A Nonparametric Test For Independence Of A Multivariate Time Series

Khung G. Baek
_Iowa State University_

William A. Brock
_Iowa State University_

Follow this and additional works at: http://lib.dr.iastate.edu/econ_las_staffpapers

Part of the _Business Administration, Management, and Operations Commons, Economic History Commons, Economic Theory Commons_, and the _Growth and Development Commons_

Recommended Citation
http://lib.dr.iastate.edu/econ_las_staffpapers/197

This Report is brought to you for free and open access by the Economics at Iowa State University Digital Repository. It has been accepted for inclusion in Economic Staff Paper Series by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
A Nonparametric Test For Independence Of A Multivariate Time Series

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. The null class of the models can be extended to include weak dependence. This test can be used to test the adequacy of a fitted model. As an application of the test we show how we test diagnostically a vector auto regressive model fitted to the given data. This procedure is legitimate because the first order asymptotic distribution of the test statistic is robust to the estimated residual vector.

Disciplines
Business Administration, Management, and Operations | Economic History | Economic Theory | Growth and Development
A NONPARAMETRIC TEST FOR INDEPENDENCE OF
A MULTIVARIATE TIME SERIES

by

Ehung G. Baek
Iowa State University, Ames, IA 50011-1070

William A. Brock
University of Wisconsin, Madison, WI 53706

December, 1989

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. The null class of the models can be extended to include weak dependence. This test can be used to test the adequacy of a fitted model. As an application of the test we show how we test diagnostically a vector autoregressive model fitted to the given data. This procedure is legitimate because the first order asymptotic distribution of the test statistic is robust to the estimated residual vector.

Key Words: 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 are Brock (1986), Brock and Sayers (1988), and Baumol and Benhabib (1989). The present paper finds statistical tests that are capable of determining whether the innovations of a conventional multivariate time series model such as 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. In order to explain we need to set the stage.

A map $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$, generating a difference equation,

$$x_{t+1} = F(x_t), \quad x_0 \text{ is the initial point} \tag{1}$$

will be said to be chaotic if the largest Lyapunov exponent

$$L(x,v) = \lim \left\{ \frac{\ln \left( \|DF^t(x)\cdot v\| \right)}{t} \right\} > 0, \tag{2}$$

where the limit is taken as $t$ goes to infinity, "ln" denotes the natural logarithm, $x$ is an initial condition, $D$ denotes derivative, $v$ in $\mathbb{R}^n$ is a direction vector, $F^t$ is $F$ applied $t$ times. Eckmann and Ruelle (1985) discuss the notion of attractor set $A$, (the smallest future invariant set under $F$) such that on $A$ there is a unique "natural" invariant measure, $\mu$, such that the limit in (2) exists, and is independent of $(x,v)$, for $x$ $\mu$-almost surely and for $v$ Lebesgue
almost surely. An attractor $A$ for $F$ is said to be \textit{chaotic} if $L$ is positive $\mu$-almost surely on $A$.

Chaotic maps can generate time series $\{x_t\}$ by (1) that look random to conventional statistical tests such as the autocovariance function (ACF) or the spectrum.\(^1\) For example Sakai and Tokumaru (1980) consider the family

$$x_{t+1} = \begin{cases} 
\frac{1-x_t}{1-a}, & x_t \in [a,1], \ 0 < a < 1 \\
x_t/a, & x_t \in [0,a]
\end{cases} \quad (3)$$

which generates the same ACF as the linear autoregression,

$$y_{t+1} = (2a - 1)y_t + v_{t+1} \quad (4)$$

where $\{v_t\}$ is Independently and Identically Distributed (IID), with mean zero and finite variance. The special case $a = 0.5$ in (3) is called the \textit{tent map}. One may directly verify that the Lyapunov exponent $L$ has a value of $\ln(2)$. Here the ACF and the spectrum are the same as second order white noise.

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, $\mu$-almost surely, $0.5 = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x_t = \int_{0}^{1} x \mu(dx)$ where $\mu$ 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

\[ \phi(L)y_t = \psi(L)e_t \]  \hspace{1cm} (5)

where \( L \) is the backward operator.

A statistician using Box-Jenkins methods will be hard put to
detect that (5) 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 IID rather than deterministic chaos with a white ACF.

Other types of deterministic chaoses are generated by taking the
"innovations" of nonlinear models such as threshold autoregressive
(Tong (1983)), bilinear (Subba Rao and Gabr (1984)) or Generalized
Autoregressive Conditionally Heteroscedastic (GARCH), (Bollerslev
(1986)) and replacing them with tent chaos.

It is clear one can create a rich variety of deterministic
processes by replacing the innovations or driving uncertainty of
conventional nonlinear stochastic models with tent chaos innovations.
Conventional model based statistical methods will be hard put to
detect the fact that the ultimate source of uncertainty is
deterministic chaos and not a stochastic process.

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

\[ y_t = F(\Omega_{t-1}, e_t; \theta), \]  \hspace{1cm} (6)
where $\Omega_{t-1}$ denotes past information which can include past $y$'s but no past $e$'s, $\theta$ is a vector of parameters to be estimated, and $e_t$ is IID. Moreover $F$ is known and given function. We ask that (6) be solvable for $e_t$, i.e. related to invertibility condition of Granger and Anderson (1978, p. 69) or Subba Rao and Gabr (1983, p. 27). Assume that $\theta$ can be estimated $\hat{\theta}$ consistently. Denote this estimator by $\theta_T$ in sample of size $T$. We assume the function $F$ evaluated at $\theta$ is invertible in $e_t$ in this paper. Then estimated innovations are denoted by $\hat{e}_t$ and satisfy

$$\hat{e}_t = F^{-1}(\hat{\Omega}_{t-1}, y_t; \hat{\theta}_T)$$

where $F^{-1}$ denotes inverse with respect to the $e_t$ argument. We come to the main problem addressed in this paper:

**Problem**: How can one tell from $\{\hat{e}_t\}$ 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 $\{e_t\}$ when the process $\{e_t\}$ is "embedded" in $m$-dimensional space by constructing "m-futures" $e^m_t = (e_t, e_{t+1}, \ldots)$,
One looks at the process \( \{ e_t^m \} \) and measures how well it fills \( m \)-space relative to a comparison IID process that has all the same unconditional moments as the original process. Let \( e_t \) be scalar valued. This comparison is performed by looking at the following measure of spatial correlation, called the correlation integral,

\[
C(m, e, T) = \frac{\# \{ (t, s) \mid 1 \leq t, s \leq T \mid \| e_t^m - e_s^m \| < \epsilon \}}{T^2}
\]

where \( T = T - (m - 1) \) and \( \| \cdot \| \) is the sup norm. (8)

If \( \{ e_t \} \) were IID it was shown in Brock and Dechert (1988a) that

\[
C(m, e, T) \longrightarrow C(1, e)^m \quad \text{a.s. as } T \longrightarrow \infty
\]

where \( C(1, e) = \lim_{T \to \infty} C(1, e, T) \).

This result suggests looking at the following statistic, called W or BDS (Brock-Dechert-Scheinkman, 1987) statistic,

\[
W(m, e, T) = T^{1/2} (C(m, e, T) - C(1, e, T)^m)
\]

which converges in distribution to a normal distribution which has 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 that the form of the nonlinear or linear model that you estimate
is correct. This type of test is highly sensitive to the deterministic chaos and nonlinear dependence between variables. The justification for the test at an heuristic level is that the correlation integral should increase with power $m$ when the data is embedded in $m$ dimensional space if $\{e_t\}$ is truly IID.

Unfortunately Brock, et al. (1988) only treated the scalar case. In economics, we face often multivariate dynamic time series models rather than univariate models, for instance VAR or multivariate ARCH. Therefore we need to develop the vector version of the $W$ test which the current paper aims at. The idea follows.

Let $\{e_t\}$ be a stochastic process of $N$-dimensional vectors. Each $e_t$ is a random vector from some common probability triple $(\mathbb{R}, \mathbb{B}, \mathbb{P})$ to $N$-dimensional space. Let $\{e_t\}$ be IID with identity covariance matrix. Intuitively speaking if you formed $m$-futures $e_t^m$ they should fill $Nm$-dimensional space for each embedding dimension, $m$. A similar comparison as was done in the original development of the $W$ statistic should work in our case. Unfortunately there are several rather tedious technical problems that must be dealt with in implementing this strategy. One of the most vexing is to find a tractable version of (6) in the multidimensional case. Nevertheless it is possible as the current paper shows.

2. A Test for Independence of a Vector of Time Series
The BDS type tests may be extended to vectors of time series in the following way. Let \{u_{i,t}\}, i=1,2,...,N; t=1,2,... be strictly stationary. Let the null hypothesis be \(H_0: \{u_{i,t}\} \) is independent across all \(i, t; i=1,...,N; t=1,2,...\).

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), we design the notation to suggest this. Ignoring endpoint problems to save notation, we define following concepts and notations.

(i) \(h(u,v): \mathbb{R}^2 \rightarrow \mathbb{R}\) is a symmetric kernel and a twice continuously differentiable \(C^2\) function where \(u,v \in \mathbb{R}\). Define a projection of a kernel as \(h(u) = \mathbb{E}[h(u,v)|u]\). Generally a symmetric kernel and a projection of it in higher dimensional space is defined in a similar way.

(ii) \(m_i\)-futures of \(i\)th element of an \(N\)-vector \(u\) is \(u_{i,t} = (u_{i,t},u_{i,t+1},...,u_{i,t+m_i-1})\), and the collection of these futures for all \(i\) at time \(t\) is \(u_t = (u_{1,t},u_{2,t},...,u_{N,t})\).

(iii) The time \(t\) past of an \(N\)-vector stochastic process \(\{y_s\}_{s=0}^\infty\) is written as \(y_t = (y_t,y_{t-1},...)\); \(\{y_t\}\) are generated by, \(y_t = \)
F(Y_{t-1}',\theta_1) + e_t where $\theta_1$ is the $N \times k_1$ coefficient parameter matrix of the data generating function (DGF) $F$. Moreover $E[e_t|Y_{t-1}] = 0$ and $E[e_t e_t'|Y_{t-1}] = H_t$ where $H_t = H(Y_{t-1}, \theta_2)$ and $\theta_2$ is a parameter matrix in the covariance matrix $H_t$. If $u_t = H_t^{-1/2} e_t$, then $E[u_t|Y_{t-1}] = 0$ and $E[u_t u_t'|Y_{t-1}] = I$ where $I$ is an $N \times N$ identity matrix. Here $H_t^{-1/2}$ denotes the inverse of the positive (semi) definite symmetric matrix $H^{1/2}$ whose elements are continuous functions of $H_t$. (White, 1984, p. 65). For later use, write $u_t$ as $G(Y_t, \theta) = H_t^{-1/2} (y_t - F(Y_{t-1}, \theta_1))$ where $\theta = (\text{vec}(\theta_1, \theta_2))$. Since we use the column stacking operation vec(·), $\theta$ is a vector. We assume that $F \in c^2$ and $\theta$ is a finite dimensional vector of parameters to be estimated.

(iv) $u_t = G(Y_t, \theta)$ is the standardized actual innovation vector and $u_T = G(Y_t, \theta_T)$ is the standardized estimated innovation vector from solving the data generating process for $e_t$. The DGF $F$ is assumed invertible in the $e$ argument with $c^2$ class inverse function $G$.

(v) The correlation integral is defined like (8) in a vector case. That is

$$C(m, e, T) = (1/T^2) \sum_{t=1}^{T} \sum_{s=1}^{T} \zeta(u_t^m, u_s^m; e)$$

(11)

where $\zeta$ is a symmetric kernel defined on $\mathbb{R}^i \times \mathbb{R}^i$. Since kernels $\zeta, h$ are the indicator functions in this paper, (11) becomes
The indicator function \( \mathbb{1} \) is taken to be
\[
\mathbb{1}(u_t^m, u_s^m; \varepsilon) = \mathcal{N}(\| h(u_t^m, u_s^m, \varepsilon) \|),
\]
where \( \| \cdot \| \) is the sup norm and
\[
\mathbb{1}(x) = \begin{cases} 
0 & \text{if } x \leq 0 \\
1 & \text{otherwise.}
\end{cases}
\]
Furthermore
\[
C(i, \varepsilon, T) = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} h(u_t^i, u_s^i, \varepsilon),
\]
\[
C_i = \mathbb{E}[h(u_t^i, u_s^i, \varepsilon)],
\]
\[
C_m = \mathbb{E}\left\{ \mathbb{E}[h(u_t^i, u_s^i, \varepsilon)] \right\} = \mathbb{E}\left\{ \sum_{i=1}^{N} (C_i)^m \right\}.
\]
For simplicity \( C(m, \varepsilon, T), C(i, \varepsilon, T) \) and \( h(a, b; \varepsilon) \) will be written as \( C(m), C(i) \) and \( h(a, b) \) for fixed \( \varepsilon, i, \) and a given sample of size \( T \) without any confusion.

(vi) \( \Theta_{1T} \) and \( \Theta_{2T} \) are consistent estimators of \( \Theta_1 \) and \( \Theta_2 \) respectively. Let \( \Theta_T = \text{vec}(\Theta_{1T}, \Theta_{2T}) \).

(vii) \( W(C(m), C(i)) = C(m) - \frac{1}{T} \sum_{i=1}^{N} C(i)^m \).
"
(viii) Let $x_t = u_t^m$, and $x_s = u_s^m$ where $u_t^m$ is defined in (ii). Define a kernel
\[
K(x_t, x_s) = \sum_{i=1}^{N} \sum_{j=0}^{m_i-1} h(u_{i,t+j}, u_{i,s+j}) - C_m^m
\]
\[
- \sum_{i=1}^{N} \sum_{j=0}^{m_i-1} \sum_{j' \neq i} m_i \cdot m_j \cdot (h(u_{i,t}, u_{i,s}) - C_i) \] (17)

(ix) $v_u K$, and $v_\theta G$ are gradient vectors of the kernel $K$ and the vector function $G$.

(x) The kernel $K$ is written as a function $J$ of all observations and parameter vectors based on the inverse function
\[
J(Y_{t+m+1}, Y_{s+m+1}; \theta)
\]
\[
= K[G(Y_{t+m+1}; \theta), G(Y_{s+m+1}; \theta)]
\]
is the kernel $K$ with all $u_{i,t+j}$ and $u_{i,s+j}$ where $m = \max(m_1, \ldots, m_N)$.

(xi) Let \( \hat{K}(u_1, u_j, u_k) = (1/3!) \sum_{\sigma} h(u_{i_1}, u_{i_2}) h(u_{j_1}, u_{j_2}) \) and the summation is over all possible permutations $\sigma$ of indices $(i, j, k)$.

(xii) Let $H(Y_r, Y_s, Y_t; \Theta, \phi) = K(G_i(Y_r; \Theta), G_i(Y_s; \Theta), G_i(Y_t; \Theta))$ and $L(Y_s, Y_t; \Theta, \phi) = h(G_i(Y_s; \Theta), G_i(Y_t; \Theta))$ where $G_i(Y, \Theta)$ is the $i$th innovation. $\partial H/\partial Y_t$, $\partial^2 L/\partial Y_t^2$, and $\partial^2 J/\partial Y_t^2$ will be abbreviated as $H'$, $L''$, and $J''$ to clarify kernel functions.
We develop the test method and properties throughout the next five theorems which are based on the following assumptions.

(A1) (Uniform Mixing Condition) The DGF generates a stochastic vector process \( \{y_t\} \) that satisfies one of the uniform mixing assumptions of Denker and Keller (1983, Theorem 1, p. 507).

\[
(A2) \text{(Moment Condition)} \quad \mu = E[\nabla \log \varphi_i G] = 0 \text{ for all } i. 
\]

\[
(A3) \text{(Asymptotic Normality)} \quad T^{1/2}(\theta_T - \theta) \xrightarrow{d} N(0, V). 
\]

(A4) (Compactness of Parameter Space) There is a compact set \( \Omega \) such that the range of \( \theta_T \) is contained in \( \Omega \) for all \( T \).

(A5) (Bounded Moment Condition) All kernels are non-degenerative, 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" is a notation for supremum of the expectation over all permutations of temporal arguments.

(A6) (Smoothness) All kernels appearing below are at least \( C^2 \). Even if the nondifferentiable indicator function is used as a kernel for practice, we show that a \( C^2 \) function always exists which approximates
it in Theorem 5.

(A7) (Continuity) For $X$ equal to $H'$, $L''$, $J''$ in (x) and (xii), \( \sup|X(\cdot, \cdot, z)| \) is continuous in $z$. Recall the sup is taken over all temporal permutations as in Denker and Keller (1983, p. 507).

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 $i$, and A5 holds for the kernels in the $C(m)$, and $W$ statistics and in the $K$ statistic in (24) below. Then

$$ T^{1/2}[W(C(m), C(i))] \xrightarrow{d} N(0, V_m) \text{ as } T \to \infty. \quad (20) $$

And

$$ V_m = 4 \left[ \prod_{i=1}^{N} \left( \frac{2m_i}{m^*} \right)^{-1} \right] + \sum_{i=1}^{N} \prod_{p=2}^{m_i-p+1} \left( (C_i - (C_i)) \right) (C_i) - \prod_{i=1}^{m_i-p+1} \left( (C_i) \right)^2 \sum_{p=2}^{m_i-p+1} \left( \prod_{j=1}^{N} \left( C_j \right) \right)^{(p-1)} \right] (21) $$

where $m^* = \max(m_1, m_2, \ldots, m_N)$,

$$ I(K_i) = \begin{cases} (K_i) & \text{if } m_i-p+1 > 0 \\ 1 & \text{otherwise} \end{cases} $$
\[ J(K_1) = \begin{cases} K_i & \text{if } m_i - p + 1 > 0 \\ (C_i)^2 & \text{otherwise.} \end{cases} \]

Moreover \( C_i = \text{E}h(u_i, t+j, u_i, s+j) \) for \( j = 0, \ldots, m_i \) and
\[ K_i = \text{E}[h(u_i, r+j, u_i, s+j)h(u_i, s+j, u_i, t+j)] \] for \( j = 0, \ldots, m_i \).

Proof: See Appendix.

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

\[ C(i) = \frac{1}{T^2} \sum_{t,s} h(u_i, t, u_i, s) \] for \( C_i \) \hspace{1cm} (22)
\[ K(i) = \frac{1}{T^3} \sum_{t,s,r} [h(u_i, t, u_i, s, r)h(u_i, s, u_i, r)] \] \hspace{1cm} (23)
\[ \tilde{K}(i) = \frac{1}{T^3} \sum_{t,s,r} \tilde{k}(u_i, t, u_i, s, u_i, r) \] for \( K_i \) where \( \tilde{k} \) \hspace{1cm} (24)
\[ \tilde{k}(u_i, t, u_i, s, u_i, r) = \frac{1}{3!} \sum_{s'} h(u_i, t, u_i, s')h(u_i, s, u_i, r). \] \hspace{1cm} (25)

A consistent estimator of \( V_m \) is obtained by replacing \( K_i \) with \( \tilde{K}(i) \) and \( C_i \) with \( C(i) \). The nonsymmetric kernel in (23) can be symmetrized without loss of generality via (25). Assumption A5 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
\[ T^{1/2}(V_m)^{-1/2} \text{W} \] for \( V_m = \tilde{V}_m \) equal to any consistent estimator of \( V_m \) we
have a statistic that converges in distribution to $N(0, I)$ asymptotically under the null hypothesis. Here $I$ denotes the identity matrix. (Warning: Superscript * in $Y$'s refers to superscript on $m$ not on $Y$.)

For practical use, under $A2$, we show that the first order asymptotics of $W$ statistic in (20) are the same for $\{\mathbf{u}_{i,t}^*\}$ as for $\{\mathbf{u}_{i,t}\}$ in the next theorem. Superscript * denotes estimated values. Theorem 6 below gives practical conditions sufficient for $A2$ to hold.

Theorem 2: Assume the same conditions of Theorem 1. Additionally assume $A1$, $A2$ (on $J'$), $A3$, and $A5$ on the kernels $J'(Y_{s+m-1}, Y_{s+m-1}; \Theta)$, $L'(Y_{t}, Y_{s}; \Theta)$ and $A4$ and $A7$ with $X$ equal to $J''$. Then

$$ T^{1/2} [W(C^*(m), C^*(i)) - W(C(m), C(i))] \xrightarrow{D} 0 \text{ as } T \to \infty. \quad (26) $$

Proof: Write an exact Taylor expansion for each $W$ by using notation $\mathbf{K}$ in (viii).

$$ W(C^*(m), C^*(i)) = (1/T^2) \sum_{t} \sum_{s} [\mathbf{K}^* \mathbf{m}^i \mathbf{m}_{i-1} (C^*_i)^{m_{i-2}}]

\sum_{j \neq i} \mathbf{m}^j (C^*(i) - C^*_i)^2/2], \quad (27) $$

$$ W(C(m), C(i)) = (1/T^2) \sum_{t} \sum_{s} [\mathbf{K} \mathbf{m}^i \mathbf{m}_{i-1} (C_i)^{m_{i-2}}]

\sum_{j \neq i} \mathbf{m}^j (C(i) - C_i)^2/2] \quad (28) $$

where $C^*_i$, $C_i$ 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 (A5) on \( L' \) will be used. We must show

\[
A = T^{1/2} \left[ \sum_{t,s} \left( \frac{X^* - X}{T} \right)^2 \right] \xrightarrow{P} 0 \text{ as } T \to \infty.
\]

(29)

Insert the formula (30)

\[
u_t = G(Y_t, \theta), \quad u^{T} = G(Y_t, \theta)
\]

into (29), expand in a Taylor series about \( \theta \) with exact second order remainder to obtain (31) below. Noting \( J' = \sum_{u,i} X \cdot v_{u,i} \), we have

\[
T^{1/2} \left[ \sum_{t,s} \left( \frac{G(Y_t, \theta_{T}) + \sum_{i} G(Y_t, \theta_{T})}{T^2} \right) - \sum_{t,s} \left( \frac{G(Y_t, \theta_{T})}{T^2} \right) \right]
\]

\[
= T^{1/2} \left[ \sum_{t,s} \left( J(Y_{t,m-1}, Y_{s,m-1}; \theta_{T}) - J(Y_{t,m-1}, Y_{s,m-1}; \theta) \right) \right] + \frac{1}{2} \sum_{i} J''(Y_{t,m-1}, Y_{s,m-1}; \theta)(\theta_{T} - \theta_{i})^2 / T^2.
\]

(31)

Note \( \sum_{t,s} \frac{E J'/T^2}{\mu} \xrightarrow{P} 0 \) where by A2, \( \mu = E J' = E \left( \sum_{u,i} X \cdot v_{u,i} \right) = 0 \).

(32)

Therefore it is sufficient to show

\[
\sum_{t,s} \sum_{u,i} \left( \frac{X^* - X}{T} \right)^2 \xrightarrow{P} 0 \text{ and the second order terms in (31)}
\]

converge in probability to zero. Observe \( \sum_{u,i} X \cdot v_{u,i} = J' \) is the derivative of a symmetric kernel \( J(\cdot, \cdot; \theta) \) with respect to a vector \( \theta \),
hence $\sum J'/T^2$ is a $V$-statistic. Under the mixing condition A1 and the nondegeneracy, and bounded second moment condition A5 on the kernel $J'$, Denker and Keller Theorem 1 asserts

$$\sum \sum J'/T^2 \xrightarrow{P} EJ' \text{ as } T \to \infty.$$  \hspace{1cm} (33)

But $EJ' = \mu = 0$ by A2. The second order terms in (31) converge in probability to zero. Since each component of $\Theta_T$ converges in distribution to a random vector it is enough to show for each element $i, j$

$$\sum \sum J''(\cdot, \cdot; \Theta_T)/T^{5/2} \xrightarrow{P} 0 \text{ as } T \to \infty.$$  \hspace{1cm} (34)

where $\Theta_T$ is in between $\Theta$ and $\Theta_T$. To show (34) 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 $i, j$

$$\sup \{E|J''(\cdot, \cdot; \Theta)| \text{ all nonnegative t and s} \} < B$$  \hspace{1cm} (35)

where the sup is taken over $\Theta$ in some compact set $K$. But (31) follows from A4 which states that the values of $\Theta_T$ lie in a compact set $K$ which is independent of $T$, and A6 which states

$$\sup \{E|J''(\cdot, \cdot; \Theta)|, \text{ all nonnegative t and s} \} \text{ is continuous in } \Theta.$$ 

Q.E.D.

To complete the proof of Theorem 2 the second order terms of (27), (28) are disposed of in the following Lemma.

Lemma 3: Consider the second order terms $M_1, M_2$ in (27), and (28).
Assume the kernel $h(u,v)$ is bounded between 0 and $B<\infty$, and assume $A4$, $A5$, $A6$, and $A7$. Then $T^{1/2}M_i \xrightarrow{p} 0$ as $T \to \infty$ for $i=1$ and 2.

Proof:

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

$$T^{1/2}(C^*(i) - C_i) = O_p(1), \quad (36)$$

$$T^{1/2}(C(i) - C_i) = O_p(1) \quad (37)$$

Convergence of the second term (37) follows from the same type of argument is that used in Theorem 1. The first term (36) requires attention. Taylor series expansion gives

$$C^*(i) = \sum \sum h(u_{i,t,u_{i,s}})/T^2 = \sum \sum h(G_i(Y_t,\Theta_T),G_i(Y_s,\Theta_T))/T^2$$

$$= \sum \sum h(Y_{t,s};\Theta_T,i)/T^2$$

$$= C(i) + \sum \sum L'(Y_{t,s};\Theta_T,i)(\Theta_T - \Theta)/T^2 +$$

$$\quad \left(\frac{1}{2}\right) \sum \sum (\Theta_T - \Theta)''(Y_{t,s};\Theta_T,i)(\Theta_T - \Theta)/T^2$$

$$= C(i) + S + U \quad (38)$$

where $S$ is the second term and $U$ is the third term of (38).

Here $\Theta_T$ is in between $\Theta$ and $\Theta_T$ and $T^{1/2}(\Theta_T - \Theta) \xrightarrow{d} N(0,V_\Theta)$ as $T \to \infty$. So $T^{1/2}S = O_p(1)$. Since $A5$ allows application of Denker and Keller Theorem 1 to $\sum \sum L'/T^2$ so that $\sum \sum L'/T^2 \xrightarrow{p} EL'$. Furthermore $T^{1/2}U \xrightarrow{p} 0$ by using $A4$, $A6$ and $A7$ with $X$ equal to $L''$. Therefore
\[ T^{1/2}(C^*(i) - C_1) = O_p(1) \text{ and } T^{1/2}(C(i) - C_1) = O_p(1). \] It follows that \( M_1 \) and \( M_2 \) converge to zero in probability, since

\[ T^{1/2}\left[ \frac{1}{T^2} \sum_{m_i} \sum_{j \neq i} (m_i - 1) C_i \sum_j (C_j^*(i) - C_i)^2/2 \right] \]

\[ = O_p(1)/T^{1/2} \rightarrow 0 \text{ as } T \rightarrow \infty. \text{ Q.E.D.} \]

While Theorem 2 shows that, under A2, the first order asymptotic distribution of \( W \) is invariant to evaluation at \( u_{T_1} \) or \( u_{T_2} \) we still need a practical estimator of the variance \( V_m \) for the statistic to be of practical use.

Theorem 4: The estimator of the variance of Theorem 1 evaluated at all estimated innovations \( u_{i,t}^* \), call it \( V_m^* \), is consistent for \( V_m \).

Proof:

Noting \( V_m = V(C_i,K_i) \), we must show

\[ V_m^* \xrightarrow{D} V_m \text{ as } T \rightarrow \infty \text{ where } V_m^* = V(C^*(i),K^*(i)). \] (40)

The estimate \( K_m^* \) can be written

\[ K_m^*(i) = \sum \sum H(Y_r,Y_t,Y_s;\theta_T,i)/T^3. \] (41)

Expand (41) in an exact first order Taylor series about \( \theta \). One gets \( K(i) \) plus first order terms in \( (\theta_T - \theta) \). \( K(i) \) is a consistent estimator of \( K_i \) almost surely. We have \( T^{1/2}(\theta_T - \theta) = O_p(1) \) as \( T \rightarrow \infty \). By A2 there is a bound \( B<\infty \) such that
sup E{|H'(Y_r,Y_s,Y_t;z,i)|, all nonnegative r, s, t } < B < ∞. (42)

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 $K^*(i) \overset{P}{\rightarrow} K_i$ and $C^*(i) \overset{P}{\rightarrow} C_i$.

Q.E.D.

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 $V_m$ for the variance is continuous in its arguments $C_i$, and $K_i$. The next theorem provides the approximation of the nonsmooth kernel with smooth kernel.

Theorem 5: For every $q>0$ we can find a kernel $h(u,v)$ that satisfies A5, A6, and A7 such that the absolute value of the difference between the variance computed at the indicator kernel and the variance computed at the smooth kernel $h$ is less than $q$.

Proof: See Brock and Dechert (1988b, p. 74)

There are some related work showing small sample properties. Baek (1988) studied the special case $N=2$, $m_1=1$, $m_2=1$ of the W test. In Baek's case the kernel $h(u_{it},u_{is})$ is the indicator function of the
event \(|u_{it} - u_{is}| < \varepsilon\). He did preliminary Monte Carlos on size and power against several contemporaneously dependent alternatives. 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) 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) \(v_t\) is a piecewise linear transformation of \(u_t\); (ii) \(v_t\) is a sine function of \(u_t\); (iii) contemporaneous version of ARCH, 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 a lot of wiggles which confuse the other tests into thinking that the two series are independent. See the Table 1.2 in Baek (1988) for contemporaneous independence.

The BDS test is also a special case, \(N=1\), of the above test. It is a test of IID for a univariate series. Encouraging Monte Carlo results on performance are reported in Hsieh (1989, Table 8). Since the performance of the BDS test was quite good we hazard the guess that the vector version of BDS propounded here will exhibit similar good performance.

3. An application of the multivariate test of independence to VAR
We show here how the multivariate test of independence is
practically applied. There has been much interest in autoregressive
models. To operationalize our test we need to show a nice property.
That is to say if you estimate the correct null autoregressive model
to your vector of time series then the W statistic in (20) evaluated
at the estimated standardized residuals has the same asymptotic
distribution as the W statistic evaluated at the true
standardized innovations. This is true provided your estimation
procedure is \( \mathcal{A} \) consistent.

We will make assumptions to show the above property. The true
covariance matrix is \( \Gamma \) and the estimated covariance matrix is \( \Gamma^* \).
Suppose
\[
y_t = A y_{t-1} + u_t,
\]
with
\[
E[u_t \mid y_{t-s}, \text{ all } s > 0 ] = 0 \text{ and } E[u_t u_t' \mid y_{t-s}, \text{ all } s > 0 ] = \Gamma,
\]
and
\[
\{ u_t \} \text{ is a stationary } N \times 1 \text{ vector IID stochastic process with finite fourth moments.}
\]

If (41) is the true data generating process and we use the right
model to fit given data, the standardized estimated VAR residuals are
asymptotically IID across time and independent across variables. The
standardized VAR residuals are given by
\[
v_t^* = (\Gamma^*)^{-1/2} (I - A^*(L))y_t
\]
where \( L \) is the lag operator and \( A^* \) is estimate of \( A \). Since \( \{ v_t^* \} \) is a
sequence of \( N \)-vectors we can use them as arguments for the \( W \) test.
For illustration we show that the above property holds in the 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 \( \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \sim N(0_{2 \times 1}, \Sigma) \) and \( \Sigma = \begin{bmatrix} \sigma_1 & \sigma \sigma_3 \\ \sigma_3 & \sigma_2 \end{bmatrix} \) and positive definite.

Let \( \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} = (\Sigma)^{-1/2} (y_t - A(L)y^*) = G(Y_t, B) \) where \( B = \text{vec}(A, \Sigma) \) and

\[ v_{1,t}^* = (\Sigma^*)^{-1/2} (y_t - A^*(L)y^*) = G(Y_t, B^*) \] where \( B^* = \text{vec}(A^*, \Sigma^*) \).

Now we are ready to establish the following theorem.

**Theorem 6:** If (41), (42), and (43) are true, and \( \Sigma \) is symmetric and positive definite then the independence test \( W \) has the same first order asymptotic distribution whether it is evaluated at \( \{v_{1,t}^*\} \) or \( \{v_t\} \).

**Proof:** If we replace all \( u_{i,t} \) variables with \( v_{i,t} \) in the proof of Theorem 2, it is sufficient to show the moment condition, A2. The idea of the proof is on the same line as the previous proofs. We will represent \( v_{t}^* \) by \( Y_t \) including current \( y \) and past values of \( y \) and \( \Sigma^* \), and \( A^* \). The next step is to conduct a Taylor expansion of the \( W \)
statistic about a true parameter. The first order terms will
disappear by the moment condition 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 holds we use (31), and (32).
Here since \( \mathbf{x}(v_t^2, v_s^2) = h(v_{1,t}, v_{1,s})h(v_{1,t+1}, v_{1,s+1})h(v_{2,t}, v_{2,s}) \)
\[
\begin{align*}
    h(v_{2,t+1}, v_{2,s+1}) - C_2 - 2C_2(C_2)^2(h(v_{1,t}, v_{1,s}) - C_1) \\
    - 2C_2(C_1)^2(h(v_{2,t}, v_{2,s}) - C_2) \text{ and}
\end{align*}
\]
\[
\begin{align*}
    v(v_t^2, v_s^2) = [\partial x/\partial v_{1,t}, \partial x/\partial v_{1,t+1}, \partial x/\partial v_{2,t}, \partial x/\partial v_{2,t+1}, \\
    \partial x/\partial v_{1,s}, \partial x/\partial v_{1,s+1}, \partial x/\partial v_{2,s}, \partial x/\partial v_{2,s+1}].
\end{align*}
\]
Now we must show that \( E[ (v_t^2, v_s^2) B_i G] = 0 \) for all elements of the
parameter vector where \( G = G(Y_t, B) \) to express the \( v_t \) vector. We will
show the moment condition when \( B_i = \sigma_1 \) or \( a_1 \). Other cases will be
shown in a similar way.
Define \( r^{-1/2} = [\eta_{ij}] \) for \( i, j = 1 \) or \( 2 \). Then \( \eta_{ij} = \eta_{ij}(\sigma_1, \sigma_2, \sigma_3) \). Thus \( v_t = [v_{1,t}] = r^{-1/2}u_t = [\eta_{ij}]u_t \). However
\[
\begin{align*}
    u_t = r^{-1/2}v_t \text{ implies } \partial v_{1,t}/\partial \sigma_1, \text{ and } \partial v_{2,t}/\partial \sigma_1 \text{ are both linear functions}
    \text{ of } v_{1,t}, \text{ and } v_{2,t}. \text{ Let } \partial v_{1,t}/\partial \sigma_1 = av_{1,t} + bv_{2,t} \text{ and }
    \partial v_{2,t}/\partial \sigma_1 = cv_{1,t} + dv_{2,t}. \text{ Therefore}
\end{align*}
\]
\[
\begin{align*}
    v_{\sigma_1} G = \left[ a v_{1,t} + b v_{2,t}, a v_{1,t+1} + b v_{2,t+1}, c v_{1,t} + d v_{2,t}, c v_{1,t+1} + d v_{2,t+1}, \\
    a v_{1,s} + b v_{2,s}, a v_{1,s+1} + b v_{2,s+1}, c v_{1,s} + d v_{2,s}, c v_{1,s+1} + d v_{2,s+1} \right] \text{,}
\end{align*}
\]
where $a, b, c, \text{ and } d$ are all functions of $\eta_{ij}$ and $\partial \eta_{ij}/\partial \sigma_1$ for $i, j = 1$ or 2.

If we expand a Taylor series about $\sigma_1$,

$$
E[v^M \cdot v^M_G] = 2E\{h'(v_1, t, v_1, s)h(v_1, t+1, v_1, s+1)h(v_2, t, v_2, s)h(v_2, t+1, v_2, s+1)\}
$$

$$
h(v_2, t+1, v_2, s+1) = 2C_1(C_2)h'(v_1, t, v_1, s)\{av_1, t+bv_2, t\} +
$$

$$
2E\{h'(v_1, t+1, v_1, s+1)h(v_1, t, v_1, s)h(v_2, t, v_2, s)h(v_2, t+1, v_2, s+1)\}
$$

$$
\{av_1, t+1+bv_2, t+1\} + 2E\{h'(v_2, t+1, v_2, s+1)h(v_2, t, v_2, s)h(v_1, t+1, v_1, s+1)\}
$$

$$
h(v_1, t+1, v_1, s+1) = 2C_2(C_1)h'(v_2, t, v_2, s)\{cv_1, t+dv_2, t\} +
$$

$$
2E\{h'(v_2, t+1, v_2, s+1)h(v_1, t, v_1, s)h(v_1, t+1, v_1, s+1)h(v_2, t, v_2, s)\}
$$

$$
\{cv_1, t+1+dv_2, t+1\}. \quad (45)
$$

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 $t$. Using the independence of $\{v_i, t+j\}$, $Eh'(v_i, t+j) = 0$, and strict stationary condition, we can show that right hand side of (45) vanishes. \footnote{This is so because}

$$
Eh(v_1, t+j, v_1, s+j) = C_1, \quad Eh(v_2, t+j, v_2, s+j) = C_2 \text{ for } j=0 \text{ and } 1,
$$

$$
Eh'(v_1, t, v_1, s)v_1, t = Eh'(v_1, t+1, v_1, s+1)v_1, t+1, \text{ and}
$$

$$
Eh'(v_2, t, v_2, s)v_2, t = Eh'(v_2, t+1, v_2, s+1)v_2, t+1.
$$

If $B_1$ is the VAR(1) coefficient $a_1$, we can use the fact

$$
Eh'(v_i, t+j) = 0 \text{ to show the moment condition. } E[v^M \cdot v^M_G] = 0 \text{ since all } E[\partial h_1(v_i, t)/\partial v_i, t] = 0 \text{ for the indicator function and is independent of } y_{t-k} \text{ for positive } k. \quad Q.E.D.
4. Summary and Suggestions for Future Research

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) and Baek (1988) indicate that the vector generalization of BDS proposed here should have good power properties against broad classes of alternatives. Hence we believe that our diagnostic test has promise of being machined into a diagnostic test of wide usefulness to applied econometricians.

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 $T$, $\epsilon$, $m$. We have not done this.

Second, we have not developed a theory of the optimal choice of $\epsilon$, $m$ 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, our test can be used to do a crude version of nonlinear Granger causality testing. If one series helps forecast another then they can't be independent. Hence rejection of the null hypothesis of independence for two series is consistent with dependence but not necessarily existence of a uni-directional Granger ordering.

For example, suppose \( \{X_t\} \) and \( \{Y_t\} \) are stationary processes whose autocorrelations and cross-correlations are zero. Our \( H_0 \) is that \( Y \) does not Granger cause \( X \). Letting \( p \) denote probability our null hypothesis is

\[
H_0: \{p[|x_{t+1} - x_{s+1}|<\epsilon | |X_t - X_s|<\epsilon, |Y_t - Y_s|<\epsilon]\}
\]

\[
= p[|x_{t+1} - x_{s+1}|<\epsilon | |X_t - X_s|<\epsilon]].
\]

In macroeconomics the null hypothesis is "\( Y \) does not Granger cause \( X \)". Under the null hypothesis we build a test statistic

\[
T^{1/2}[\hat{p}[|x_{t+1} - x_{s+1}|<\epsilon | |X_t - X_s|<\epsilon, |Y_t - Y_s|<\epsilon] - p[|x_{t+1} - x_{s+1}|<\epsilon | |X_t - X_s|<\epsilon)] \overset{d}{\rightarrow} N(0,V)
\]

where \( V \) is the finite variance and \( \hat{p} \) is the probability measure in a given sample. For practical use, we can replace \( p \) with the correlation integral \( C \) in the sample (as in Brock and Dechert (1988a, Lemma 2.1, p. 257).

The test statistic is

\[
T^{1/2}[C(x_{t+1}^t Y_t^t)/C(X_t Y_t) - C(x_{t+1}^t X_t)/C(X_t^t)] \overset{d}{\rightarrow} N(0,V)
\]

where \( X_t = (x_t, x_{t-1}, \ldots, x_{t-L}) \) and \( Y_t = (y_t, y_{t-1}, \ldots, y_{t-M}) \) for certain lags \( L \) and \( M \). Here \( C(z) = \sum_{t} \sum_{s} \sum_{i} \sum_{h} h(z, i, t, h, s) \) for \( z = t s i h \).
Even if this idea is natural from the test developed in this paper an adaptation of our test for nonlinear Granger causality testing remains for future research.

Fourth, 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, A2, holds for any vector of kernels such that \( E h_1' = 0 \). This includes the indicator kernel used in Baek (1988) and Brock, et al. (1988). In these papers the kernels were indicator functions of events like \( \{|u_{i,t} - u_{i,s}| < \epsilon \} \) for some choice of \( \epsilon \).

Strictly speaking these indicator function kernels do not satisfy the smoothness condition needed for the theorems. However this problem is easily fixed by approximation of the indicator function to a smooth function. Here is the intuition. The limit distribution is normal and, thus, only depends upon the mean and variance. The mean is zero. Turn to the variance. The variance of the asymptotic distribution is a mean square continuous function of the kernel. Therefore a mean square approximation to the original kernel can be chosen to obtain a variance as close to the original variance as you like. In view of this approximation result we will act as if the theorems apply directly to indicator function kernels. We do this for expositional simplicity. Turn now to the question of how to choose a vector of kernels \( h(\cdot, \cdot) \).

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

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

Proof of Theorem 1

This follows from use of the delta method (Serfling, 1980, p. 118) and Denker and Keller (1983, Theorem 1). Denker and Keller's uniform mixing conditions are trivially satisfied for the stochastic process \( \{u_{i,t}\} \) 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.

\[
C(m) - C_m = \frac{2}{T} \sum_{t} \left[ \sum_{i} \sum_{j} h_1(u_{i,t+j}) - C_m \right] + R_1, \quad (A1)
\]

\[
C(i) - C_i = \frac{2}{T} \sum_{t} \left[ h_1(u_{i,t}) - C_i \right] + R_2 \quad (A2)
\]

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 \( C(i) \), \( i=1, \ldots, N \). I.e.

\[
W(C(m), C(i)) = g(C(m), C(1), \ldots, C(N)) = C(m) - \prod_{i}^{m} C(i). \]

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:

\[
W = \frac{2}{T} \sum_{t} \left[ \sum_{i,j} h_1(u_{i,t+j}) - C_m \right] - \sum_{i} \Phi_i \left[ h_1(u_{i,t}) - C_i \right] + R
\]

where \( \Phi_i = \prod_{j \neq i}^{m_i-1} \prod_{j}^{m_j} \). We know that \( T^{1/2} R \to 0 \) in probability as \( T \) goes to infinity. Write the function \( g \) evaluated at all arguments \( u_{i,t}^m \) which is equal to
Then \( W = \frac{1}{T} \sum \sigma(u^m_t) + R \). Therefore

\[ T^{1/2} W \xrightarrow{d} N(0, V^*_m) \text{ where } V^*_m = \text{Var} \left[ \sum \sigma(u^m_t) \right] + 2 \sum P=2 \sigma(u^m_L) \sigma(u^m_R) \]

and \( m^* = \max(m_1, \ldots, m_N) \).

Now we show that all the second order terms go to zero in probability.

\[ W(C(m), C(i)) = g(C(m), C(1), \ldots, C(N)) = C(m) - \prod_{i} C(i)^{m_i}. \]

\[ W = [C(m) - C_m] - \sum_{m_i}^{m} (C_i) - \sum_{j \neq i}^{m_j} (C(i) - C_i)^{m_j} \]

\[ m_i(m_i - 1)(C_i)^{m_i - 2} \prod_{j \neq i} (C_j)^{m_j} (C(i) - C_i)^{2/2} \text{ where } C_i \text{ is an intermediate point between } C(i) \text{ and } C_i. \]

Then the second order terms are

\[ T^{1/2} \left( \sum \sigma(m_i - 1)(C_i)^{m_i - 2} (C_j) \prod_{j \neq i} (C(i) - C_i)^{2/2} + \right) \]

\[ R_{C_i}^2/2/T^2. \]

Based on the Lemma 3 the second order terms go to zero in probability.

The above proof enables us to derive the variance formula in case of the indicator kernel function. The variance formula of (21) is straightforward from (A4).
Endnotes

1. The ACF function is defined by using time average and ergodic property in the deterministic process. For example, suppose $e_t = x_{t-1/2}$ and $x_t$ is generated from a tent map, $F$, like (3) when $a = 1/2$. Then the estimate of the 1st lag autocorrelation is defined in the following way:

$$
\frac{1}{T} \sum_{t=2}^{T} e_t e_{t-1} = \frac{1}{T} \sum_{t=2}^{T} (x_{t-1/2})(x_{t-1/2})
$$

where $\mu$ is the invariant measure so that $\mu(dx) = dx$. Then the right hand side is 0. One can show that $\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T-k} e_t e_{t-k} = 0$ for any positive integer $k$. See Sakai and Tokumaru (1980) for details.

2. The test (10) is a family of tests $W(m, \varepsilon, T)$ that depend upon the parameters $\varepsilon, m$. The parameter $m$ reflects the number of lags for which you are testing for dependence. For example if $W$ is small for $m = 2, 3$ and jumps at $m = 4, 5, \ldots$ then you can be quite sure that there is little dependence at 2 and 3 lags but a "lot" of dependence at 4 lags and more.

The choice parameter $\varepsilon$ has been investigated by Monte Carlo by Hsieh and LeBaron (1988a, b). They find that size and power performance are best when $\varepsilon$ is chosen between 0.5 and 1.5 times the standard deviation of the estimated innovations being tested for non IID.


4. The proposed test in this paper has power against non-independent stochastic processes where the rank test does not have any power. See Table 1.2 of Baek (1988).

5. If $h$ is the indicator kernel, $h(u,v) = 1$ whenever $|u-v| \leq \varepsilon$, and 0 otherwise. Then $E h_1^*(u) = \int \frac{d}{du} [f(x(u,v;\varepsilon))dF(v)]dF(u)$

$$
= \int[f(u+\varepsilon)-f(u-\varepsilon)]f(u)du = 0
$$

where $F$ is the distribution function and $f$ is the density function of $u$. 

References


