Testing for independence between two covariance stationary time series

Hong, Yongmiao

September 1996

Biometrika (1996), **83**, 3, *pp.* 615–625
Printed in Great Britain

Testing for independence between two covariance stationary time series

By YONGMIAO HONG
Department of Economics, Cornell University, Ithaca, New York 14853, U.S.A.

Summary

A one-sided asymptotically normal test for independence between two stationary time series is proposed by first prewhitening the two time series and then basing the test on the residual cross-correlation function. The test statistic is a properly standardised version of the sum of weighted squares of residual cross-correlations, with weights depending on a kernel function. Haugh's (1976) test can be viewed as a special case of our approach in the sense that it corresponds to the use of the truncated kernel. Many kernels deliver better power than Haugh's test. A simulation study shows that the new test has good power against short and long cross-correlations.

Some key words: Coherency; Cross-correlation; Independence; Kernel function; Multivariate time series.

1. Introduction

Recently there has been growing interest in testing serial dependence within a univariate time series, e.g. Chan & Tran (1992), Robinson (1991), Skaug & Tjostheim (1993a, b). In contrast, relatively few attempts have been made to test dependence between time series. Dependence between time series is important in multivariate time series analysis. In economics, for example, elucidation of various causalities between time series is vital to forecasting and prediction.

In exploiting dependence between two covariance stationary time series, say (X_t) and (Y_t), one is often interested in testing whether they are mutually independent. Here we propose a test for uncorrelatedness between (X_t) and (Y_t) by first prewhitening X_t and Y_t and then basing the test on the residual cross-correlation function. Our test statistic is a properly standardised version of the sum of weighted squares of residual cross-correlations, with weights depending on a kernel function. The test is asymptotically normally distributed under the null hypothesis.

Haugh (1976) proposed an asymptotically χ^2 test based on the sum of finitely many squares of residual cross-correlations. Haugh's test can be viewed as a special case of our approach with the use of the truncated kernel. In an influential paper, Pierce (1977) used Haugh's test to investigate relationships between a number of aggregate economic time series, and found little or no relationships between most of the economic series. From an econometric point of view, this might be partly due to low power of Haugh's test. Indeed, Geweke (1981a, b) finds that Haugh's test often has low power. In this paper, we find that many kernels deliver better power than Haugh's test or the truncated kernel based test. Within a suitable class of kernel functions, the Daniell kernel maximises the power of our test under both local and fixed alternatives. In addition, we avoid Haugh's assumption that X_t and Y_t have an ARMA, autoregressive-moving average, representation, which, if misspecified, will invalidate the asymptotic distribution of the test statistic.

In § 2, we introduce the test statistic. Asymptotic normality is established in § 3. In §§ 4 and 5, we investigate asymptotic local and global power. In § 6, we examine finite sample performance of the new test in comparison with Haugh's (1976) test via Monte Carlo methods. All mathematical proofs are available from the author upon request.

2. THE TEST STATISTIC

Throughout, we impose the following assumption on X_t and Y_t.

Assumption 1. The stochastic sequence (X_t, Y_t) is a bivariate jointly stationary linear process such that

$$X_t = \sum_{j=0}^{\infty} a_j u_{t-j}, \quad Y_t = \sum_{j=0}^{\infty} b_j v_{t-j} \quad (t = 1, \ldots, N),$$

where (i) (u_t) and (v_t) are each an identically and independently distributed sequence, with $E(u_t) = 0$, $E(v_t) = 0$, $E(u_t^2) = \sigma_u^2$, $E(v_t^2) = \sigma_v^2$, $E(u_t^4) < \infty$ and $E(v_t^4) < \infty$; (ii) (a_j) and (b_j) are sequences of real numbers such that $\sum_{j=0}^{\infty} |a_j| < \infty$, $\sum_{j=0}^{\infty} |b_j| < \infty$ with $a_0 = b_0 = 1$. Furthermore, $|\sum_{j=0}^{\infty} a_j z^j|$ and $|\sum_{j=0}^{\infty} b_j z^j|$ are bounded away from zero for $|z| \leqslant 1$.

This includes as special cases AR, autoregressive, MA, moving average, and ARMA models of finite but possibly unknown orders. For such linear processes, it is well known (Haugh, 1976, p. 379) that (X_t) and (Y_t) are uncorrelated if and only if the innovations (u_t) and (v_t) are uncorrelated. Consequently, one can test independence between (X_t) and (Y_t) by first prewhitening X_t and Y_t and then testing independence between the residuals, say $(\hat{u}_t)$ and $(\hat{v}_t)$. This approach, as pointed out by Haugh (1976), is much easier to handle and interpret, because it filters out the autocorrelation of X_t and Y_t.

Assumption 1 implies that X_t and Y_t have an AR(∞) representation:

$$A(L)X_t = u_t, \quad B(L)Y_t = v_t,$$

where

$$A(L) = 1 - \sum_{j=1}^{\infty} \alpha_j L^j = \left(\sum_{j=0}^{\infty} a_j L^j\right)^{-1}, \quad B(L) = 1 - \sum_{j=1}^{\infty} \beta_j L^j = \left(\sum_{j=0}^{\infty} b_j L^j\right)^{-1},$$

with L a lag operator. We fit X_t by an AR(p) model. The ordinary least squares residual is

$$\hat{u}_t = X_t - \hat{\alpha}(p)' X_t(p),$$

where $X_t(p) = (X_{t-1}, \ldots, X_{t-p})'$, and $\hat{\alpha}(p)$ is the ordinary least squares estimator

$$\hat{\alpha}(p) = \left\{\sum_{t=p+1}^{N} X_t(p) X_t(p)'\right\}^{-1} \sum_{t=p+1}^{N} X_t(p) X_t.$$

When X_t is an AR(p_0) process, $\hat{u}_t$ will be consistent for u_t if $p \geqslant p_0$. In general, there exists no p_0 such that $\alpha_j = 0$ for every $j > p_0$. Hence, we must let $p = p(N)$ grow with N properly in order for $\hat{u}_t$ to be consistent for u_t. We will provide proper conditions on p and (α_j) to ensure asymptotic normality of our test statistic.

Similarly, we fit Y_t by an AR(q) model, with the ordinary least squares residual

$$\hat{v}_t = Y_t - \hat{\beta}(q)' Y_t(q).$$

We define the residual cross-correlation function

$$\hat{\rho}_{uv}(j) = \hat{R}_{uv}(j)/\{\hat{R}_{uu}(0)\hat{R}_{vv}(0)\}^{\frac{1}{2}},$$

where the residual cross-covariance function

$$\hat{R}_{uv}(j) = \begin{cases} N^{-1}\sum_{t=j+1}^{N} \hat{u}_t\hat{v}_{t-j} & (j \geqslant 0), \\ N^{-1}\sum_{t=-j+1}^{N} \hat{u}_{t+j}\hat{v}_t & (j < 0), \end{cases}$$

$\hat{R}_{uu}(0) = N^{-1}\sum_{t=1}^{N} \hat{u}_t^2$, and $\hat{R}_{vv}(0) = N^{-1}\sum_{t=1}^{N} \hat{v}_t^2$.

To construct our statistics, we introduce a kernel function k satisfying the following.

Assumption 2. The function $k: \mathbb{R} \to [-1, 1]$ is symmetric, continuous at 0 and at all but a finite number of other points, with $k(0) = 1$ and $\int_{-\infty}^{\infty} k^2(z)\,dz < \infty$.

This includes such commonly-used kernels as the Bartlett, Daniell, Parzen, quadratic-spectral and the truncated kernels; see e.g. Priestley (1981, pp. 446–7).

Our test statistic is

$$Q_N = \frac{N\sum_{j=1-N}^{N-1} k^2(j/M)\hat{\rho}_{uv}^2(j) - S_N(k)}{\{2D_N(k)\}^{1/2}},$$

where the smoothing parameter $M = M(N) \to \infty$, $M/N \to 0$, and

$$S_N(k) = \sum_{j=1-N}^{N-1} (1 - |j|/N)k^2(j/M),$$

$$D_N(k) = \sum_{j=2-N}^{N-2} (1 - |j|/N)(1 - (|j|+1)/N)k^4(j/M).$$

Under some additional conditions on k and M, we can obtain

$$Q_N^* = \frac{N\sum_{j=1-N}^{N-1} k^2(j/M)\hat{\rho}_{uv}^2(j) - MS(k)}{\{2MD(k)\}^{1/2}},$$

where

$$S(k) = \int_{-\infty}^{\infty} k^2(z)\,dz, \quad D(k) = \int_{-\infty}^{\infty} k^4(z)\,dz.$$

Both Q_N and Q_N^* have the same asymptotic null distribution and power properties. We will investigate their finite sample performances by simulation methods in § 6.

Both Q_N and Q_N^* are essentially coherency-based tests because

$$\|\hat{C}_{uv}\|^2 = \sum_{j=1-N}^{N-1} k^2(j/M)\hat{\rho}_{uv}^2(j),$$

where and hereafter

$$\|.\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |.|^2\,d\omega,$$

and

$$\hat{C}_{uv}(\omega) = \sum_{j=1-N}^{N-1} k(j/M)\hat{\rho}_{uv}(j)e^{-ij\omega}$$

is a kernel estimator for coherency $C_{uv}(\omega)$ between u_t and v_t, which is a measure of cross-correlation between u_t and v_t in the frequency domain and has the invariance property that $|C_{uv}(\omega)| = |C_{xy}(\omega)|$ given Assumption 1 (Priestley, 1981, pp. 660–2). Hong (1996) used an analogous frequency domain approach to test autocorrelation for the residual from a linear regression model that includes both lagged dependent variables and exogenous variables.

Haugh (1976) proposed an asymptotic χ^2 test statistic

$$S = N \sum_{j=-M}^{M} \hat{\rho}_{uv}^2(j).$$

Apart from standardisation factors $S_N(k)$ and $D_N(k)$, S can be viewed as a special case of Q_N with the choice of the truncated kernel $k(z) = 1$ for $|z| \leqslant 1$ and $k(z) = 0$ for $|z| > 1$. As will be seen below, many choices of k yield better power than Haugh's test.

On the other hand, the residuals $\hat{u}_t$ and $\hat{v}_t$ used by Haugh (1976) are obtained by fitting a univariate ARMA model of finite order for X_t and Y_t respectively. As pointed out by Haugh (1976), this approach is of somewhat 'parametric nature', because the assumption of an ARMA model is rather unrealistic in practice. Model misspecification may lead to misleading conclusions because it will invalidate the asymptotic distribution of the test statistic. In contrast, we approximate X_t and Y_t by truncated autoregressions with lag truncation numbers growing properly as the sample size increases (Berk, 1974). This ensures that $\hat{u}_t$ and $\hat{v}_t$ are consistent for u_t and v_t.

3. Asymptotic null distribution

We now derive the asymptotic null distribution of Q_N, and thus Q_N^*. For simplicity, we assume that (u_t) and (v_t) are mutually independent under the null hypothesis.

THEOREM 1. *Suppose Assumptions* 1 *and* 2 *hold. Let* $M \to \infty$, $M/N \to 0$. *Let* p *and* q *satisfy*

$$p = o\left(\frac{N^{1/2}}{M^{1/4}}\right), \quad N \sum_{j=p+1}^{\infty} \alpha_j^2 = o\left(\frac{N^{1/2}}{M^{1/4}}\right), \quad q = o\left(\frac{N^{1/2}}{M^{1/4}}\right), \quad N \sum_{j=q+1}^{\infty} \beta_j^2 = o\left(\frac{N^{1/2}}{M^{1/4}}\right).$$

If u_t *is independent of* v_s *for all* t, s, *then* $Q_N \to N(0, 1)$ *in distribution.*

Under the conditions on p and q, the sampling effects of $\hat{\alpha}(p)$ and $\hat{\beta}(p)$ are asymptotically irrelevant to the limiting distribution of Q_N. The condition $p = o(N^{1/2}/M^{1/4})$ requires that p not grow too fast; in particular, p must grow more slowly than $N^{1/2}$, as $M \to \infty$. This ensures that the sampling variance of $\hat{\alpha}(p)$ is asymptotically negligible. On the other hand, $N \sum_{j=p+1}^{\infty} \alpha_j^2 = o(N^{1/2}/M^{1/4})$ requires that p not grow too slowly. This ensures that the bias of the AR(p) model for X_t vanishes sufficiently fast so that it has negligible impact. When α_j decays to zero sufficiently quickly, the conditions on p will be satisfied. The discussion for q is exactly the same.

For practical implications of the conditions on p, consider first the case where X_t is an AR(p_0) process. This implies $\sum_{j=p_0+1}^{\infty} \alpha_j^2 = 0$. If p_0 is known, $p \geqslant p_0$ ensures both the conditions on p for all N. When p_0 is unknown, in general one has to let p grow in order to be larger than p_0. Next, for stationary and invertible ARMA processes of finite orders, α_j will decay at a geometric rate for large j, that is $|\alpha_j| \leqslant \Delta\alpha_{\max}^j$ for some $\alpha_{\max} \in (0, 1)$. It follows that $N \sum_{j=p+1}^{\infty} \alpha_j^2 = o(N^{1/2}/M^{1/4})$ holds provided $p \to \infty$ at any rate faster than $\ln(N)$. Finally, for the general case where X_t is an AR(∞) process, we must let p grow with N properly. Suppose $\sum_{j=p+1}^{\infty}\alpha_j^2 = O(p^{-\nu})$ for some $\nu > 2$. Then $N^{1/2}M^{1/4}/p^{\nu-1} \to 0$ will suffice for $N\sum_{j=p+1}^{\infty}\alpha_j^2 = o(N^{1/2}/M^{1/4})$.

4. Asymptotic local power

We now investigate the asymptotic power of Q_N under a class of local alternatives. For simplicity, we maintain the assumption that u_t is independent of v_s, and consider the following sequence of completely specified models:

$$H_{aN}:\ C_{uv}^0(\omega) = a(N)g(\omega), \quad \omega \in [-\pi, \pi],$$

where $C_{uv}^0(\omega)$ is the coherency function between u_t and v_t, g is a complex-valued continuous function on $[-\pi, \pi]$, and $a(N) \to 0$ so that the local alternative H_{aN} converges to H_0 as $N \to \infty$. Here, the dependence of C_{uv}^0 on N has been made implicit for notational simplicity. This approach is similar to those of Gallant & Jorgenson (1979) and Gallant & White (1988), who also let the specified model approach the data generating process rather than vice versa. This leads to a much simpler analysis and delivers conclusions identical to those that would be reached by fixing the model C_{uv}^0 and moving the data generating process properly.

Theorem 2. *Suppose Assumptions* 1 *and* 2 *hold, and u_t is independent of v_s for all t, s. Let $M \to \infty$, $M/N \to 0$. Let p and q satisfy*

$$p = o\left(\frac{N^{1/2}}{M^{1/4}}\right), \quad N\sum_{j=p+1}^{\infty}\alpha_j^2 = o\left(\frac{N^{1/2}}{M^{1/4}}\right), \quad q = o\left(\frac{N^{1/2}}{M^{1/4}}\right), \quad N\sum_{j=q+1}^{\infty}\beta_j^2 = o\left(\frac{N^{1/2}}{M^{1/4}}\right).$$

Define

$$Q_N^a = \{N\|\hat{C}_{uv} - C_{uv}^0\|^2 - S_N(k)\}/\{2D_N(k)\}^{\frac{1}{2}},$$

where C_{uv}^0 is as in H_{aN}. If $a(N) = M^{1/4}/N^{1/2}$, then $Q_N^a \to N\{\mu(k), 1\}$ in distribution, where

$$\mu(k) = \|g\|^2/\{2D(k)\}^{1/2}, \quad D(k) = \int_{-\infty}^{\infty} k^4(z)\,dz.$$

The test Q_N^a is able to detect a class of local alternatives converging to H_0 at rate $a(N) = M^{1/4}/N^{1/2}$. The slower is M, the more powerful is the test. This is in contrast to the fact that approximation of asymptotic normality improves when p grows fast.

Because $M^{1/4}/N^{1/2}$ grows more slowly than the parametric rate $N^{-1/2}$, our test is less efficient than Haugh's (1976) test under H_{aN}; Haugh assumes an arbitrary but fixed M. This is the price we have to pay for achieving consistency against a larger class of alternatives. Of course, the claim that $M^{1/4}/N^{1/2}$ is slower than $N^{-1/2}$ should not be taken too literally. When $M = N^{1/5}$, for example, we have $M^{1/4}/N^{1/2} = N^{1/20-1/2}$, which is very close to $N^{-1/2}$ even for fairly large N.

By Theorem 2, the asymptotic power of a test based on Q_N^a with size $\alpha \in (0, 1)$ is

$$\lim_{N\to\infty} \mathrm{pr}(Q_N^a > Z_\alpha) = 1 - \Phi\{Z_\alpha - \mu(k)\},$$

where Φ is the cumulative distribution function of $N(0, 1)$, and Z_α is the upper-tail standard normal critical value at level α. This power is a function of k. Suppose $M = N^\nu$ $(0 < \nu < 1)$. Then following an analogous derivation of Pitman (1979, Ch. 7), we obtain that for two tests using kernels k_1 and k_2, the Pitman's asymptotic relative efficiency of k_2 with respect to k_1 is

$$\text{ARE}_P(k_2;\, k_1) = \{D(k_1)/D(k_2)\}^{1/(2-\nu)}.$$

For example, the relative efficiency of the Bartlett kernel $k_B(z) = (1-|z|)1(|z| \leqslant 1)$ to the truncated kernel $k_T(z) = 1(|z| \leqslant 1)$ is

$$\text{ARE}_P(k_B;\, k_T) = 5^{1/(2-\nu)} > 5^{\frac{1}{2}} \simeq 2{\cdot}23$$

for all $0 < \nu < 1$, where $1(.)$ denotes the indicator function. Thus, k_B is about 120% more efficient than k_T; the latter delivers a Haugh's (1976) type test. Many other kernels also deliver better power than the truncated kernel.

We now consider the optimal kernel that maximises the power of Q_N^a over a suitable class of kernel functions. Let $r > 0$ be the largest positive integer such that

$$k^{(r)} := \lim_{z \to 0} \left\{ \frac{1 - k(z)}{|z|^r} \right\}$$

exists, is finite and nonzero. This r is called the 'characteristic exponent' of the function $k(z)$. We consider the following class of kernels with $r = 2$:

$$\kappa(\tau) = \{k \text{ satisfies Assumption 2, } k^{(2)} = \tfrac{1}{2}\tau^2,\ K(\lambda) \geqslant 0 \text{ for } \lambda \in (-\infty, \infty)\},$$

where

$$K(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} k(z) e^{-iz\lambda}\, dz.$$

This includes the Daniell, Parzen and quadratic-spectral kernels, but rules out the truncated and Bartlett kernels.

THEOREM 3. *Suppose the conditions of Theorem 2 hold. Then the Daniell kernel*

$$k_D(z) = \sin(3^{\frac{1}{2}}\tau z)/(3^{\frac{1}{2}}\tau z), \quad z \in (-\infty, \infty)$$

maximises the asymptotic power of Q_N^a over $\kappa(\tau)$.

This conclusion is in contrast to the quadratic-spectral kernel, which is optimal for estimation of f_{uv} using various mean squared error criteria, e.g. Andrews (1991), Priestley (1962). In fact, as shown in Hong (1996), the Daniell kernel is also optimal for entropy and Hellinger metric-based tests for autocorrelation of the residual from a linear dynamic regression model. For hypothesis testing, the quadratic-spectral kernel may be worse than many other kernels.

Three commonly-used kernels, Daniell, Parzen and quadratic-spectral, have $D(k) = 1{\cdot}209200/\tau$, $1{\cdot}325414/\tau$ and $1{\cdot}218851/\tau$ respectively. Thus, while the Daniell kernel is optimal, we expect little power difference among these kernels. Of course, kernels outside $\kappa(\tau)$ may have $D(k)$ significantly different from that of the Daniell kernel.

5. ASYMPTOTIC GLOBAL POWER

Next, we turn to examine asymptotic global power of Q_N. To state the consistency theorem, we impose the following condition on the dependence between (u_t) and (v_t).

Assumption 3. The innovations (u_t) and (v_t) are fourth order stationary processes with

$$\sum_{j=-\infty}^{\infty} R_{uv}^2(j) < \infty, \quad \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\kappa_{uvuv}(0, i, j, l)| < \infty,$$

where $\kappa_{uvuv}(0, i, j, l)$ is the fourth order cumulant of $u_t v_{t+i} u_{t+j} v_{t+l}$.

Here, $R_{uv}(j)$ need not be absolutely summable, as is the case for the alternatives that have such long cross-correlations that the cross-spectral densities do not exist at frequency 0. The cumulant condition is standard in multivariate time series; it characterises the temporal dependence of $(u_t v_t)$. When (u_t, v_t) is a bivariate jointly Gaussian process, the cumulant condition holds trivially because $\kappa_{uvuv}(0, i, j, l) = 0$ for all i, j, l.

THEOREM 4. *Suppose Assumptions* 1–3 *hold. Let* $M \to \infty$, $M/N \to 0$. *Let*

$$p = o(N/M), \quad \sum_{j=p+1}^{\infty} \alpha_j^2 = o(M^{-1}), \quad q = o(N/M), \quad \sum_{j=q+1}^{\infty} \beta_j^2 = o(M^{-1}).$$

Then

$$(M^{1/2}/N) Q_N \to \| C_{xy} \|^2 / \{2D(k)\}^{1/2}$$

in probability.

Theorem 4 implies $Q_N \to \infty$ at rate $N/M^{1/2}$ under fixed alternatives. Asymptotically, the slower M grows, the faster will Q_N diverge to infinity, and so the more powerful is Q_N. This conclusion is analogous to that reached under H_{aN}.

To compare the efficiencies of two tests under fixed alternatives, Pitman's criterion is inappropriate because the asymptotic power of Q_N will approach unity as $N \to \infty$ at any given level $\alpha \in (0, 1)$. Instead, we use Bahadur's (1960) asymptotic slope criterion, which is pertinent for large sample tests under fixed alternatives. Bahadur's asymptotic slope is the rate at which the asymptotic p-value goes to zero as $N \to \infty$. Because Q_N is asymptotically $N(0, 1)$ under the null hypothesis, its asymptotic p-value is $1 - \Phi(Q_N)$. Now define

$$\Gamma_N(k) = -2 \ln \{1 - \Phi(Q_N)\}.$$

Because $\ln \{1 - \Phi(\xi)\} = -\tfrac{1}{2}\xi^2 \{1 + o(1)\}$ as $\xi \to +\infty$ (Bahadur, 1960), we have

$$(M/N^2) \Gamma_N(k) \to \| C_{xy} \|^2 / \{2D(k)\}^{1/2}$$

in probability under fixed alternatives as $M \to \infty$, $M/N \to 0$. Following Bahadur, we call $\| C_{xy} \|^2 / \{2D(k)\}^{1/2}$ the 'asymptotic slope' of Q_N. A large asymptotic slope implies a fast rate at which the asymptotic p-value of the test converges to zero as $N \to \infty$. Furthermore, the rate at which $\Gamma_N(k)$ diverges to infinity is N^2/M; this rate is faster than the rate for parametric tests including asymptotic normal and χ^2 tests, the latter equal to N (Bahadur, 1960). When $M = \ln(N)$, for example, N^2/M is close to the square of N. Consequently, Q_N has an infinitely larger asymptotic slope than parametric tests, including Haugh's S test. This conclusion on relative efficiency under fixed alternatives is in sharp contrast to that reached under H_{aN}.

It can also be shown that Bahadur's relative efficiency comparing two kernels is the same as Pitman's efficiency. Thus, all the discussions on k in § 4 apply.

6. FINITE SAMPLE PERFORMANCE

We now examine finite sample performance of Q_N and Q_N^* in comparison with Haugh's (1976) tests using Monte Carlo methods. We consider two processes for X_t and Y_t:

(a) $X_t = 0{\cdot}5X_{t-1} + u_t$ and $Y_t = 0{\cdot}5Y_{t-1} + v_t$,

(b) $X_t = u_t + 0{\cdot}5u_{t-1}$ and $Y_t = v_t + 0{\cdot}5v_{t-1}$,

where u_t and v_t are identically and independently distributed $N(0, 1)$ random variables. Three alternatives are considered.

Alternative 1:

$$\rho_{uv}(j) = \begin{cases} 0{\cdot}2 & \text{for } j = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Alternative 2:

$$\rho_{uv}(j) = \begin{cases} 0{\cdot}125 & \text{for } j = 0, \\ \sin(0{\cdot}125\pi j)/(\pi j) & \text{for } 1 \leqslant j \leqslant 8, \\ 0 & \text{otherwise.} \end{cases}$$

Alternative 3:

$$\rho_{uv}(j) = \begin{cases} 0{\cdot}3 & \text{for } j = 3, \\ 0 & \text{otherwise.} \end{cases}$$

Under Alternative 1, (u_t) and (v_t) are correlated simultaneously but not otherwise, and the coherency is a nonzero constant $(1/5\pi)$ for all frequencies. This pattern of very short cross-correlation is similar to those of many financial time series. In contrast, the cross-correlation function of Alternative 2 has a maximum at $j = 0$, and then decays slowly and smoothly to 0 at $j = 8$. The coherency has a large nonzero value for all positive frequencies near 0 but is zero otherwise. The correlation is long and smooth; this pattern might be exhibited by two time series that are observed weekly and have strong quarterly relationships, but whose weekly motions are only weakly related. As pointed out by Geweke (1981a), this pattern is similar to the cross-correlations of many estimated innovations that have exhibited to substantiate a finding of little or no relationships between time series, e.g. Pierce (1977). For Alternative 3, (u_t) and (v_t) are correlated only at lag $j = 3$.

The simulation experiment was carried out using a GAUSS random number generator on a 486 PC. Two sample sizes are used: $N = 100$ and 200. For each N, we generate $N + 50$ observations and then discard the first 50 to reduce the effects of initial values. We use AR(p)/AR(q) to fit X_t/Y_t, with $p, q = 3$ for $N = 100$, and $p, q = 6$ for $N = 200$. To examine effects of using different k and M, we use three kernels from the class $\kappa(\pi/3^{\frac{1}{2}})$, and three rates for M. The three kernels are Daniell, Parzen and quadratic-spectral kernels. The three rates are $M = \lfloor \ln(N) \rfloor$, $\lfloor 3N^{0{\cdot}2} \rfloor$ and $\lfloor 3N^{0{\cdot}3} \rfloor$, where $\lfloor a \rfloor$ denotes the integer part of a. These rates deliver $M = 5, 8, 12$ for $N = 100$ and $M = 5, 9, 15$ for $N = 200$.

We also compute Haugh's (1976) two statistics

$$S = N \sum_{j=-M}^{M} \hat{\rho}_{uv}^2(j), \quad S^* = N^2 \sum_{j=-M}^{M} (N - j)^{-1} \hat{\rho}_{uv}^2(j),$$

where the above three rules for M also apply. Both S and S^* are asymptotically χ^2_{2M+1} under the null hypothesis.

Because the performances of each test are much the same whether X_t/Y_t follow AR(1) or MA(1) processes, we only report results when X_t/Y_t follow AR(1) processes. Table 1 reports size performances of all the tests at 10% and 5% nominal significance levels, based on 1000 replications. Both Q_N and Q_N^* have reasonable sizes, and there is no clear evidence

Table 1. *Rejection rates out of* 1000 *replications under the null hypothesis of independence,* $X_t = 0{\cdot}5X_{t-1} + u_t$, $Y_t = 0{\cdot}5Y_t + v_t$, *where* u_t, $v_t \sim N(0, 1)$, *and* $\rho_{uv}(j) = 0$ *for all* j

		N = 100						N = 200					
		M = 5		M = 8		M = 12		M = 5		M = 9		M = 15	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Q_N	DN	10·2	6·4	9·6	5·4	8·6	4·8	9·0	6·0	9·9	6·2	9·7	6·5
	PZ	10·6	5·9	9·0	5·6	8·0	4·6	9·1	5·7	10·1	6·8	9·5	6·3
	QS	10·4	6·2	9·3	5·5	8·1	5·0	8·8	5·5	8·5	5·5	8·8	5·7
Q_N^*	DN	9·6	5·1	7·9	5·4	5·6	3·7	8·7	6·0	8·0	5·7	7·7	5·1
	PZ	9·4	5·4	7·6	4·8	6·7	4·0	8·8	5·6	9·5	6·1	8·3	5·4
	QS	10·0	5·6	8·5	5·0	6·5	4·1	8·8	5·5	8·5	5·5	8·8	5·7
S		7·8	3·6	6·5	2·2	3·9	1·1	8·3	4·4	7·7	3·7	6·7	2·7
S^*		8·7	4·5	7·8	3·3	7·7	3·2	8·6	4·8	9·2	4·5	8·7	4·2

DN, Daniell kernel; PZ, Parzen kernel; QS, quadratic-spectral kernal.

favouring either one. At the 10% level, Q_N and Q_N^* have better sizes than S and S^*. At the 5% level, Q_N and Q_N^* exhibit a little over-rejection in some cases, while S and S^* exhibit a little under-rejection in some cases. The test S^* has better size than S at both the 10% and 5% levels, especially for $N = 100$.

Table 2 reports power performances under Alternative 1 at the 5% level, based on 500 replications. We use both asymptotic and empirical critical values, the latter obtained from the 1000 replications under the null hypothesis. Both Q_N and Q_N^* perform similarly. The three kernels deliver similar power. For each kernel, the more slowly M grows, the better is the power of the test. In fact, because Alternative 1 is a simultaneous cross-correlation, including extra terms will sacrifice efficiency of the tests. On the other hand, S and S^* perform similarly, and smaller M gives better power. We see that Q_N and Q_N^* are about twice as powerful as Haugh's tests.

Table 2. *Rejection rates out of* 500 *replications at the* 5% *level under Alternative* 1: $X_t = 0{\cdot}5X_{t-1} + u_t$, $Y_t = 0{\cdot}5Y_t + v_t$, *where* u_t, $v_t \sim N(0,1)$, $\rho_{uv}(0) = 0{\cdot}2$ *and* $\rho_{uv}(j) = 0$ *for all* $j \neq 0$

		N = 100						N = 200					
		M = 5		M = 8		M = 12		M = 5		M = 9		M = 15	
		ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV
Q_N	DN	38·8	35·4	29·0	28·4	24·0	24·6	66·6	62·4	55·8	51·8	44·4	40·4
	PZ	37·0	32·6	28·8	27·2	23·8	25·2	65·0	63·4	54·4	50·2	42·8	38·4
	QS	39·0	34·2	29·4	27·8	24·6	24·6	66·8	64·4	55·8	53·0	44·4	40·2
Q_N^*	DN	36·6	35·2	26·8	28·4	19·2	24·6	65·8	62·2	53·6	52·0	40·8	40·6
	PZ	36·0	32·6	26·0	27·2	20·6	25·2	64·6	63·4	52·8	50·2	39·8	38·4
	QS	36·6	34·2	27·8	27·8	21·4	24·6	65·8	64·4	54·6	53·0	42·6	40·0
S		13·4	16·2	10·4	16·4	6·0	15·2	35·0	36·8	24·0	28·8	17·6	23·6
S^*		14·4	15·8	12·8	15·8	10·6	14·6	35·8	36·0	26·6	28·0	21·4	22·4

DN, Daniell kernel; PZ, Parzen kernel; QS, quadratic-spectral kernel; ACV, asymptotic critical value; ECV, empirical critical value.

Table 3 reports power performances under Alternative 2. Again, for Q_N and Q_N^*, the three kernels perform roughly the same. The effects of M, however, are somewhat different from those under Alternative 1. Now the three choices of M give similar power, and there is only weak evidence that smaller M delivers better power. In contrast, for S and S^*, smaller M still delivers better power. Both Q_N and Q_N^* are more powerful than S and S^*, especially for small N and/or medium and large M.

Table 4 reports power performances under Alternative 3. For Q_N and Q_N^*, the Parzen kernel is a little more powerful than the Daniell and quadratic-spectral kernels when M is small, but the three kernels perform similarly when M is medium and large. For each kernel, small M gives low power, while medium and large M give good power. On the other hand, smaller M gives better power for S and S^*. Interestingly, Haugh's tests have better power than Q_N and Q_N^* when M is small. This can be expected because Haugh's tests put more weight than Q_N and Q_N^* on $j=3$ for which $\rho_{uv}(j) \neq 0$. When M is large,

Table 3. *Rejection rates out of 500 replications at the 5% level under Alternative 2: $X_t = 0{\cdot}5X_{t-1} + u_t$, $Y_t = 0{\cdot}5Y_t + v_t$, where u_t, $v_t \sim N(0, 1)$, and $\rho_{uv}(j) = \sin(0{\cdot}125\pi j)/(\pi j)$ for $0 \leqslant j \leqslant 8$, and $\rho_{uv}(j) = 0$ otherwise.*

		$N=100$						$N=200$					
		$M=5$		$M=8$		$M=12$		$M=5$		$M=9$		$M=15$	
		ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV
Q_N	DN	34·4	32·0	34·4	32·0	29·6	29·6	65·4	62·2	70·4	65·8	65·6	59·8
	PZ	35·0	31·4	32·0	30·6	28·8	29·6	69·0	66·0	70·8	66·0	63·2	58·6
	QS	34·4	31·8	34·2	31·4	29·2	29·8	66·0	63·6	70·8	67·0	65·4	58·4
Q_N^*	DN	32·0	32·0	29·4	31·8	23·2	29·6	64·8	62·2	68·4	65·8	60·2	60·0
	PZ	33·6	31·4	29·2	30·6	23·0	29·6	67·6	65·6	68·6	66·0	60·2	58·2
	QS	33·6	31·8	31·4	31·4	26·2	29·4	65·8	63·6	69·4	67·0	63·2	58·4
S		19·4	24·6	13·4	21·4	9·2	18·0	58·6	61·8	40·4	46·2	25·8	37·2
S^*		22·8	24·8	17·6	21·0	13·4	18·0	60·0	61·6	43·0	45·0	33·6	35·8

DN, Daniell kernel; PZ, Parzen kernel; QS, quadratic-spectral kernel; ACV, asymptotic critical value; ECV, empirical critical value.

Table 4. *Rejection rates out of 500 replications at the 5% level under Alternative 3: $X_t = 0{\cdot}5X_{t-1} + u_t$, $Y_t = 0{\cdot}5Y_t + v_t$, where u_t, $v_t \sim N(0, 1)$, and $\rho_{uv}(3) = 0{\cdot}3$ and $\rho_{uv}(j) = 0$ for all $j \neq 3$*

		$N=100$						$N=200$					
		$M=5$		$M=8$		$M=12$		$M=5$		$M=9$		$M=15$	
		ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV
Q_N	DN	15·8	14·6	42·8	41·6	46·8	46·8	40·2	33·2	83·6	80·2	83·6	81·4
	PZ	26·0	21·6	44·2	42·8	46·2	48·4	61·4	58·6	84·2	81·6	82·6	80·2
	QS	17·4	14·6	42·8	41·2	47·0	47·0	45·0	40·8	83·8	81·2	83·2	81·2
Q_N^*	DN	14·6	14·6	39·0	41·6	40·2	46·8	39·6	33·0	81·8	80·4	81·4	81·4
	PZ	25·0	22·0	42·4	42·8	41·8	48·4	61·0	58·4	83·2	81·6	81·4	80·2
	QS	16·4	14·6	41·2	41·2	43·6	47·0	44·4	41·2	82·8	81·2	82·2	81·2
S		40·0	45·6	28·2	38·0	18·2	34·2	82·6	83·2	64·6	70·8	47·6	57·2
S^*		44·2	46·2	33·2	37·6	26·6	33·8	82·6	83·4	68·6	70·6	53·0	54·8

DN, Daniell kernel; PZ, Parzen kernel; QS, quadratic-spectral kernel; ACV, asymptotic critical value; ECV, empirical critical value.

however, Haugh's tests become less powerful than Q_N and Q_N^*. This is because, although Haugh's tests put more weight on $j = 3$, they also, inefficiently, put more weights than Q_N and Q_N^* on many lags for which $\rho_{uv}(j) = 0$.

In summary, the simulation study shows that the new tests perform reasonably well, having good power against short and long cross-correlations. Different choices of kernel, other than the truncated kernel, give similar power. In most cases, the new tests have better power than Haugh's tests or the truncated kernel based tests.

Acknowledgement

I would like to thank a referee, an Associate Editor, the Editor, D. Easley, J. Geweke, G. Jakubson, N. Kiefer, H. Pesaran and M. Wells for providing useful comments. The research was carried out with a faculty research summer support from the Department of Economics, Cornell University.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–58.
Bahadur, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31**, 276–95.
Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2**, 489–502.
Chan, N. H. & Tran, L. T. (1992). Nonparametric tests for serial dependence. *J. Time Ser. Anal.* **13**, 102–13.
Gallant, A. R. & Jorgenson, D. (1979). Statistical inference for a system of simultaneous nonlinear implicit equations in the context of instrumental variables estimation. *J. Economet.* **11**, 275–302.
Gallant, A. R. & White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.
Geweke, J. (1981a). The approximate slopes of econometric tests. *Econometrica* **49**, 1427–42.
Geweke, J. (1981b). A comparison of tests of independence of two covariance stationary time series. *J. Am. Statist. Assoc.* **76**, 363–73.
Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross correlation approach. *J. Am. Statist. Assoc.* **71**, 378–85.
Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–64.
Pierce, A. (1977). Lack of dependence among economic variables. *J. Am. Statist. Assoc.* **72**, 11–22.
Pitman, E. J. G. (1979). *Some Basic Theory for Statistical Inference*. London: Chapman Hall.
Priestley, M. B. (1962). Basic considerations in the estimation of spectra. *Technometrics* **4**, 551–64.
Priestley, M. B. (1981). *Spectral Analysis and Time Series*, **1**, **2**. London: Academic Press.
Robinson, P. M. (1991). Consistent nonparametric entropy-based testing. *Rev. Econ. Studies* **58**, 437–53.
Skaug, H. J. & Tjostheim, D. (1993a). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **80**, 591–602.
Skaug, H. J. & Tjostheim, D. (1993b). Nonparametric tests of serial independence. In *Developments in Time Series Analysis*, the M. B. Priestley Birthday Volume, Ed. T. Subba Rao, pp. 207–29. London: Academic Press.

[*Received August* 1994. *Revised August* 1995]