. To show that the moment condition, (II), holds we use (A9) and (A10).

Here since

$$\begin{aligned}\mathcal{K}(\boldsymbol{v}_t^{(2,2)}, \boldsymbol{v}_s^{(2,2)}) &= h(v_{1,t}, v_{1,s})h(v_{1,t+1}, v_{1,s+1})h(v_{2,t}, v_{2,s})h(v_{2,t+1}, v_{2,s+1}) \\ &\quad - C_{(2,2)} - 2C_1\left(C_{(2,2)}\right)^2\left(h(v_{1,t}, v_{1,s}) - C_1\right) \\ &\quad - 2C_{(2,2)}(C_1)^2\left(h(v_{2,t}, v_{2,s}) - C_{(2,2)}\right) \text{ and} \\ \nabla_{\boldsymbol{v}}\mathcal{K}(\boldsymbol{v}_t^{(2,2)}, \boldsymbol{v}_s^{(2,2)}) &= [\partial\mathcal{K}/\partial v_{1,t}, \partial\mathcal{K}/\partial v_{1,t+1}, \partial\mathcal{K}/\partial v_{2,t}, \partial\mathcal{K}/\partial v_{2,t+1}, \\ &\quad \partial\mathcal{K}/\partial v_{1,s}, \partial\mathcal{K}/\partial v_{1,s+1}, \partial\mathcal{K}/\partial v_{2,s}, \partial\mathcal{K}/\partial v_{2,s+1}]. \end{aligned} \tag{A23}$$

Now we must show that $E[(\nabla_{\boldsymbol{v}}\mathcal{K} \cdot D_{\theta_i}G)] = 0$ for all elements of the parameter vector, where $G = G(\boldsymbol{Y}_t, \boldsymbol{\theta})$, in order to express the $\boldsymbol{v}_t$ vector. We will show the moment condition when $\theta_i = \sigma_1$ or a_1 in VAR(1) model. Other cases will be shown in a similar way.

Define $\boldsymbol{\Gamma}^{-1/2} = [\eta_{ij}]$ for $i, j = 1$ or 2. Then $\eta_{ij} = \eta_{ij}(\sigma_1, \sigma_2, \sigma_3)$. Thus

$$\boldsymbol{v}_t = \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} = \boldsymbol{\Gamma}^{-1/2}\boldsymbol{u}_t = [\eta_{ij}]\boldsymbol{u}_t.$$

However $\boldsymbol{u}_t = \boldsymbol{\Gamma}^{1/2}\boldsymbol{v}_t$ implies $\partial v_{1,t}/\partial\sigma_1$, and $\partial v_{2,t}/\partial\sigma_1$ are both linear functions of $v_{1,t}$ and $v_{2,t}$. Let $\partial v_{1,t}/\partial\sigma_1 = av_{1,t} + bv_{2,t}$ and $\partial v_{2,t}/\partial\sigma_1 = cv_{1,t} + dv_{2,t}$. Therefore

$$\begin{aligned} D_{\sigma_1}G_{t,s} = [&av_{1,t}+bv_{2,t},\ av_{1,t+1}+bv_{2,t+1},\ cv_{1,t}+dv_{2,t},\ cv_{1,t+1}+dv_{2,t+1}, \\ &av_{1,s}+bv_{2,s},\ av_{1,s+1}+bv_{2,s+1},\ cv_{1,s}+dv_{2,s},\ cv_{1,s+1}+dv_{2,s+1}] \end{aligned} \tag{A24}$$

where a, b, c and d are all functions of η_{ij} and $\partial\eta_{ij}/\partial\sigma_1$ for i, $j = 1$ or 2. If we expand a Taylor series about σ_1,

$$\begin{aligned} E[\nabla_{\boldsymbol{v}}\mathcal{K}\cdot D_{\sigma_1}G] = 2E\big[\{h'(v_{1,t},v_{1,s})h(v_{1,t+1},v_{1,s+1})h(v_{2,t},v_{2,s}) \\ h(v_{2,t+1},v_{2,s+1}) - 2C_1(C_2)^2h'(v_{1,t},v_{1,s})\}\{av_{1,t}+bv_{2,t}\}\big] \\ + 2E\big[\{h'(v_{1,t+1},v_{1,s+1})h(v_{1,t},v_{1,s})h(v_{2,t},v_{2,s})h(v_{2,t+1},v_{2,s+1})\} \\ \{av_{1,t+1}+bv_{2,t+1}\}\big] + 2E\big[\{h'(v_{2,t},v_{2,s})h(v_{1,t},v_{1,s})h(v_{1,t+1},v_{1,s+1}) \\ h(v_{2,t+1},v_{2,s+1}) - 2C_2(C_1)^2h'(v_{2,t},v_{2,s})\}\{cv_{1,t}+dv_{2,t}\}\big] \\ + 2E\big[\{h'(v_{2,t+1},v_{2,s+1})h(v_{1,t},v_{1,s})h(v_{1,t+1},v_{1,s+1})h(v_{2,t},v_{2,s})\} \\ \{cv_{1,t+1}+dv_{2,t+1}\}\big] \end{aligned} \tag{A25}$$

where $h'(x,y) = \partial h(x,y)/\partial x$. Since $h(\cdot)$ is a symmetric kernel, we need only calculate the partial derivatives with respect to x. Using the independence of $\{v_{i,t+j}\}$, $Eh_1'(v_{i,t+j}) = 0$, and strict stationary condition, we can show that right hand side of (A25) vanishes if $t \neq s \pm 1$. For $t = s \pm 1$, the expectation is nonzero, but the contribution is $o_p(1)$.

If θ_i is the VAR(1) coefficient a_1, we can use the assumption to show the moment condition, (II), holds. We have $Eh_1'(v_{i,t+j}) = 0$ to show the moment condition. $E[\nabla_{\boldsymbol{v}}\mathcal{K}\cdot D_{a_1}G] = 0$ since all $E[\partial h_1(v_{i,t})/\partial v_{i,t}] = 0$ for the indicator function and is independent of $\boldsymbol{y}_{t-k}$ for positive k.

References

Baek, E. G. (1988). Three essays on nonlinearity of economics, Ch. 1. Ph.D. dissertation, Department of Economics, University of Wisconsin–Madison.

Baek, E. G. and Brock, W. A. (1988). A nonparametric test for temporal dependence in a vector of time series. SSRI W.P. #8816, Department of Economics, University of Wisconsin–Madison.

Baumol, W. J. and Benhabib, J. (1989). Chaos: Significance, mechanism, and economic applications. *J. Econom. Perspectives* **3**, 77-105.

Box, G. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, Oakland.

Brock, W. A. (1986). Distinguishing random and deterministic systems: Abridged version. *J. Econom. Theory* **40**, 168-195.

Brock, W. A. and Dechert, W. D. (1988a). Theorems on distinguishing deterministic random systems. *Dynamic Econometric Modeling: Proceedings of the Third International Symposium in Economic Theory & Econometrics* (Edited by W. A. Barnett, E. R. Berndt and H. White), 247–265. Cambridge University Press, New York.

Brock, W. A. and Dechert, W. D. (1988b). A general class of specification tests: The scalar case. *1988 Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 70–79.

Brock, W. A., Dechert, W. D., Scheinkman, J. and LeBaron, B. (1988). A test for independence based on the correlation dimension. Revised version of SSRI W.P. #8702, Department of

Economics, University of Wisconsin–Madison, University of Houston, and University of Chicago.

Brock, W. A. and Sayers, C. (1988). Is the business cycle characterized by deterministic chaos? *J. Monetary Econom.* **21**, 71–90.

Denker, M. and Keller, G. (1983). On *U*-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. verw. Gebiete* **64**, 505–522.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* **50**, 251–276.

Hsieh, D. and LeBaron, B. (1988a). Finite sample properties of the BDS statistic I: Size. *A test for Nonlinear Dynamics and Chaos* (Edited by W. A. Brock, D. Hsieh and B. LeBaron (1991)) forthcoming. MIT Press, Cambridge.

Hsieh, D. and LeBaron, B. (1988b). Finite sample properties of the BDS statistic II: Distribution under alternative hypotheses. *A test for Nonlinear Dynamics and Chaos* (Edited by W. A. Brock, D. Hsieh and B. LeBaron (1991)) forthcoming. MIT Press, Cambridge.

Hsieh, D. (1989). Testing for nonlinear dependence in daily foreign exchange rates. *J. Bus.* **62**, 339–368.

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

White, H. (1984). *Asymptotic Theory for Econometricians*. Academic Press, New York.

Korea Development Institute, P. O. Box 113, Cheong-Ryang, Seoul 130-012, Korea.
Department of Economics, University of Wisconsin, Madison, WI 53706, U.S.A.

(Received December 1989; accepted July 1991)