Econometrica, Vol. 64, No. 4 (July, 1996), 837–864

CONSISTENT TESTING FOR SERIAL CORRELATION OF UNKNOWN FORM[1]

BY YONGMIAO HONG

This paper proposes three classes of consistent one-sided tests for serial correlation of unknown form for the residual from a linear dynamic regression model that includes both lagged dependent variables and exogenous variables. The tests are obtained by comparing a kernel-based normalized spectral density estimator and the null normalized spectral density, using a quadratic norm, the Hellinger metric, and the Kullback-Leibler information criterion respectively. Under the null hypothesis of no serial correlation, the three classes of new test statistics are asymptotically $N(0,1)$ and equivalent. The null distributions are obtained without having to specify any alternative model. Unlike some conventional tests for serial correlation, the null distributions of our tests remain invariant when the regressors include lagged dependent variables. Under a suitable class of local alternatives, the three classes of the new tests are asymptotically equally efficient. Under global alternatives, however, their relative efficiencies depend on the relative magnitudes of the three divergence measures. Our approach provides an interpretation for Box and Pierce's (1970) test, which can be viewed as a quadratic norm based test using a truncated periodogram. Many kernels deliver tests with better power than Box and Pierce's test or the truncated kernel based test. A simulation study shows that the new tests have good power against an AR(1) process and a fractionally integrated process. In particular, they have better power than the Lagrange multiplier tests of Breusch (1978) and Godfrey (1978) as well as the portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978). The cross-validation procedure of Beltrao and Bloomfield (1987) and Robinson (1991a) works reasonably well in determining the smoothing parameter of the kernel spectral estimator and is recommended for use in practice.

KEYWORDS: Consistent tests, cross-validation, entropy, Hellinger metric, local and global alternatives, quadratic norm, serial correlation of unknown form, spectral estimation, strong dependence.

1. INTRODUCTION

CONSIDER A LINEAR AUTOREGRESSIVE distributed lag dynamic regression (AD) model

$$(1)\qquad \alpha^{(0)}(B)Y_t = c + \alpha^{(1)}(B)X_{1t} + \cdots + \alpha^{(q)}(B)X_{qt} + u_t \qquad (t = 1, 2, \ldots, n),$$

where the $\alpha^{(j)}(B) = \sum_{l=0}^{m_j} \alpha_{lj} B^l$ are polynomials of order m_j in lag operator B associated with the dependent variable Y_t and the q exogenous variables X_{jt}, c

[1] This is a revised version of earlier manuscripts entitled "Consistent, Regression-based Testing for Serial Correlation of Unknown Form" and "Consistent Spectrum-Based Testing for Serial Correlation of Unknown Form." I am most grateful to a co-editor, who went well beyond the call of duty in providing constructive suggestions and references. I would also like to thank D. Easley, J. Geweke, G. Jakubson, N. M. Kiefer, M. Pesaran, R. Shehadeh, T. Vogelsang, H. White, F. Xu, two referees, and seminars participants at Hong Kong University of Science and Technology, Queen's and Yale Universities, and University of Guelph for providing helpful comments and references on the earlier versions. A faculty research summer support from the Department of Economics at Cornell University is gratefully acknowledged. All errors are attributed solely to the author.

is a constant, and u_t is an unobservable disturbance. The polynomial $\alpha^{(0)}(B)$ is assumed to have all roots outside the unit circle, and is normalized by setting $\alpha_{00} = 1$. Throughout it is also assumed that the X_{jt} are covariance stationary with $E(X_{jt}^2) < \infty$.

Put $\alpha_0 = (\alpha_{10}, \ldots, \alpha_{m_0 0})'$ and $\alpha_j = (\alpha_{0j}, \alpha_{1j}, \ldots, \alpha_{m_j j})'$, $j = 1, \ldots, q$. Then $\alpha = (c, \alpha_0', \ldots, \alpha_q')'$ is a $\sum_{j=0}^{q}(m_j + 1) \times 1$ vector consisting of all unknown coefficients in (1). We can estimate α by (e.g.) the ordinary least squares (OLS) method. A key condition for the consistency of the OLS estimator for α in (1) is that $\{u_t\}$ be serially uncorrelated. Serial correlation of $\{u_t\}$ may occur due to misspecification of (1), such as omitting relevant variables, choosing too low a lag order for Y_t or the X_{jt}, or using inappropriate transformed variables. In general, any form of serial correlation will render inconsistent the OLS estimator for α and/or its usual standard covariance matrix estimator.

Most existing tests are not consistent against serial correlation of unknown form. Consistent tests are useful when no prior information about the true alternative is available. Andrews and Ploberger (1994) recently proposed a class of consistent tests against weakly stationary strong mixing alternatives for the residual from a static regression model. In this paper, we propose some consistent tests for serial correlation of unknown form for the residual from (1), with possibly good power against strong dependence. We compare a kernel-based normalized spectral density estimator to the null normalized spectral density, using a quadratic norm, the Hellinger metric, and the Kullback-Leibler information criterion respectively. The null limit distributions of our tests are all $N(0,1)$ and are derived without having to specify an alternative model. Unlike some conventional tests (e.g. Box and Pierce (1970) (BP), Ljung and Box (1978) (LB), Durbin and Watson (1950, 1951)), the null distributions of our tests remain invariant whether the regressors include lagged dependent variables. Our approach also provides an interpretation for BP, which can be viewed as a test based on a quadratic norm with the use of a truncated periodogram.

Our tests are asymptotically equivalent under a suitable class of local alternatives, but their relative efficiencies under global alternatives depend on the relative magnitudes of the divergence measures. Within a suitable class of kernel functions, the Daniell kernel maximizes the power of our tests under both local and global alternatives. Many kernels deliver tests with better power than the truncated kernel-based test or the tests of BP and LB. Because the Lagrange multiplier (LM) tests of Breusch (1978) and Godfrey (1978) are similar in spirit to BP and LB in the sense that they all put uniform weights on the autocorrelations under tested, we expect that the new tests may have better power than the LM tests. Simulation shows that the new tests indeed have better power than the LM tests as well as BP and LB, against an AR(1) process and a fractionally integrated process.

In Section 2, we describe our method and test statistics. The null asymptotic normality is derived in Section 3. In Sections 4 and 5, we investigate asymptotic local and global power properties respectively. In Section 6, we conduct a simulation study of finite sample performance of our tests in comparison to some commonly used tests. The last section concludes the paper. All proofs are given in the Appendix.

2. METHOD AND TEST STATISTICS

Suppose $\{u_t\}$ is a stationary real-valued process with $E(u_t)=0$, autocovariance function $R(j)$, autocorrelation function $\rho(j)$, and normalized spectral density function

$$f(\omega)=(2\pi)^{-1}\sum_{j=-\infty}^{\infty}\rho(j)\cos j\omega, \qquad \omega\in[-\pi,\pi].$$

The hypotheses of interest are

$$H_0\colon \rho(j)=0 \quad \text{for all } j\neq 0 \quad \text{v.s.} \quad H_A\colon \rho(j)\neq 0 \quad \text{for some } j\neq 0.$$

The hypothesis H_0 is equivalent to $f(\omega)=f_0(\omega)$ for all $\omega\in[-\pi,\pi]$, where $f_0(\omega)=1/2\pi$ for $\omega\in[-\pi,\pi]$. Let $D(f_1;f_2)$ be a divergence measure for two spectral densities f_1, f_2 such that $D(f_1;f_2)\geq 0$ and $D(f_1;f_2)=0$ if and only if $f_1=f_2$. Then a consistent test for H_0 can be based on $D(\hat{f}_n;f_0)$, where $\hat{f}_n$ is a kernel estimator for f. Examples of D include the quadratic norm of f from f_0,

$$Q(f;f_0)=\left[2\pi\int_{-\pi}^{\pi}(f(\omega)-f_0(\omega))^2\,d\omega\right]^{1/2},$$

the Hellinger metric,

$$H(f;f_0)=\left[\int_{-\pi}^{\pi}(f^{1/2}(\omega)-f_0^{1/2}(\omega))^2\,d\omega\right]^{1/2},$$

and the Kullback-Leibler information criterion (relative entropy),

$$I(f;f_0)=-\int_{\Omega(f)}\ln(f(\omega)/f_0(\omega))f_0(\omega)\,d\omega,$$

where $\Omega(f)=\{\omega\in[-\pi,\pi]\colon f(\omega)>0\}$. These measures are intuitively appealing and have their own merits. The quadratic norm delivers a computationally convenient statistic that is simply a weighted average of squared sample autocorrelations with weights depending on the kernel. BP can be viewed as based on $Q(\hat{f}_n;f_0)$ with $\hat{f}_n$ being a truncated periodogram. Note that $H(f;f_0)$ is a quadratic norm between $f^{1/2}$ and $f_0^{1/2}$. Unlike $Q(f;f_0)$, which gives the same weight to the difference between f and f_0 whether the smaller of the two is large or small, $H(f;f_0)$ is relatively robust to outliers and is thus particularly suitable for contaminated data (cf. Pitman (1979)). Finally, entropy-based tests have an appealing information-theoretic interpretation.

We consider three classes of consistent tests for H_0 based on $Q(\hat{f}_n;f_0)$, $H(\hat{f}_n;f_0)$, and $I(\hat{f}_n;f_0)$ respectively. Let $\hat{\alpha}$ be an estimator for α. Then the residual of (1) is

$$\hat{u}_t=\hat{\alpha}^{(0)}(B)Y_t-\hat{c}-\hat{\alpha}^{(1)}(B)X_{1t}-\cdots-\hat{\alpha}^{(q)}(B)X_{qt}.$$

Define the residual-based sample autocorrelation function

$$\hat{\rho}(j) = \hat{R}(j)/\hat{R}(0) \qquad (j = 0, \pm 1, \ldots, \pm (n-1)),$$

where $\hat{R}(j) = n^{-1} \sum_{t=|j|+1}^{n} \hat{u}_t \hat{u}_{t-|j|}$. A kernel estimator for f is given by

$$\hat{f}_n(\omega) = (2\pi)^{-1} \sum_{j=-n+1}^{n-1} k(j/p_n) \hat{\rho}(j) \cos(j\omega), \quad \omega \in [-\pi, \pi],$$

where the bandwidth $p_n \to \infty$, $p_n/n \to 0$; the kernel k satisfies the following conditions.

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

The conditions that $k(0) = 1$ and k is continuous at 0 imply that for j small relative to n, the weight given to $\hat{\rho}(j)$ is close to unity (the maximum weight). Assumption A.1(a) includes the Bartlett, Daniell, general Tukey, Parzen, Quadratic-Spectral (QS), and truncated kernels (e.g. Priestley (1981, p. 441)). Of them, the Bartlett, general Tukey, and Parzen kernels are of compact support, i.e. $k(z) = 0$ for $|z| > 1$. For these kernels, p_n is called the "lag truncation number," because the lags of order $j > p_n$ receive zero weight. In contrast, the Daniell and QS kernels are of unbounded support; here p_n is not a "truncation point," but determines the "degree of smoothing" for $\hat{f}_n$.

The first class of tests are a proper standardized version of $Q(\hat{f}_n; f_0)$:

$$(2) \qquad M_{1n} = \left(\tfrac{1}{2} n Q^2\left(\hat{f}_n; f_0\right) - C_n(k)\right) \Big/ (2D_n(k))^{1/2}$$
$$= \left(n \sum_{j=1}^{n-1} k^2(j/p_n) \hat{\rho}^2(j) - C_n(k)\right) \Big/ (2D_n(k))^{1/2},$$

where the second equality follows by Parseval's identity, and

$$C_n(k) = \sum_{j=1}^{n-1} (1 - j/n) k^2(j/p_n),$$
$$D_n(k) = \sum_{j=1}^{n-2} (1 - j/n)(1 - (j+1)/n) k^4(j/p_n).$$

Given $p_n \to \infty$ and $p_n/n \to 0$, $p_n^{-1} D_n(k) \to D(k) = \int_0^{\infty} k^4(z)\, dz$. Thus, one can replace $D_n(k)$ by $p_n D(k)$ without affecting the asymptotic distribution of M_{1n}. Under some additional conditions on k and/or p_n (cf. Robinson (1994, p. 73)), we have $p_n^{-1} C_n(k) = C(k) + o(p_n^{-1/2})$, where $C(k) = \int_0^{\infty} k^2(z)\, dz$. In this case, we can also replace $C_n(k)$ by $p_n C(k)$. Thus, a more compact expression of M_{1n} can be given as

$$M_{1n}^* = \left(n \sum_{j=1}^{n-1} k^2(j/p_n) \hat{\rho}^2(j) - p_n C(k)\right) \Big/ (2p_n D(k))^{1/2}.$$

Based on raw observed data, Anderson (1993, p. 841) shows that the Cramer-von Mises test statistic for H_0 (which is based on a quadratic norm between the integrated sample spectrum and the null integrated spectrum) is asymptotically equal to $n\sum_{j=1}^{n-1}\hat{\rho}^2(j)/j^2$. It appears that M_{1n} may be more flexible because we allow more room to choose weights via k and p_n. In particular, M_{1n} may have better power against strong dependence than the Cramer-von Mises test and other integrated spectrum based tests (e.g. Durlauf (1991)) because the latter put rather heavy weights on low order sample autocorrelations. In addition, the Cramer-von Mises test and other integrated spectrum based tests have nonstandard distributions. For these tests, the critical regions cannot be found in the basic statistical tables, and tabulations of the percentiles of the distributions are required.

When k is the truncated kernel, i.e. $k(z)=1$ for $|z|\leq 1$ and 0 for $|z|>1$, we obtain

$$(3)\qquad M_{1n}^T=\left(n\sum_{j=1}^{p_n}\hat{\rho}^2(j)-p_n\right)\Bigg/(2p_n)^{1/2},$$

a generalized BP's test when $p_n\to\infty$. Under H_0, M_{1n}^T is asymptotically equivalent to

$$(4)\qquad M_R=(nR^2-p_n)/(2p_n)^{1/2},$$

where R^2 is the squared multi-correlation coefficient from the AR(p_n) regression

$$(5)\qquad \hat{u}_t=\beta_1\hat{u}_{t-1}+\beta_2\hat{u}_{t-2}+\cdots+\beta_{p_n}\hat{u}_{t-p_n}+\epsilon_{p_nt}\qquad (t=1,2,\ldots,n),$$

where initial values $\hat{u}_{t-p_n}=0$, $1\leq t\leq p_n$. Hence, M_{1n}^T can be viewed as a test for the hypothesis that the p_n coefficients of the AR(p_n) model are jointly equal to zero. Because any stationary invertible linear process with continuous f can be approximated well by a truncated AR model with sufficiently high order (cf. Berk (1974)), M_R will eventually capture all possible autocorrelations as long as more and more lags of $\hat{u}_t$ are included as n increases. When M_R rejects H_0, t statistics in (5) may provide useful information about the pattern of serial correlation. We note that the power of M_R may be different from that of M_{1n}^T, because in general they are not asymptotically equivalent under H_A.

Like BP, M_{1n}^T and M_R put equal weight for all p_n sample autocorrelations. Intuitively, this might not be the optimal weighting because for most stationary processes the autocorrelation decays to zero as the lag increases. Therefore, we expect that tests based on kernels other than the truncated kernel may give better power than BP, M_{1n}^T and M_R.[2] Because the LM tests of Breusch (1978) and Godfrey (1978) are similar in spirit to BP (they are asymptotically equiva-

[2] Another argument against the truncated kernel is that in practice one may be interested only in low order autocorrelations. Thus, a better test should put more weights on low order lags rather than put uniform weights on all p_n sample autocorrelations.

lent under a static regression model), we expect that the LM tests may also be less powerful than tests based on kernels rather than the truncated kernel.

The other two classes of test statistics are[3]

$$(6) \qquad M_{2n} = \left(2nH^2\left(\hat{f}_n; f_0\right) - C_n(k)\right) \Big/ (2D_n(k))^{1/2},$$

$$(7) \qquad M_{3n} = \left(nI\left(\hat{f}_n; f_0\right) - C_n(k)\right) \Big/ (2D_n(k))^{1/2}.$$

For (6) and (7), we impose the following additional condition on k.

ASSUMPTION A.1(b): *$\int_{-\pi}^{\pi} |k(z)|\, dz < \infty$ and $K(\lambda) = (1/2\pi)\int_{-\infty}^{\infty} k(z) e^{-iz\lambda}\, dz \geq 0$ for $\lambda \in (-\infty, \infty)$.*

The absolute integrability of k ensures that its Fourier transform K exists. Assumption A.1 implies that K is a symmetric density. Because $\hat{f}_n$ can be written as an integral of the convolution of $p_n \sum_{j=-\infty}^{\infty} K(p_n(\cdot + 2\pi j))$ with the periodogram of $\{\hat{u}_t\}$ (cf. Robinson (1991a, (2.5))), it follows that $\hat{f}_n(\omega) \geq 0$ for all ω. Assumption A.1 includes the Bartlett, Daniell, Parzen, and QS kernels, but rules out the truncated and general Tukey kernels.

3. ASYMPTOTIC NULL DISTRIBUTION

To establish asymptotic null normality of our tests, we assume the following conditions.

ASSUMPTION A.2: *$\{u_t\}$ is identically and independently distributed (i.i.d.) with $E(u_t) = 0$, $E(u_t^2) = \sigma_0^2$, and $E(u_t^4) = \mu_4 < \infty$.*

ASSUMPTION A.3: $n^{1/2}(\hat{\alpha} - \alpha) = O_P(1)$.

Most earlier works assume i.i.d. normality.[4] Circumstances occur where this assumption may fail. For example, it is well-known that stock price innovations often have highly leptokurtic distributions. Although Ljung and Box (1978, Section 4.4) conjecture that BP and LB are robust to nonnormality and support this by simulation, no formal justification was available in the literature. The fourth moment condition helps ensure asymptotic normality of the M_{jn}. While our tests are consistent against H_A, we note that there is a gap between A.2 and H_0. The use of A.2 simplifies much of the asymptotic analysis because it involves higher moments of spectral estimates. An earlier version of this paper imposes the martingale condition under H_0, but with rather restrictive moment conditions.

[3] Under suitable additional conditions on k, we can also have more compact expressions $M_{2n}^* = (2nH^2(\hat{f}_n; f_0) - p_n C(k))/(2p_n D(k))^{1/2}$ and $M_{3n}^* = (nI(\hat{f}_n; f_0) - p_n C(k))/(2p_n D(k))^{1/2}$.

[4] Robinson (1991b) proposes a general class of tests for serial correlation, assuming only the martingale difference sequences for $\{u_t\}$ under H_0.

Assumption A.3 includes such estimators as OLS, adaptive asymptotically efficient weighted least squares, and maximum likelihood. It also includes estimators that are $n^{1/2}$-consistent but not asymptotically normal, as may arise if $\{u_t\}$ exhibits strong dependence.

We first establish asymptotic normality of M_{1n}.

THEOREM 1: *Suppose Assumptions A.1(a), A.2–A.3 hold. Let* $p_n \to \infty$, $p_n/n \to 0$. *Then*

$$M_{1n} \to^d N(0,1).$$

The estimator $\hat{\alpha}$ (i.e. the use of $\hat{u}_t$ in place of u_t) has no impact on the limiting distribution of M_{1n}. This is true even when the regressors include lagged dependent variables. This conclusion is in sharp contrast to some conventional tests, such as those of BP, LB, and Durbin and Watson (1950, 1951). Intuitively, the use of $\hat{u}_t$ may induce some finite adjustments for degrees of freedom when (1) includes lagged dependent variables, but these finite adjustments become negligible for M_{1n} as p_n becomes large.

When the truncated kernel is used, we obtain M_{1n}^T. Comparing between M_{1n}^T and BP, we have extended Box and Pierce's results in some directions. First, M_{1n}^T applies to the residual from (1), whereas BP only applies to ARM A models of finite orders. As Breusch and Pagan (1980, p. 245) point out, BP is inappropriate when the regressors include both lagged dependent variables and exogenous variables. Second, the limiting distribution of M_{1n}^T remains invariant whether lagged dependent variables are present, while BP has to adjust degrees of freedom according to the number of regressors. Finally, the null distribution of BP is derived under the normality assumption, while we do not impose it.

Next, we establish the asymptotic equivalence between M_{1n}^T and M_R under H_0.

THEOREM 2: *Suppose Assumptions A.1(a), A.2–A.3 hold. Let* $p_n \to \infty$, $p_n^5/n \to 0$. *Then*

$$M_R - M_{1n}^T = o_P(1), \quad \textit{and} \quad M_R \to^d N(0,1).$$

The conditions on p_n are much more restrictive for M_R than for M_{1n}, but it includes the logarithm rates delivered by such information criteria as Akaike's Information Criterion and the Bayesian Information Criterion for (5) (cf. Ng and Perron (1994)).

Finally, we establish asymptotic equivalence among the M_{jn} under H_0.

THEOREM 3: *Suppose Assumptions A.1–A.3 hold. Let* $p_n \to \infty$, $p_n^3/n \to 0$. *Then*

$$M_{2n} - M_{1n} = o_P(1), \quad M_{3n} - M_{1n} = o_P(1), \quad M_{2n} \to^d N(0,1), \quad \textit{and} \quad M_{3n} \to^d N(0,1).$$

4. ASYMPTOTIC LOCAL POWER

For consistent tests, the asymptotic power approaches unity as $n \to \infty$ under H_A at any given level $0 < \zeta < 1$. To get a power value less than unity, we can either fix the size and move the alternative hypothesis closer to H_0 as $n \to \infty$, or fix the alternative and let the size or Type II error decrease to zero as $n \to \infty$. The first approach is the familiar Pitman's local analysis, and the second is the nonlocal analysis (e.g. Bahadur (1960), Hodges and Lehmann (1956, Section 3)). They give different insights and conclusions for the M_{jn}.

We first focus on local power analysis. For simplicity, we maintain Assumption A.2 and consider a sequence of specified models $\{f_n^0\}$ such that $f_n^0 \to f_0$ as $n \to \infty$. This leads to much simpler analysis and delivers conclusions identical to those that would be reached by fixing the model and moving the data generating process. Specifically, we consider

$$H_{an}: f_n^0(\omega) = f_0(\omega) + a_n g(\omega), \quad \omega \in [-\pi, \pi],$$

where $a_n \to 0$ as $n \to \infty$, and $g: R \to R$ is a symmetric periodic (with periodicity 2π) bounded continuous function with $\int_{-\pi}^{\pi} g(\omega)\, d\omega = 0$. The condition $\int_{-\pi}^{\pi} g(\omega)\, d\omega = 0$ ensures that f_n^0 is a normalized spectral density for all n sufficiently large.

The next theorem delivers the rate a_n at which our tests have nontrivial power.

THEOREM 4: *Suppose Assumptions A.1(a) and A.2–A.3 hold. Define*

$$M_{1n}^a = \left(\tfrac{1}{2} n Q^2(\hat{f}_n; f_n^0) - C_n(k)\right) \Big/ (2D_n(k))^{1/2},$$

$$M_{2n}^a = \left(2nH^2(\hat{f}_n; f_n^0) - C_n(k)\right) \Big/ (2D_n(k))^{1/2},$$

$$M_{3n}^a = \left(nI(\hat{f}_n; f_n^0) - C_n(k)\right) \Big/ (2D_n(k))^{1/2},$$

where f_n^0 is as in H_{an} with $a_n = p_n^{1/4}/n^{1/2}$. Let $p_n \to \infty$, $p_n/n \to 0$. Then $M_{1n} \to^d N(\mu(k), 1)$, where $\mu(k) = 2\pi \int_{-\pi}^{\pi} g^2(\omega)\, d\omega / (2D(k))^{1/2}$. If in addition Assumption A.1(b) holds and $p_n^3/n \to 0$. Then $M_{2n}^a - M_{1n}^a = o_P(1)$, $M_{3n}^a - M_{1n}^a = o_P(1)$, $M_{2n}^a \to^d N(\mu(k), 1)$, and $M_{3n}^a \to^d N(\mu(k), 1)$.

The asymptotic power of the M_{jn}^a is $\lim_{n \to \infty} \Pr(M_{jn}^a > C_\zeta) = 1 - \Phi(C_\zeta - \mu(k))$, where Φ is the cdf of $N(0,1)$ and C_ζ is the upper-tailed critical value at level ζ. Therefore, the M_{jn}^a are asymptotically equally efficient under H_{an} with $a_n = p_n^{1/4}/n^{1/2}$. Asymptotically, there is no optimal choice of p_n: the slower p_n grows, the more powerful are the M_{jn}^a, because H_{an} is farther away from H_0.[5] Because a_n is slower than the parametric rate $n^{-1/2}$, the M_{jn}^a are less efficient than

[5] If a higher order approximation to the distribution is made, we expect that there will be a trade-off between size and power in choosing p_n. Thus, one can choose an optimal p_n to minimize both Type I and Type II errors using a suitable criterion.

parametric tests under H_{an}. Of course, this should not be taken too literally. For example, if $p_n = n^{1/5}$, $a_n = n^{-1/2+1/20}$ is only slightly slower than $n^{-1/2}$.

Because $\mu(k)$ depends on k, different kernels may deliver different power. Suppose $p_n = cn^{\nu}$, $0 < c < \infty$, $0 < \nu < 1/3$. Then, following an analogous reasoning of Pitman (1979), we obtain that for any M_{jn} that uses k_1 and k_2 respectively, the relative asymptotic efficiency of k_2 with respect to k_1 is

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

Therefore, the relative efficiency of the Bartlett kernel (k_B) to the truncated kernel (k_T) is $\text{ARE}_P(k_B; k_T) = 5^{1/(2-\nu)} > 2.23$.[6] In fact, other commonly used kernels are all more powerful than the truncated kernel.

We now derive the optimal kernel that maximizes the power of our tests over some proper class of kernel functions. Let r be the largest integer such that

$$k^{(r)} = \lim_{z \to 0} (1 - k(z))/|z|^r$$

exists, and is finite and nonzero. We consider a class of kernels with $r = 2$:

$$\kappa(\tau) = \{k(\cdot) \text{ satisfies Assumptions A.1 with } k^{(2)} = \tau^2/2 > 0\}.$$

The condition $k^{(2)} = \tau^2/2$ plays a role of normalization. For any $k \in \kappa(\tau)$, the bias $E\hat{f}_n(\omega) - f(\omega) = p_n^{-2} k^{(2)} f^{(2)}(\omega)(1 + o(1))$ under proper conditions (i.e. f is twice continuously differentiable on $[-\pi, \pi]$). Thus, with the same p_n, any two kernels in $\kappa(\tau)$ will deliver estimates with the same asymptotic bias. This normalization excludes some meaningless comparisons. Otherwise, two kernels that have the same shape but are scaled differently will deliver different powers when the same p_n is used. The class $\kappa(\tau)$ includes the Daniell, Parzen, and QS kernels, but rules out the truncated, Bartlett, and general Tukey kernels.

THEOREM 5: *Suppose conditions of Theorem* 4 *hold, and the* M_{jn}^a *are defined as in Theorem* 4. *Then under* H_{an} *with* $a_n = p_n^{1/4}/n^{1/2}$, *the Daniell kernel* $k_D(z) = \sin(\sqrt{3}\,\tau z)/(\sqrt{3}\,\tau z)$, $z \in (-\infty, \infty)$, *maximizes the power of the* M_{jn}^a *over* $\kappa(\tau)$.

The Daniell kernel is different from the QS kernel, which is optimal within $\kappa(\tau)$ in the context of spectral density estimation using various mean squared error criteria (e.g. Andrews (1991) and Priestley (1962)). For hypothesis testing, the QS kernel can be worse than many other kernels. This conclusion is not peculiar to the M_{jn}^a considered here (e.g. Hong (1996)). While Theorem 5 is of some theoretical interest, some commonly used kernels have rather close values of $D(k)$. For example, the Daniell, Parzen, and QS kernels have $D(k) = 0.6046/\tau$, $0.6627/\tau$, and $0.6094/\tau$ respectively. Therefore, we expect little difference in power among these kernels.

[6] The Bartlett kernel $k_B(z) = (1 - |z|)1[|z| \leq 1]$, and the truncated kernel $k_T(z) = 1[|z| \leq 1]$.

5. ASYMPTOTIC GLOBAL POWER

While local power analysis provides useful insights, it is by no means a complete account of asymptotic power properties. In particular, Theorem 4 implies that the M^a_{jn} are asymptotically equivalent under H_{an}, so it is difficult to differentiate these tests using Pitman's criterion. In practice, however, these tests may lead to different decisions. To examine their relative efficiencies under H_A, we have to use nonlocal power analysis.

We first establish consistency of our tests under H_A. Under H_A, $\hat{\alpha}$ is generally not consistent for α in (1). Consequently, the residual $\{\hat{u}_t\}$ is biased for $\{u_t\}$, because it contains specification errors as well as $\{u_t\}$. This would complicate the analysis in establishing consistency of our tests. For simplicity, we consider the static regression model

$$Y_t = c + \alpha^{(1)}(B)X_{1t} + \cdots + \alpha^{(q)}(B)X_{qt} + u_t \qquad (t = 1, 2, \ldots, n). \tag{8}$$

Now α is a $(q + 1 + \sum_{j=1}^{q} m_j) \times 1$ vector consisting of all the unknown coefficients in (8). The OLS estimator $\hat{\alpha}$ is consistent for α under H_A, although inefficient. Hence, $\{\hat{u}_t\}$ is consistent for $\{u_t\}$. The following assumptions describe the dependence structure of $\{u_t\}$.

ASSUMPTION A.4: *$\{u_t\}$ is a mean zero fourth order stationary process with $\sum_{j=-\infty}^{\infty} R^2(j) < \infty$ and $\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} |\kappa_4(i,j,l)| < \infty$, where $\kappa_4(i,j,l)$ is the fourth order joint cumulant of the distribution of $\{u_t, u_{t+i}, u_{t+j}, u_{t+l}\}$.*

ASSUMPTION A.5: *(a) There exists some $\delta > 0$ such that $f(\omega) \geq \delta$ for all $\omega \in [-\pi, \pi]$; (b) $f(\omega)$ is continuous on $[-\pi, \pi]$.*

Assumption A.4 allows for fractionally integrated processes $I(d)$ with $d < 1/4$, thus including some long memory processes. The fourth order joint cumulant $\kappa_4(i,j,l)$ is defined as

$$\kappa_4(i,j,l) = E(u_t u_{t+i} u_{t+j} u_{t+l}) - E\left(\tilde{u}_t \tilde{u}_{t+i} \tilde{u}_{t+j} \tilde{u}_{t+l}\right),$$

where $\{\tilde{u}_t\}$ is a Gaussian sequence with the same mean and covariance as $\{u_t\}$. The cumulant condition is a standard assumption in time series (e.g. Anderson (1971) and Hannan (1970)); it characterizes the temporal dependence of $\{u_t\}$. When $\{u_t\}$ is a Gaussian process, the cumulant condition holds trivially because $\kappa_4(i,j,l) = 0$ for all i, j, l. If $\{u_t\}$ is a fourth order stationary linear process with absolute summable coefficients and innovations whose fourth moments exist, the cumulant condition also holds (e.g., Hannan (1970, p. 211)). More primitive conditions (e.g. strong mixing) can be imposed to ensure the cumulant condition, but such primitive conditions will rule out strongly dependent processes.[7]

[7] I thank Professor P. M. Robinson and one referee for pointing out this.

Assumption A.5(a), used for M_{2n} and M_{3n}, allows for spectral densities that are infinity at $\omega = 0$, as with "fractional noise" and fractionally differenced ARIMA processes, the two most popular long memory models (cf. Robinson (1991b, Section 5)). Assumption A.5(b) helps ensure consistency of M_{3n}; it rules out long memory processes.

THEOREM 6: *Suppose A.1(a), A.3–A.4 and H_A hold. Let $p_n \to \infty$, $p_n/n \to 0$. Then*

$$(p_n^{1/2}/n)M_{1n} \to^p \tfrac{1}{2}Q^2(f;f_0)/(2D(k))^{1/2}.$$

Suppose in addition A.1(b) and A.5(a) hold. Then

$$(p_n^{1/2}/n)M_{2n} \to^p 2H^2(f;f_0)/(2D(k))^{1/2}.$$

If furthermore A.5(b) holds and $p_n^2/n \to 0$, then

$$(p_n^{1/2}/n)M_{3n} \to^p I(f;f_0)/(2D(k))^{1/2}.$$

Thus, the M_{jn} are all consistent against H_A; in particular, M_{1n} and M_{2n} are consistent against long memory processes $I(d)$, $d < 1/4$. We note that Robinson (1993) proposes an F-test that is also consistent against long memory processes and has many other appealing features, especially having an exact null distribution under the Gaussian case.

The rate at which the M_{jn} diverge to infinity is $n/p_n^{1/2}$. The slower p_n grows, the faster will the M_{jn} diverge to infinity, and so the more powerful will be the M_{jn}. This conclusion on p_n is the same as that reached under H_{an}. We now use Bahadur's (1960) asymptotic slope criterion to investigate relative efficiencies among the M_{jn}. The basic idea of Bahadur is to hold the power fixed and compare the resulting test sizes. Bahadur's efficiency is defined as the limit ratio of the sample sizes required by two tests to achieve the same asymptotic significance level (p-value) under a fixed alternative. Geweke (1981), among others, has applied this criterion in the econometric literature. Extending Bahadur's (1960) approach, we obtain the following result.

THEOREM 7: *Suppose Assumptions A.1, A.3–A.5 hold. Let $p_n = cn^{\nu}$, where $0 < c < \infty$, $0 < \nu < 1/2$. Then Bahadur's asymptotic relative efficiencies among the M_{jn} under H_A are*

$$\mathrm{ARE}_B(M_{1n};M_{2n}) = \left[Q^2(f;f_0)/4H^2(f;f_0)\right]^{1/(2-\nu)},$$

$$\mathrm{ARE}_B(M_{2n};M_{3n}) = \left[2H^2(f;f_0)/I(f;f_0)\right]^{1/(2-\nu)},$$

$$\mathrm{ARE}_B(M_{3n};M_{1n}) = \left[2I(f;f_0)/Q^2(f;f_0)\right]^{1/(2-\nu)}.$$

When $f \to f_0$, the three Bahadur's efficiencies converge to unity, delivering the same conclusion as Pitman's criterion. When f is far away from f_0, however, the three divergence measures are not equal in general. It would be interesting to characterize the interrelationships and inequalities among the three divergence measures so that relative power ranking is possible. We expect that such characterizations will depend on dependence patterns of $\{u_t\}$. Studies (e.g. Ullah (1993) and references therein) on interrelationships and inequalities on various divergence measures between probability distributions seem useful here.

Bahadur's (1960) asymptotic slope of a test statistic is the rate at which minus twice the logarithm of the asymptotic significance level goes to infinity as $n \to \infty$. A larger asymptotic slope implies a faster rate at which the asymptotic significance level decreases to zero as $n \to \infty$. It can be shown that the rate at which minus twice the logarithm of the asymptotic significance level of the M_{jn} goes to infinity is n^2/p_n. This rate is faster than the rate n of parametric tests (including both asymptotic normal and χ^2 tests; see Bahadur (1960)). Therefore, the M_{jn} have an infinitely larger asymptotic slope than parametric tests under H_A, i.e. the asymptotic significance level of the M_{jn} decreases to zero faster than that of parametric tests. This conclusion about the relative efficiency between the M_{jn} and parametric tests under H_A is in sharp contrast to that reached under H_{an}.

It can be shown that for any M_{jn} Bahadur's relative efficiency of k_2 to k_1 is $\text{ARE}_B(k_2; k_1) = [D(k_1)/D(k_2)]^{1/(2-\nu)}$ under the conditions of Theorem 7. This is the same as that obtained for the Pitman's criterion. Hence, the discussions on the kernel in Section 4 apply here.

6. MONTE CARLO EVIDENCE

We now examine finite sample performance of our tests in comparison to some commonly used tests for serial correlation. Consider the data generating process

$$Y_t = c + \alpha_1 Y_{t-1} + \alpha_2 X_t + u_t, \tag{9}$$

where the exogenous variable $X_t = 0.8X_{t-1} + v_t$, and the v_t are NID(0,3). We set $\alpha = (c, \alpha_1, \alpha_2)' = (1, 0.5, 0.5)'$. Let $\{\epsilon_t\}$ be NID(0,1) and $\{e_t\}$ be UID(0,1). We consider four processes: (a) $u_t = \epsilon_t$; (b) $u_t = e_t$; (c) $u_t = 0.3u_{t-1} + \epsilon_t$; (d) $(1-B)^{0.35}u_t = \epsilon_t$. Both (a) and (b) permit us to examine size performances under normal and non-normal (uniform) white noise errors. Process (c) is the widely used AR(1); and (d) is the fractionally differenced ARIMA(0, 0.35, 0) process, whose autocorrelation decays at a hyperbolic rate as the lag increases. This long memory process may arise from aggregation of time series data and has become more and more popular in time series modeling. We generate (d) using Davies and Harte's (1987) algorithm. Two sample sizes $n = 64$, 128 are investigated. For each n, we set the initial value $Y_0 = 0$ and generate $2n + 1$ observations using (9), with the first $n + 1$ observations being discarded to reduce the effects of the initial value. The simulation is conducted using the GAUSS random number generator on an IBM RISCsystem Workstation, Model 570, at Cornell University.

We compare our tests with those of BP, LB, and Breusch (1978) and Godfrey (1978). Both M_{2n} and M_{3n} are computed using the integration procedure INTEQUAD1 in the GAUSS program, with the order of the integration being set equal to 40. To examine the effects of using different kernels, we choose four kernels for the M_{jn}:

$$\text{Daniell (DAN):}\quad k(z)=\sin(\pi z)/\pi z,\qquad z\in(\infty,\infty);$$

$$\text{Parzen (PAR):}\quad k(z)=\begin{cases}1-6(\pi z/6)^2+6|\pi z/6|^3, & |z|\le 3/\pi,\\ 2(1-|\pi z/6|)^3, & 3/\pi\le|z|\le 6/\pi,\\ 0, & \text{otherwise;}\end{cases}$$

$$\text{QS:}\quad k(z)=(9/5z^2)\left\{\sin\left(\sqrt{5/3}\,\pi z\right)/\sqrt{5/3}\,\pi z-\cos\left(\sqrt{5/3}\,\pi z\right)\right\},$$

$$z\in(-\infty,\infty);$$

$$\text{Bartlett (BAR):}\quad k(z)=\begin{cases}1-|z|, & |z|\le 1,\\ 0, & \text{otherwise.}\end{cases}$$

Here, DAN, PAR, and QS belong to $\kappa(\pi/\sqrt{3})$. BAR does not belong to $\kappa(\tau)$.

To examine the effects of using different p_n, we first use three rates: (i) $p_n=[\ln(n)]$; (ii) $p_n=[3n^{0.2}]$, and (iii) $p_n=[3n^{0.3}]$, where $[a]$ denotes the integer closest to a. These rates deliver $p=4,7,10$ for $n=64$ and $5,8,13$ for $n=128$. The $\ln(n)$ rate, up to some proportionality, is the rate delivered by information-based criteria for (5). The rate $n^{0.2}$, up to some proportionality, is the optimal rate minimizing the mean squared error of $\hat{f}_n$ when the kernel with $r=2$ is used; and the rate $n^{0.3}$ is close to the upper bound on p_n for M_{2n} and M_{3n}. Both (ii) and (iii) violate the conditions of Theorem 2 (for M_R), but we include them to examine the performance of M_R with these rates.

The above three deterministic rules allow us to investigate the effects of choices of p_n, but they are to some degree unmotivated. In practice, it would be desirable to choose p_n via data-driven methods. Beltrao and Bloomfield (1987) propose a form of cross-validation based on a pseudo log-likelihood type criterion under the Gaussian case. In an important paper, Robinson (1991a) considerably extends their results to non-Gaussian situations, showing that such chosen p_n is consistent for an optimal integrated mean squared error bandwidth. Such a global bandwidth is more appropriate here than the narrow band ones stressed in the econometric literature in autocorrelation-consistent variance estimation (e.g. Andrews (1991)). The procedure can be conveniently implemented using fast Fourier transforms. In our application, we use a grid search for the optimal integer-valued p_n over the range from 2 to 20, with the grid interval equal to 1.[8] It is possible to choose a real-valued p_n with a finer grid interval, but this is likely to have negligible impact.

[8] In our simulation we find that the lower bound "2" (the smallest integer that ensures $\hat{f}_n\neq f_0$) works well for the samples under study. Robinson (1991a, p. 1346) points out that as n increases the cross-validation will tend to choose the per-set lower bound as the optimal p_n, when u_t is a white noise. Therefore, for large n, one might use a lower bound that is slowly increasing as n increases.

TABLE I

REJECTION RATES IN PERCENTAGE UNDER NORMAL WHITE NOISES

n:		64								128							
p_n:		4		7		10		CV		5		8		13		CV	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
M_{1n}	DAN	6.4	4.3	7.5	4.8	8.0	5.1	9.0	6.1	6.0	4.0	7.5	4.7	8.2	5.2	9.0	5.7
	PAR	6.6	4.2	7.6	4.9	8.2	5.2	9.6	6.3	6.4	4.2	7.7	4.7	8.3	5.4	9.6	6.0
	QS	6.5	4.2	7.5	4.8	8.0	5.1	9.7	6.6	6.1	4.1	7.4	4.7	8.0	5.3	9.7	6.1
	BAR	6.2	4.0	6.9	4.5	7.6	5.0	9.9	6.6	5.7	3.8	7.0	4.4	7.6	5.0	9.7	6.4
M_{2n}	DAN	7.0	4.9	8.3	5.2	8.8	5.3	10.1	6.6	6.6	4.0	7.9	4.9	8.2	5.3	9.2	5.9
	PAR	7.0	4.8	8.0	5.0	8.1	4.6	10.2	6.7	6.7	4.2	7.7	4.7	8.1	4.8	9.7	6.1
	QS	6.9	4.8	8.2	5.1	8.5	5.1	10.6	7.1	6.4	3.9	7.5	4.8	8.3	5.1	9.9	6.3
	BAR	6.6	4.5	7.1	4.5	7.3	4.6	10.0	6.4	5.9	3.8	6.8	4.2	7.3	4.5	9.8	6.1
M_{3n}	DAN	7.7	5.2	9.6	6.5	10.8	7.3	11.4	7.8	6.9	4.5	8.9	5.5	9.6	6.2	10.0	6.9
	PAR	7.5	5.1	9.2	6.0	9.8	6.4	11.3	7.7	7.2	4.4	8.3	5.3	8.9	6.0	10.5	6.7
	QS	7.5	5.1	9.3	6.3	10.0	6.7	12.0	8.3	6.9	4.2	8.3	5.2	9.1	6.1	10.8	7.0
	BAR	6.9	4.7	7.8	5.1	8.3	5.1	10.7	7.0	6.0	3.7	7.2	4.5	7.9	4.8	10.1	6.5
M_{1n}^T		7.2	4.4	6.6	4.1	6.4	3.9			7.7	4.9	7.6	4.6	7.2	4.4		
M_R		8.1	5.1	6.8	4.1	5.6	3.0			7.9	5.1	7.9	5.0	7.0	3.5		
BP		13.6	6.6	10.5	5.2	8.8	4.4			12.7	6.6	10.8	5.6	9.4	4.9		
LB		16.5	8.5	14.4	7.6	14.1	8.2			14.2	7.5	13.0	6.9	13.0	7.1		
LM		11.1	5.5	9.6	4.3	8.0	3.6			9.8	4.8	9.2	4.4	8.0	3.5		

Notes: (i) Model $Y_t = 1.0 + 0.5Y_{t-1} + 0.5X_t + u_t$, where $X_t = 0.8X_{t-1} + v_t$, $v_t = \text{NID}(0,3)$ and $u_t = \text{NID}(0,1)$.
(ii) 5000 Replications.
(iii) CV = Cross-Validation.
(iv) DAN, PAR, QS, BAR = Daniell, Parzen, Quadratic-Spectral, and Bartlett kernels.

For comparison, we use the same deterministic rules of p_n for BP, LB, and LM tests, where $\text{BP} = n\sum_{j=1}^{p_n} \hat{\rho}^2(j)$ and $\text{LB} = n(n+2)\sum_{j=1}^{p_n}(n-j)^{-1}\hat{\rho}^2(j)$. Both BP and LB are inappropriate for (9), but we still treat BP and LB as asymptotically $\chi^2_{p_n-1}$ under H_0.[9] The LM test is computed as $\text{LM} = nR^2$, where R^2 is obtained from the OLS regression of $\hat{u}_t$ on $1, Y_{t-1}, X_t, \hat{u}_{t-1}, \dots, \hat{u}_{t-p_n}$. Under H_0, LM is asymptotically $\chi^2_{p_n}$. (Strictly speaking, the asymptotic χ^2 of the LM tests is valid only when p_n is fixed.)

Table I reports rejection rates (in percentage) under normal white noise errors at 10% and 5% nominal levels, based on 5000 replications. The new tests have reasonable sizes at the 5% level, but have greater difficulties of getting it right at the 10% level for small p_n. Faster p_n gives better size. The cross-validation works reasonably well, especially at the 10% level. For each M_{jn}, DAN, PAR, and QS perform similarly, but BAR performs slightly differently. Among the M_{jn}, M_{3n} exhibits a little overrejection at the 5% level for large and cross-validated p_n, while M_{1n} and M_{2n} perform similarly. Both M_{1n}^T and M_R

[9] We also use empirical critical values for BP and LB, so a power comparison based on these empirical critical values is appropriate.

have similar sizes. Unlike the M_{jn}, the rejection rates of M_{1n}^T and M_R decrease as p_n increases; they have best sizes when p_n is small. LM works reasonably well. In contrast, LB exhibits strong overrejection for all p_n. The rejection rates of BP decrease as p_n increases, and has better sizes than LB. These findings differ from those found in the earlier literature.

Table II reports sizes under non-normal (uniform) white noise errors. The results show that the performances of all the tests are similar to those under normal errors.

We now examine power using both asymptotic critical values (ACV) and empirical critical values (ECV) at the 5% level. The ECV are obtained from the 5000 replications under (a). The use of ECV permits us to compare power of all the tests on an equal basis. Table III reports the number of rejections out of 1000 replications under the AR(1) alternative. For each pair k and p_n, the M_{jn} have roughly the same power. For each M_{jn}, smaller p_n gives better power. The cross-validation gives better power than the deterministic rules in terms of ACV, and its ECV-based power is good. Given each p_n, DAN, PAR, and QS perform similarly for each M_{jn}, with DAN and QS very slightly more powerful than PAR. For deterministic p_n, BAR has slightly better power than DAN, PAR, and QS in most cases, but this is not inconsistent with Theorem 5, because Theorem 5

TABLE II

REJECTION RATES IN PERCENTAGE UNDER NONNORMAL (UNIFORM) WHITE NOISES

n:		64								128							
p_n:		4		7		10		CV		5		8		13		CV	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
M_{1n}	DAN	6.1	3.9	7.9	5.0	8.9	5.2	8.8	5.4	6.2	4.1	7.2	4.5	9.1	6.0	9.4	6.2
	PAR	6.3	3.8	8.1	5.0	8.8	5.3	9.2	5.9	6.2	4.4	7.5	4.9	9.3	6.1	9.8	6.9
	QS	6.0	3.8	7.9	4.7	8.7	5.1	9.5	5.9	6.1	4.2	7.2	4.5	8.9	5.8	10.2	7.0
	BAR	5.9	3.7	7.4	4.4	8.2	4.8	9.7	6.2	6.1	4.1	6.6	4.4	8.3	5.3	10.2	6.7
M_{2n}	DAN	6.9	4.4	8.7	5.5	9.1	5.7	9.9	6.2	6.5	4.4	7.2	4.8	9.1	6.0	9.9	6.7
	PAR	6.9	4.3	8.5	5.2	8.6	5.2	9.9	6.1	6.4	4.3	7.4	4.6	8.5	5.4	9.9	6.8
	QS	6.6	4.3	8.4	5.3	9.0	5.5	10.4	6.5	6.4	4.4	7.4	4.7	8.5	5.5	10.7	7.1
	BAR	6.2	3.9	7.2	4.4	8.1	4.6	9.8	5.8	6.2	4.2	6.5	4.3	7.3	4.7	10.7	6.6
M_{3n}	DAN	7.4	5.0	10.0	6.9	10.9	7.5	11.0	7.6	7.5	5.0	8.9	5.4	10.7	7.2	11.0	7.4
	PAR	7.4	4.8	9.4	6.2	10.2	6.7	10.7	7.2	6.8	4.5	8.0	5.1	9.6	6.3	10.5	7.4
	QS	7.2	4.7	9.6	6.4	10.5	7.1	11.4	7.6	6.7	4.6	7.8	5.0	9.8	6.3	11.0	7.7
	BAR	6.5	4.2	8.0	4.9	8.8	5.4	10.4	6.4	6.3	4.3	6.8	4.5	7.8	5.2	10.4	6.9
M_{1n}^T		7.6	4.8	7.2	4.6	6.4	4.0			8.6	5.3	8.9	5.7	8.4	5.0		
M_R		8.9	5.6	7.8	4.0	6.0	3.1			8.5	5.9	9.0	5.5	7.4	3.8		
BP		13.7	7.0	10.8	5.4	8.7	4.5			13.2	7.1	12.5	6.6	10.7	5.3		
LB		16.5	8.9	14.8	8.3	14.2	8.0			15.0	8.4	14.9	8.2	14.4	8.2		
LM		11.9	6.4	10.2	4.7	8.4	3.6			10.3	5.5	10.6	5.3	8.4	3.7		

Notes: (i) Model $Y_t = 1.0 + 0.5Y_{t-1} + 0.5X_t + u_t$, $X_t = 0.8X_{t-1} + v_t$, $v_t = \text{NID}(0,3)$ and $u_t = \text{UID}(0,1)$.
(ii) 5000 Replications.
(iii) CV = Cross-Validation.
(iv) DAN, PAR, QS, BAR = Daniell, Parzen, Quadratic-Spectral, and Bartlett kernels.

TABLE III

NUMBER OF REJECTIONS UNDER AR(1) ALTERNATIVE USING ASYMPTOTIC AND EMPIRICAL CRITICAL VALUES AT THE 5% LEVEL

n:		64								128							
p_n:		4		7		10		CV		5		8		13		CV	
		ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV
M_{1n}	DAN	336	361	277	293	247	249	360	328	674	718	601	608	521	521	713	699
	PAR	323	352	270	280	240	238	378	339	654	683	584	593	501	496	738	716
	QS	335	359	279	296	242	242	383	339	670	712	598	607	516	516	742	719
	BAR	340	372	300	319	275	284	384	338	687	742	635	659	578	582	738	705
M_{2n}	DAN	358	363	293	287	256	249	372	331	688	726	614	620	523	516	727	709
	PAR	343	357	277	280	235	251	386	348	668	698	596	608	503	514	747	725
	QS	354	363	291	291	251	251	393	341	687	722	614	628	524	525	745	723
	BAR	351	372	306	314	275	291	380	346	703	745	643	665	579	606	738	718
M_{3n}	DAN	370	365	311	282	286	244	390	331	704	725	627	613	547	509	735	698
	PAR	355	356	295	274	267	246	398	340	676	695	608	606	530	511	753	718
	QS	354	363	291	291	251	251	405	339	692	721	627	623	544	518	750	716
	BAR	351	372	306	314	275	291	388	332	706	743	651	678	590	606	745	714
M_{1n}^T		181	205	164	180	151	170			440	450	379	396	288	315		
M_R		203	203	152	179	106	151			468	468	337	346	218	262		
BP		236	205	181	180	156	170			501	450	402	396	308	313		
LB		274	204	234	175	216	160			525	441	433	385	357	301		
LM		232	224	173	193	122	160			503	518	399	416	272	332		

Note: (i) Model $Y_t = 1.0 + 0.5Y_{t-1} + 0.5X_t + u_t$, $X_t = 0.8X_{t-1} + v_t$, $v_t = \text{NID}(0,3)$, and $u_t = 0.3u_{t-1} + \varepsilon_t$, $\varepsilon_t = \text{NID}(0,1)$.
(ii) 1000 replications.
(iii) ACV = asymptotic critical value, ECV = empirical critical value.
(iv) CV = cross-validation.
(v) DAN, PAR, QS, BAR = Daniell, Parzen, Quadratic-Spectral and Bartlett kernels.

rules out BAR. As expected, the M_{jn} are more powerful than LM, M_{1n}^T, BP, M_R and LB. The latter five tests have similar power against the AR(1) alternative.

Table IV reports power against the fractionally differenced ARIMA(0, 0.35, 0) process. Given each pair k and p_n, the M_{jn} have similar power, although there seems to be rather weak evidence that M_{1n} has very slightly better power than M_{2n}, which in turn has very slightly better power than M_{3n}. For all M_{jn}, a slower p_n gives better power. Again, the cross-validation delivers better power than deterministic rules in terms of ACV, and its ECV-based power is also good. DAN, PAR, and QS have similar power. For deterministic p_n, BAR has very slightly better power. The M_{jn} are more powerful than M_{1n}^T, BP, LB, LM, and M_R.

To summarize, (i) for the new tests, the choice of kernels (other than the truncated kernel) has little impact on size and power; the truncated kernel, which delivers a generalized BP test, has lower power than other kernels. (ii) The choice of p_n has relatively significant effects on size and power. A faster p_n gives better size, while a slower p_n gives better power. The cross-validation procedure of Beltrao and Bloomfield (1987) and Robinson (1991a) works rea-

TABLE IV

NUMBER OF REJECTIONS UNDER FRACTIONAL ARIMA(0, d, 0) ALTERNATIVE USING ASYMPTOTIC AND EMPIRICAL CRITICAL VALUES AT THE 5% LEVEL

n:		64								128							
p_n:		4		7		10		CV		5		8		13		CV	
		ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV	ACV	ECV
M_{1n}	DAN	448	471	391	412	362	363	452	431	844	869	823	831	780	780	861	855
	PAR	441	465	391	396	353	351	472	440	839	855	815	824	770	770	864	857
	QS	448	472	397	411	360	360	473	441	844	863	821	828	778	778	870	856
	BAR	449	483	411	435	396	399	466	430	848	881	836	851	808	810	869	852
M_{2n}	DAN	432	439	379	371	344	339	445	414	837	848	796	798	726	723	848	835
	PAR	421	433	365	365	324	341	464	429	830	846	786	793	711	717	857	847
	QS	429	441	377	377	343	343	468	422	834	851	792	799	724	724	863	842
	BAR	435	462	393	408	358	375	449	418	838	868	811	831	772	787	858	843
M_{3n}	DAN	441	435	391	364	352	323	457	403	832	847	783	776	720	690	844	812
	PAR	425	426	371	357	341	313	472	411	826	837	781	778	700	684	854	840
	QS	431	430	385	365	355	324	478	414	830	846	792	791	720	701	861	835
	BAR	434	454	396	399	363	363	457	405	839	864	803	822	768	775	857	835
M_{1n}^T		311	331	254	278	213	247			738	744	672	686	586	615		
M_R		275	275	184	211	122	189			656	656	552	565	388	434		
BP		367	331	279	278	231	247			770	743	691	685	606	615		
LB		398	326	332	265	291	230			781	736	713	676	658	598		
LM		315	307	238	263	177	228			730	739	649	660	530	573		

Notes: (i) Model: $Y_t = 1.0 + 0.5Y_{t-1} + 0.5X_t + u_t$, $X_t = 0.8X_{t-1} + v_t$, $v_t = \text{NID}(0,3)$, and $u_t = (1-B)^{-0.35}\varepsilon_t$, $\varepsilon_t = \text{NID}(0,1)$.
(ii) 1000 replications.
(iii) ACV = asymptotic critical value; ECV = empirical critical value.
(iv) CV = cross-validation.
(v) DAN, PAR, QS, BAR = Daniell, Parzen, Quadratic-Spectral, and Bartlett kernels.

sonably well in balancing Type I and II errors and is therefore recommended for use in practice. (iii) Among the new tests, M_{1n} and M_{2n} have relatively robust sizes, while M_{3n} exhibits a little bit of overrejection in some cases. The M_{jn} have similar power under each of the two alternatives. (iv) The M_{jn} have better power than LM, BP, LB, M_{1n}^T, and M_R against both alternatives.

7. CONCLUSION

This paper proposes three classes of consistent tests for serial correlation of unknown form for the residuals from a linear dynamic regression model. The tests are based on comparison between a kernel-based spectral density estimator with the null spectral density, using a quadratic norm, Hellinger metric, and Kullback-Leibler information criterion respectively. The three classes of tests are asymptotically equivalent under a class of local alternatives, but their relative efficiencies under the global alternatives depend on the relative magnitudes of the divergence measures. The asymptotic distributions of our tests are all standard normal, and remain invariant when the regressors include lagged dependent variables. Many kernels are more efficient than the truncated kernel, the latter delivering Box and Pierce's (1970) type tests. A simulation study shows that the new tests have better power than the Lagrange multiplier tests of Breusch (1978) and Godfrey (1978) and the portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978) against an AR(1) process and a fractionally integrated process. The cross-validation procedure of Beltrao and Bloomfield (1987) and Robinson (1991a) works reasonably well in choosing the smoothing parameter of the spectral estimates and is thus recommended for use in practice.

Department of Economics, Cornell University, Uris Hall 454, Ithaca, NY 14853, U.S.A.

Manuscript received March, 1994; final revision received June, 1995.

MATHEMATICAL APPENDIX

For notational simplicity, we consider here the regression model

$$Y_t = c + \alpha_1 Y_{t-1} + \alpha_2 X_t + u_t \qquad (t = 1,2,\ldots,n), \tag{A1}$$

where $|\alpha_1| < 1$ and the exogenous variable X_t is covariance stationary with $EX_t^2 < \infty$. The proof for (1) is completely analogous, with more tedious notations.

Throughout, we put $k_{nj} = k(j/p_n)$, $Z_{jt} = u_t u_{t-j}$, and denote Δ, Δ_1, Δ_2 as generic constants that may differ from equation to equation. We define $\tilde{\rho}(j)$, $\tilde{R}(j)$, and $\tilde{f}_n$ exactly as $\hat{\rho}(j)$, $\hat{R}(j)$, and $\hat{f}_n$ respectively, with $\{u_t\}$ replacing $\{\hat{u}_t\}$. We also use $\|f_1 - f_2\|_\infty = \sup_{\omega \in [-\pi,\pi]} |f_1(\omega) - f_2(\omega)|$.

PROOF OF THEOREM 1: The proof consists of Theorems A.1–A.2 below.

THEOREM A.1: *Suppose* A.1(*a*) *and* A.2 *hold. Let* $p_n \to \infty$, $p_n/n \to 0$. *Then*

$$\left(n \sum_{j=1}^{n-1} k_{nj}^2 \tilde{\rho}^2(j) - C_n(k) \right) \Bigg/ (2D_n(k))^{1/2} \to^d N(0,1).$$

THEOREM A.2: *Suppose* A.1(*a*) *and* A.2–A.3 *hold. Let* $p_n \to \infty$, $p_n/n \to 0$. *Then*

$$\sum_{j=1}^{n-1} k_{nj}^2 (\hat{\rho}^2(j) - \tilde{\rho}^2(j)) = o_P(p_n^{1/2}/n).$$

PROOF OF THEOREM A.1: Given $p_n/n \to 0$, $\tilde{R}(0) - \sigma_0^2 = O_P(n^{-1/2})$ by Chebyshev's inequality, and $\sum_{j=1}^{n-1} k_{nj}^2 \tilde{R}^2(j) = O_P(p_n/n)$ by Markov's inequality, we obtain

$$\sum_{j=1}^{n-1} k_{nj}^2 \tilde{\rho}^2(j) = \sigma_0^{-4} \sum_{j=1}^{n-1} k_{nj}^2 \tilde{R}^2(j) + o_P(p_n^{1/2}/n) = n^{-1}\sigma_0^{-4}(\tilde{C}_n + \tilde{W}_n) + o_P(p_n^{1/2}/n),$$

where $\tilde{C}_n = n^{-1}\sum_{j=1}^{n-1}\sum_{t=j+1}^{n} k_{nj}^2 Z_{jt}^2$ and $\tilde{W}_n = n^{-1}\sum_{j=1}^{n-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} 2k_{nj}^2 Z_{jt}Z_{js}$. Noting $E\tilde{C}_n = \sigma_0^4 C_n(k)$ and $E[\sum_{t=j+1}^{n}(Z_{jt}^2-\sigma_0^4)]^2 = O(n)$, we obtain

$$\operatorname{var}(\tilde{C}_n) = n^{-2}E\left[\sum_{j=1}^{n-1} k_{nj}^2 \sum_{t=j+1}^{n}(Z_{jt}^2-\sigma_0^4)\right]^2 = O(p_n^2/n)$$

by Minkowski's inequality. Hence, $p_n^{-1/2}(\sigma_0^{-4}\tilde{C}_n - C_n(k)) = o_P(1)$ by Chebyshev's inequality and $p_n/n \to 0$.

It remains to show $(2D_n(k))^{-1/2}\sigma_0^{-4}\tilde{W}_n \to^d N(0,1)$. Put $w_{jts} = 2Z_{jt}Z_{js}$. Then

$$\tilde{W}_n = \left(n^{-1}\sum_{j=1}^{l_n} k_{nj}^2 + n^{-1}\sum_{j=l_n+1}^{n-2} k_{nj}^2\right)\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} w_{jts} = W_{1n} + V_{1n}, \quad \text{say,}$$

where we choose l_n such that $l_n/p_n \to \infty$, $l_n/n \to 0$. Next, we partition

$$\begin{aligned} W_{1n} &= n^{-1}\sum_{j=1}^{l_n} k_{nj}^2 \\ &\quad\times\left(\sum_{t=2l_n+3}^{n}\sum_{s=l_n+2}^{t-l_n-1} + \sum_{t=2l_n+3}^{n}\sum_{s=t-l_n}^{t-1} + \sum_{t=l_n+3}^{2l_n+2}\sum_{s=l_n+2}^{t-1} + \sum_{s=j+1}^{l_n+1}\sum_{t=s+1}^{n}\right) w_{jts} \\ &= U_n + V_{2n} + V_{3n} + V_{4n}, \quad \text{say.} \end{aligned}$$

It follows that $\tilde{W}_n = U_n + \sum_{j=1}^{4} V_{jn}$. We shall show (i) $p_n^{-1/2}V_{jn} = o_P(1)$ for $j = 1,2,3,4$ and (ii) $(2\sigma_0^8 D_n(k))^{-1/2}U_n \to^d N(0,1)$. We now verify (i). Given Assumption A.2, we have that for $t_1 > s_1$, $t_2 > s_2$ and $1 \le j_1,\ j_2 \le n-2$,

$$\text{(A2)} \qquad E(w_{j_1t_1s_1}w_{j_2t_2s_2}) = E(w_{j_1t_1s_1}w_{j_2t_2s_2})\delta_{t_1,t_2}\delta_{s_1,t_1-j_2}\delta_{s_2,t_1-j_1}.$$

Using (A2), we have $EV_{1n}^2 \le 2\mu_4^2\sum_{j=l_n+1}^{n-2} k_{nj}^4 + 2n^{-1}\sigma_0^8(\sum_{j=l_n+1}^{n-2} k_{nj}^2)^2 = o(p_n)$, where $p_n^{-1}\sum_{j=l_n+1}^{n-2} k_{nj}^4 \to 0$ given $p_n \to \infty$, $l_n/p_n \to \infty$. Similarly, $EV_{2n}^2 = O(p_n l_n/n + p_n^2 l_n/n)$, $EV_{3n}^2 = O(p_n l_n^2/n^2 + p_n^2 l_n/n^2)$, $EV_{4n}^2 = O(p_n l_n/n + p_n^2/n)$. Hence, (i) follows by Chebyshev's inequality.

Next, we verify (ii). Put $U_{nt} = 2u_t\sum_{j=1}^{l_n} k_{nj}^2 u_{t-j}H_{jt-l_n-1}$, where $H_{jt-l_n-1} = \sum_{s=l_n+2}^{t-l_n-1} Z_{js}$. Then $U_n = n^{-1}\sum_{t=2l_n+3}^{n} U_{nt}$ and $\{U_{nt}, F_{t-1}\}$ is a martingale difference sequence, where F_t is the σ-field consisting of $u_s, s \le t$. Put $\sigma^2(n) = EU_n^2$. We apply Brown's (1971) theorem, which implies $\sigma^{-1}(n)U_n \to^d N(0,1)$ if (a) $\sigma^{-2}(n)n^{-2}\sum_{t=2l_n+3}^{n} E(U_{nt}^2 1[|U_{nt}| > \epsilon n\sigma(n)]) \to 0$ for every $\epsilon > 0$ and (b) $\sigma^{-2}(n)n^{-2}\sum_{t=2l_n+3}^{n}\ddot{U}_{nt}^2 \to^p 1$, where $\ddot{U}_{nt}^2 = E(U_{nt}^2 \mid F_{t-1})$.

We first show $2\sigma_0^8 p_n D(k)/\sigma^2(n) \to 1$. Because u_{t-i} is independent of H_{jt-l_n-1} for $1 \le i, j \le l_n$, we obtain $EU_{nt}^2 = 4\sigma_0^8(t-2l_n-2)\sum_{j=1}^{l_n} k_{nj}^4$, $t \ge 2l_n+3$. It follows that

$$\text{(A3)} \qquad \sigma^2(n) = 2\sigma_0^8 n^{-2}(n-2l_n-1)(n-2l_n-2)\sum_{j=1}^{l_n} k_{nj}^4 = 2\sigma_0^8 p_n D(k)(1+o(1))$$

given $p_n \to \infty$, $l_n/p_n \to \infty$, and $l_n/n \to 0$, where $p_n^{-1}\sum_{j=1}^{l_n} k_{nj}^4 \to D(k)$.

We now verify (a). Noting u_{t-i} is independent of H_{jt-l_n-1} for $1 \le i, j \le l_n$, we have

$$EU_{nt}^4 = 16\mu_4 E\left[\sum_{j=1}^{l_n} k_{nj}^2 u_{t-j}H_{jt-l_n-1}\right]^4 \le 48\mu_4^2\left[\sum_{j=1}^{l_n} k_{nj}^4 (EH_{jt-l_n-1}^4)^{1/2}\right]^2 = O(t^2p_n^2),$$

where $\sup_{1\le j\le l_n} EH_{jt-l_n-1}^4 = O(t^2)$. Hence, $\sigma^{-4}(n)n^{-4}\sum_{t=2l_n+3}^{n} EU_{nt}^4 = O(n^{-1})$. This proves (a).

To verify (b), we show $\sigma^{-4}(n)E(\ddot{U}_n^2 - \sigma^2(n))^2 \to 0$, where $\ddot{U}_n^2 = n^{-2}\sum_{t=2l_n+3}^{n}\ddot{U}_{nt}^2$. By definition, we have $\ddot{U}_{nt}^2 = 4\sigma_0^2[\sum_{j=1}^{l_n} k_{nj}^2 u_{t-j} H_{jt-l_n-1}]^2$. We decompose

$$\text{(A4)}\qquad \ddot{U}_{nt}^2 = 4\sigma_0^2 \sum_{j=1}^{l_n} k_{nj}^4 u_{t-j}^2 H_{jt-l_n-1}^2 + 8\sigma_0^2 \sum_{j=2}^{l_n}\sum_{i=1}^{j-1} k_{ni}^2 k_{nj}^2 u_{t-i}u_{t-j}H_{it-l_n-1}H_{jt-l_n-1}$$
$$= 4\sigma_0^2 B_{nt} + 4\sigma_0^2 A_{1nt}, \quad \text{say},$$

where

$$\text{(A5)}\qquad B_{nt} = \sum_{j=1}^{l_n} k_{nj}^4 u_{t-j}^2\left(\sum_{s=l_n+2}^{t-l_n-1} Z_{js}^2\right) + \sum_{j=1}^{l_n} k_{nj}^4 u_{t-j}^2\left(\sum_{s=l_n+3}^{t-l_n-1}\sum_{r=l_n+2}^{s-1} w_{jsr}\right)$$
$$= C_{nt} + A_{2nt}, \quad \text{say},$$

where, as before, $w_{jsr} = 2Z_{js}Z_{jr}$. Furthermore, we can decompose

$$\text{(A6)}\qquad C_{nt} = \frac{1}{4\sigma_0^2}EU_{nt}^2 + \sum_{j=1}^{l_n} k_{nj}^4 u_{t-j}^2 \sum_{s=l_n+2}^{t-l_n-1}(Z_{js}^2 - \sigma_0^4) + (t-2l_n-2)\sigma_0^4 \sum_{j=1}^{l_n} k_{nj}^4(u_{t-j}^2 - \sigma_0^2)$$
$$= (4\sigma_0^2)^{-1}EU_{nt}^2 + A_{3nt} + A_{4nt}, \quad \text{say}.$$

Combining (A4)–(A6) yields $\ddot{U}_n^2 - \sigma^2(n) = 4\sigma_0^2 n^{-2}\sum_{j=1}^{4}\sum_{t=2l_n+3}^{n} A_{jnt}$. Hence, it suffices to show $\sigma^{-4}(n)n^{-4}E(\sum_{t=2l_n+3}^{n} A_{jnt})^2 \to 0$ for $j = 1,2,3,4$. Noting u_{t-i} is independent of H_{jt-l_n-1}, $1 \le i, j \le l_n$, we have $EA_{1nt}^2 = 4\sigma_0^4\sum_{j=2}^{l_n}\sum_{i=1}^{j-1} k_{ni}^4 k_{nj}^4 E(H_{it-l_n-1}^2 H_{jt-l_n-1}^2) = O(t^2p_n^2)$ by Cauchy-Schwarz (C-S) inequality and $EH_{jt-l_n-1}^4 = O(t^2)$. For $s < t - l_n$, it is easy to show $E(A_{1nt}A_{1ns}) = 0$. It follows that

$$\text{(A7)}\qquad n^{-4}E\left(\sum_{t=2l_n+3}^{n} A_{1nt}\right)^2 = n^{-4}\sum_{|t-s|\le l_n} E(A_{1nt}A_{1ns}) = O(p_n^2 l_n/n) = o(p_n^2)$$

by the C-S inequality and $l_n/n \to 0$. Next, noting u_{t-i} is independent of Z_{js} for $1 \le i, j \le l_n$, $s < t - l_n$, and using (A.2), we have $EA_{2nt}^2 = O(t^2p_n + tp_n^2)$. It follows by Minkowski's inequality that

$$\text{(A8)}\qquad n^{-4}E\left(\sum_{t=2l_n+3}^{n} A_{2nt}\right)^2 \le n^{-4}\left(\sum_{t=2l_n+3}^{n}(EA_{2nt}^2)^{1/2}\right)^2 = O(p_n) = o(p_n^2).$$

Similarly, $EA_{3nt}^2 \le \mu_4(\sum_{j=1}^{l_n} k_{nj}^4[E(\sum_{s=l_n+2}^{t-l_n-1}(Z_{js}^2 - \sigma_0^4))^2]^{1/2})^2 = O(tp_n^2)$ by Minkowski's inequality and $E[\sum_{s=l_n+2}^{t-l_n-1}(Z_{js}^2 - \sigma_0^4)]^2 = O(t)$. Therefore,

$$\text{(A9)}\qquad n^{-4}E\left(\sum_{t=2l_n+3}^{n} A_{3nt}\right)^2 \le n^{-4}\left(\sum_{t=2l_n+3}^{n}(EA_{3nt}^2)^{1/2}\right)^2 = O(p_n^2/n) = o(p_n^2).$$

Finally, because $EA_{4nt}^2 = (t-2l_n-2)^2\sigma_0^8\sum_{j=1}^{l_n} k_{nj}^8 E(u_{t-j}^2 - \sigma_0^2)^2 = O(t^2p_n)$, we have

$$\text{(A10)}\qquad n^{-4}E\left(\sum_{t=2l_n+3}^{n} A_{4nt}\right)^2 \le n^{-4}\left(\sum_{t=2l_n+3}^{n}(EA_{4nt}^2)^{1/2}\right)^2 = O(p_n) = o(p_n^2).$$

Combining (A7)–(A10) yields $\sigma^{-4}(n)E(\ddot{U}_n^2 - \sigma^2(n))^2 \to 0$. If follows that $\sigma^{-1}(n)U_n \to^d N(0,1)$ by Brown's theorem. Because $p_n^{-1}D_n(k) \to D(k)$, we have $(2\sigma_0^8 D_n(k))^{-1/2}U_n \to^d N(0,1)$, i.e. (ii) holds. This completes the proof. *Q.E.D.*

PROOF OF THEOREM A.2: The result follows if $\sum_{j=1}^{n-1} k_{nj}^2(\hat{R}^2(j) - \tilde{R}^2(j)) = o_P(p_n^{1/2}/n)$. Because $\hat{R}^2(j) - \tilde{R}^2(j) = (\hat{R}(j) - \tilde{R}(j))^2 + 2\tilde{R}(j)(\hat{R}(j) - \tilde{R}(j))$, we shall show (i) $\sum_{j=1}^{n-1} k_{nj}^2(\hat{R}(j) - \tilde{R}(j))^2 = O_P(n^{-1})$ and (ii) $\sum_{j=1}^{n-1} k_{nj}^2 \tilde{R}(j)(\hat{R}(j) - \tilde{R}(j)) = o_P(p_n^{1/2}/n)$.

Put $\hat{\lambda}_{nt} = (\hat{c} - c) + (\hat{\alpha}_1 - \alpha_1)Y_{t-1} + (\hat{\alpha}_2 - \alpha_2)X_t$. Then $\hat{u}_t = u_t - \hat{\lambda}_{nt}$, and for $j \geq 1$,

$$\text{(A11)} \qquad \hat{R}(j) - \tilde{R}(j) = -n^{-1} \sum_{t=j+1}^{n} \hat{\lambda}_{nt} u_{t-j} - n^{-1} \sum_{t=j+1}^{n} u_t \hat{\lambda}_{nt-j} + n^{-1} \sum_{t=j+1}^{n} \hat{\lambda}_{nt} \hat{\lambda}_{nt-j}.$$

We first verify (i). Using (A11), we obtain

$$\text{(A12)} \qquad \sum_{j=1}^{n-1} k_{nj}^2(\hat{R}(j) - \tilde{R}(j))^2 \leq 4n^{-2} \sum_{j=1}^{n-1} k_{nj}^2 \Bigg[\Bigg(\sum_{t=j+1}^{n} \hat{\lambda}_{nt} u_{t-j} \Bigg)^2 + \Bigg(\sum_{t=j+1}^{n} u_t \hat{\lambda}_{nt-j} \Bigg)^2 + \Bigg(\sum_{t=j+1}^{n} \hat{\lambda}_{nt} \hat{\lambda}_{t-j} \Bigg)^2 \Bigg]$$

$$= 4T_{1n} + 4T_{2n} + 4T_{3n}, \quad \text{say.}$$

It suffices to show $T_{jn} = O_P(n^{-1})$ for $j = 1, 2, 3$. For the first term, we have

$$T_{1n} \leq 4(\hat{c} - c)^2 \sum_{j=1}^{n-1} k_{nj}^2 \Bigg(n^{-1} \sum_{t=j+1}^{n} u_{t-j} \Bigg)^2 + 4(\hat{\alpha}_1 - \alpha_1)^2 \sum_{j=1}^{n-1} k_{nj}^2 \Bigg(n^{-1} \sum_{t=j+1}^{n} u_{t-j} Y_{t-1} \Bigg)^2 + 4(\hat{\alpha}_2 - \alpha_2)^2 \sum_{j=1}^{n-1} k_{nj}^2 \Bigg(n^{-1} \sum_{t=j+1}^{n} u_{t-j} X_t \Bigg)^2$$

$$= 4(\hat{c} - c)^2 S_{1n} + 4(\hat{\alpha}_1 - \alpha_1)^2 S_{2n} + 4(\hat{\alpha}_2 - \alpha_2)^2 S_{3n}, \quad \text{say.}$$

Here, $S_{1n} = O_P(p_n/n)$ by Markov's inequality and $E|S_{1n}| = O(p_n/n)$. Similarly, $S_{2n} = O_P(1)$ because $E|S_{2n}| \leq \sum_{j=1}^{n-1} k_{nj}^2(\Delta_1 n^{-1} + \Delta_2 \alpha_1^{2(j-1)}) = O(1)$ by $|\alpha_1| < 1$ and Lemma A.1, which is given at the end of this proof. Finally, $S_{3n} = O_P(p_n/n)$ given $E(n^{-1} \sum_{t=j+1}^{n} u_{t-j} X_t)^2 = O(n^{-1})$ by strict exogeneity of X_t. If follows from A.3 that $T_{1n} = O_P(n^{-1})$ given $p_n/n \to 0$. Similarly, we can show $T_{2n} = o_P(n^{-1})$. We also have $T_{3n} \leq (n^{-1} \sum_{t=1}^{n} \hat{\lambda}_{nt}^2)^2 \sum_{j=1}^{n-1} k_{nj}^2 = o_P(n^{-1})$, where $n^{-1} \sum_{t=1}^{n} \hat{\lambda}_{nt}^2 = O_P(n^{-1})$, as can be shown easily. Hence (i) holds.

We now verify condition (ii). Using (A11), we have

$$\sum_{j=1}^{n-1} k_{nj}^2 \tilde{R}(j)(\hat{R}(j) - \tilde{R}(j)) = -n^{-1} \sum_{j=1}^{n-1} k_{nj}^2 \tilde{R}(j) \times \Bigg(\sum_{t=j+1}^{n} \hat{\lambda}_{nt} u_{t-j} + \sum_{t=j+1}^{n} u_t \hat{\lambda}_{nt-j} - \sum_{t=j+1}^{n} \hat{\lambda}_{nt} \hat{\lambda}_{nt-j} \Bigg)$$

$$= -T_{4n} - T_{5n} + T_{6n}, \quad \text{say.}$$

It suffices to show $T_{jn} = o_P(p_n^{1/2}/n)$ for $j = 4,5,6$. For the first term,

$$T_{4n} = (\hat{c} - c) \sum_{j=1}^{n-1} k_{nj}^2 \tilde{R}(j) \left(n^{-1} \sum_{t=j+1}^{n} u_{t-j} \right) + (\hat{\alpha}_1 - \alpha_1) \sum_{j=1}^{n-1} k_{nj}^2 \tilde{R}(j)$$

$$\times \left(n^{-1} \sum_{t=j+1} u_{t-j} Y_{t-1} \right) + (\hat{\alpha}_2 - \alpha_2) \sum_{j=1}^{n-1} k_{nj}^2 \tilde{R}(j) \left(n^{-1} \sum_{t=j+1}^{n} u_{t-j} X_t \right)$$

$$= (\hat{c} - c) S_{4n} + (\hat{\alpha}_1 - \alpha_1) S_{5n} + (\hat{\alpha}_2 - \alpha_2) S_{6n}, \quad \text{say.}$$

Here, $S_{4n} = O_P(p_n/n)$ because $E|S_{4n}| \le \sum_{j=1}^{n-1} k_{nj}^2 [E\tilde{R}^2(j)]^{1/2} [E(n^{-1} \sum_{t=j+1}^{n} u_{t-j})^2]^{1/2} = O(p_n/n)$. Similarly, $S_{6n} = O_P(p_n/n)$ by exogeneity of X_t. We also have $S_{5n} = O_P(p_n/n + n^{-1/2})$ because by the C-S inequality and Lemma A.1 below,

$$E|S_{5n}| \le \sum_{j=1}^{n-1} k_{nj}^2 (E\tilde{R}^2(j))^{1/2} \left(E \left(n^{-1} \sum_{t=j+1}^{n} u_{t-j} Y_{t-1} \right)^2 \right)^{1/2}$$

$$\le \sigma_0^2 \Delta_1^{1/2} n^{-1} \sum_{j=1}^{n-1} k_{nj}^2 + \sigma_0^2 \Delta_2^{1/2} n^{-1/2} \sum_{j=1}^{n-1} k_{nj}^2 |\alpha_1|^{j-1}$$

$$= O(p_n/n + n^{-1/2}).$$

Therefore, $T_{4n} = o_P(p_n^{1/2}/n)$ given $p_n \to \infty$, $p_n/n \to 0$, and Assumption A.3. Next, $T_{5n} = o_P(p_n^{1/2}/n)$ by the C-S inequality, $\sum_{j=1}^{n} k_{nj}^2 \tilde{R}^2(j) = O_P(p_n/n)$, and $T_{2n} = o_P(n^{-1})$. Similarly, $T_{6n} = o_P(p_n^{1/2}/n)$ given $T_{3n} = o_P(n^{-1})$. It follows that condition (ii) holds. Thus, the proof will be completed provided the following lemma is proved.

LEMMA A.1: $E(n^{-1} \sum_{t=j+1}^{n} u_{t-j} Y_{t-1})^2 \le \Delta_1 n^{-1} + \Delta_2 \alpha_1^{2(j-1)}$ *for all* $j \ge 1$.

PROOF OF LEMMA A.1: We first rewrite (A1) as $Y_t = c_0 + \alpha_2 a(B) X_t + a(B) u_t$, where $c_0 = c/(1-\alpha_1)$, $a(B) = (1 - \alpha_1 B)^{-1} = \sum_{j=0}^{\infty} \alpha_1^j B^j$. Then for each $j \ge 1$,

$$\tfrac{1}{4} E \left(\sum_{t=j+1}^{n} u_{t-j} Y_{t-1} \right)^2 \le c_0^2 E \left(\sum_{t=j+1}^{n} u_{t-j} \right)^2$$

$$+ \alpha_2^2 E \left(\sum_{t=j+1}^{n} u_{t-j} a(B) X_{t-1} \right)^2 + E \left(\sum_{t=j+1}^{n} u_{t-j} a(B) u_{t-1} \right)^2.$$

The first term is $O(n)$; the second term $E[\sum_{t=j+1}^{n} u_{t-j} a(B) X_{t-1}]^2 = \sigma_0^2 \sum_{t=j+1}^{n} E[a(B) X_{t-1}]^2 = O(n)$ by strict exogeneity of X_t, $EX_t^2 < \infty$, and $|\alpha_1| < 1$. For the third term,

$$E \left(\sum_{t=j+1}^{n} u_{t-j} a(B) u_{t-1} \right)^2 = \sum_{t=j+1}^{n} E\left[u_{t-j}^2 (a(B) u_{t-1})^2 \right]$$

$$+ 2 \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} E[u_{t-j} a(B) u_{t-1}][u_{s-j} a(B) u_{s-1}]$$

$$\le \Delta_1 n + \Delta_2 n^2 \alpha_1^{2(j-1)},$$

where $E[u_{t-j}^2 (a(B) u_{t-1})^2] \le \mu_4 \sum_{i=0}^{\infty} \alpha_1^{2i} \le \Delta_1$ and $E[u_{t-j} u_{s-j} (a(B) u_{t-1})(a(B) u_{s-1})] \le 2 \sigma_0^4 \alpha_1^{2(j-1)}$ for $s < t$, as can be shown easily. The desired result follows immediately. *Q.E.D.*

PROOF OF THEOREM 2: For $j = 0, 1, \ldots, p_n$, let $\hat{U}_j = (0, \ldots, 0, \hat{u}_1, \ldots, \hat{u}_{n-j})'$ and $U_j = (0, \ldots, 0, u_1, \ldots, u_{n-j})'$ be $n \times 1$ vectors with the first j elements being zero. Define $\hat{U}_- = (\hat{U}_1, \ldots, \hat{U}_{p_n})$

and $U_- = (U_1, \dots, U_{p_n})$. Then $R^2 = \hat{U}_0'\hat{U}_-(\hat{U}_-'\hat{U}_-)^{-1}\hat{U}_-'\hat{U}_0/n\hat{\sigma}_n^2$ by definition, where $\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^n(\hat{u}_t - \ddot{u}_n)^2$, $\ddot{u}_n = n^{-1}\sum_{t=1}^n \hat{u}_t$.

Put $\ddot{R}_n^2 = \hat{U}_0'\hat{U}_-(\hat{U}_-'\hat{U}_-)^{-1}\hat{U}_-'\hat{U}_0/n\hat{R}(0)$. Then $R^2 - \ddot{R}^2 = \ddot{R}^2(\hat{R}(0)/\hat{\sigma}_n^2 - 1) = O_P(n^{-1})$ given $\ddot{R}^2 \le 1$ and $\hat{\sigma}_n^2 - \hat{R}(0) = O_P(n^{-1})$. It follows that $M_R - M_{\ddot{R}} = o_P(1)$, where $M_{\ddot{R}} = (n\ddot{R}^2 - p_n)/(2p_n)^{1/2}$. Hence, it suffices to show $M_{\ddot{R}} - M_{1n}^T = o_P(1)$. Now,

$$\text{(A13)} \qquad \ddot{R}^2 - \sum_{j=1}^{p_n} \hat{\rho}^2(j) = \left(\hat{U}_0'\hat{U}_-(\hat{U}_-'\hat{U}_-)^{-1}\hat{U}_-'\hat{U}_0/n\hat{R}(0) - \hat{U}_0'\hat{U}_-\hat{U}_-'\hat{U}_0/(n\hat{R}(0))^2\right)$$

$$= (\ddot{R}^2/\hat{R}(0))\theta_{p_n}'\left(\hat{R}(0)I_{p_n} - n^{-1}\hat{U}_-'\hat{U}_-\right)\theta_{p_n},$$

where $\theta_{p_n} = (\hat{U}_-'\hat{U}_-)^{-1/2}\hat{U}_-'\hat{U}_0/[\hat{U}_0'\hat{U}_-(\hat{U}_-'\hat{U}_-)^{-1}\hat{U}_-'\hat{U}_0]^{1/2}$ such that $\theta_{p_n}'\theta_{p_n} = 1$. Put $\Theta_{p_n} = \{\theta \in R^{p_n}: \theta'\theta = 1\}$ and $\eta_n = \sup_{\theta_{p_n} \in \Theta_{p_n}} |\theta_{p_n}'(\hat{R}(0)I_{p_n} - n^{-1}\hat{U}_-'\hat{U}_-)\theta_{p_n}|$. Then

$$\eta_n = \sup_{\theta_{p_n} \in \Theta_{p_n}} n^{-1}\left|\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left(\sum_{t=1}^n \hat{u}_t^2\delta_{ij} - \sum_{t=\max(i,j)+1}^n \hat{u}_{t-i}\hat{u}_{t-j}\right)\theta_{p_n i}\theta_{p_n j}\right|$$

$$= \sup_{\theta_{p_n} \in \Theta_{p_n}} n^{-1}\left|\sum_{j=1}^{p_n}\theta_{p_n j}^2 \sum_{t=n-j+1}^n \hat{u}_t^2 - 2\sum_{j=2}^{p_n}\sum_{i=1}^{j-1}\theta_{p_n i}\theta_{p_n j}\sum_{t=j+1}^n \hat{u}_{t-i}\hat{u}_{t-j}\right|$$

$$\le n^{-1}\sum_{t=n-p_n+1}^n \hat{u}_t^2 + 2n^{-1}\sum_{j=2}^{p_n}\sum_{i=1}^{j-1}\left|\sum_{t=j+1}^n u_{t-i}u_{t-j}\right|$$

$$+ 2n^{-1}\sum_{j=2}^{p_n}\sum_{i=1}^{j-1}\left|\sum_{t=j+1}^n u_{t-i}\hat{\lambda}_{nt-j}\right|$$

$$+ 2n^{-1}\sum_{j=2}^{p_n}\sum_{i=1}^{j-1}\left|\sum_{t=j+1}^n u_{t-j}\hat{\lambda}_{nt-i}\right| + 2n^{-1}\sum_{j=2}^{p_n}\sum_{i=1}^{j-1}\left|\sum_{t=j+1}^n \hat{\lambda}_{nt-i}\hat{\lambda}_{nt-j}\right|$$

$$= \eta_{1n} + 2\eta_{2n} + 2\eta_{3n} + 2\eta_{4n} + 2\eta_{5n}, \quad \text{say},$$

where $\hat{u}_t = u_t - \hat{\lambda}_{nt}$, $\hat{\lambda}_{nt}$ is as in the proof of Theorem A.2. It is straightforward to show that $\eta_{1n} = O_P(p_n/n)$, $\eta_{2n} = O_P(p^2/n^{1/2})$ by Markov's inequality, $\eta_{3n} = O_P(p_n^2/n^{1/2})$, $\eta_{4n} = O_P(p_n^2/n^{1/2})$, and $\eta_{5n} = O_P(p_n^2/n)$ by the C-S inequality and $n^{-1}\sum_{t=1}^n \hat{\lambda}_{nt}^2 = O_P(n^{-1})$. It follows that $\eta_n = O_P(p_n^2/n^{1/2})$. Hence, we obtain $(\ddot{R}^2 - \sum_{j=1}^{p_n}\hat{\rho}^2(j))/\ddot{R}^2 = O_P(p_n^2/n^{1/2}) = o_P(1)$ from (A13) and $p_n^5/n \to 0$. This implies $\ddot{R}^2 = O_P(p_n/n)$ because $\sum_{j=1}^{p_n}\hat{\rho}(j) = O_P(p_n/n)$. Consequently, $\ddot{R}^2 - \sum_{j=1}^{p_n}\hat{\rho}^2(j) = o_P(p_n^{1/2}/n)$ given $p_n^5/n \to 0$, so $M_{\ddot{R}} - M_{1n}^T = o_P(1)$. Because $M_{1n}^T \to^d N(0,1)$ by Theorem 1, we have $M_{\ddot{R}}$ (and M_R) $\to^d N(0,1)$. *Q.E.D.*

PROOF OF THEOREM 3: Put $\hat{\delta}_n(\omega) = \hat{f}_n(\omega)/f_0(\omega) - 1$. We first show $\sup_{\omega \in [-\pi,\pi]}|\hat{\delta}_n(\omega)| = o_P(1)$. Consider $\|\hat{f}_n - \tilde{f}_n\|_\infty$ and $\|\tilde{f}_n - f_0\|_\infty$. We have $\|\hat{f}_n - \tilde{f}_n\|_\infty = O_P(p_n/n^{1/2})$ because from (A11)

$$\text{(A14)} \qquad \sum_{j=1}^{n-1}|k_{nj}||\hat{R}(j) - \tilde{R}(j)|$$

$$\le n^{-1}\sum_{j=1}^{n-1}|k_{nj}|\left(\left|\sum_{t=j+1}^n \hat{\lambda}_{nt}u_{t-j}\right| + \left|\sum_{t=j+1}^n u_t\hat{\lambda}_{nt-j}\right| + \left|\sum_{t=j+1}^n \hat{\lambda}_{nt}\hat{\lambda}_{nt-j}\right|\right)$$

$$\le 2\left(\sum_{j=1}^{n-1}|k_{nj}|\right)\left(n^{-1}\sum_{t=1}^n \hat{\lambda}_{nt}^2\right)^{1/2}\left(n^{-1}\sum_{t=1}^n u_t^2\right)^{1/2} + \left(\sum_{j=1}^{n-1}|k_{nj}|\right)\left(n^{-1}\sum_{t=1}^n \hat{\lambda}_{nt}^2\right)$$

$$= O_P(p_n/n^{1/2}),$$

where $p_n^{-1}\sum_{j=1}^{n-1}|k_{nj}|\to\int_0^\infty|k(z)|\,dz$ and $\sum_{t=1}^n\hat\lambda_{nt}^2=O_P(n^{-1})$. Also, $\|\tilde f_n-f_0\|_\infty=O_P(p_n/n^{1/2})$ because $\sum_{j=1}^{n-1}|k_{nj}|\,E|\tilde R(j)|=O_P(p_n/n^{1/2})$. It follows that $\|\hat f_n-f_0\|_\infty=o_P(p_n^{-1/2})$ by the triangle inequality and $p_n^3/n\to 0$.

Now, noting $H^2(\hat f_n;f_0)=2(1-\int_{-\pi}^{\pi}(1+\hat\delta_n(\omega))^{1/2}f_0(\omega)\,d\omega)$, $Q^2(\hat f_n;f_0)=\int_{-\pi}^{\pi}\hat\delta_n^2(\omega)f_0(\omega)\,d\omega$, $\int_{-\pi}^{\pi}\hat\delta_n(\omega)f_0(\omega)\,d\omega=0$, and $|(1+z)^{1/2}=1-\frac12 z+\frac18 z^2|\le|z^3|$ for small z, we obtain

$$
\begin{aligned}
&|2H^2(\hat f_n;f_0)-\tfrac12 Q^2(\hat f_n;f_0)|\\
&\quad=4\left|\int_{-\pi}^{\pi}\left[(1+\hat\delta_n(\omega))^{1/2}-1-\tfrac12\hat\delta_n(\omega)+\tfrac18\hat\delta^2(\omega)\right]f_0(\omega)\,d\omega\right|\\
&\quad\le 4\int_{-\pi}^{\pi}|\hat\delta_n^3(\omega)|\,f_0(\omega)\,d\omega\\
&\quad\le \sup_{\omega\in[-\pi,\pi]}|\hat\delta_n(\omega)|\,Q^2(\hat f_n;f_0)=o_P(p_n^{1/2}/n),
\end{aligned}
$$

where $Q^2(\hat f_n;f_0)=O_P(p_n/n)$. It follows that $M_{2n}-M_{1n}=o_P(1)$ and $M_{2n}\to^d N(0,1)$.

Next, we consider M_{3n}. Let $\Omega^c(\hat f_n)=\{\omega\in[-\pi,\pi]:\hat f_n(\omega)=0\}$, the complement of $\Omega(\hat f_n)$. Noting $|\ln(1+z)-z+\frac12 z^2|\le|z^3|$ for small z, $\int_{-\pi}^{\pi}\hat\delta_n(\omega)f_0(\omega)\,d\omega=0$ and $\hat\delta_n(\omega)=-1$ for $\omega\in\Omega^c(\hat f_n)$, we obtain

$$
\begin{aligned}
&|I(\hat f_n;f_0)-\tfrac12 Q^2(\hat f_n;f_0)|\\
&\quad=\left|\int_{\Omega(\hat f_n)}\ln(1+\hat\delta_n(\omega))f_0(\omega)\,d\omega+\tfrac12\int_{-\pi}^{\pi}\hat\delta_n^2(\omega)f_0(\omega)\,d\omega\right|\\
&\quad=\left|\int_{\Omega(\hat f_n)}(\ln(1+\hat\delta_n(\omega))-\hat\delta_n(\omega)+\tfrac12\hat\delta_n^2(\omega))f_0(\omega)\,d\omega\right.\\
&\qquad\left.+\int_{\Omega^c(\hat f_n)}(\hat\delta_n(\omega)+\tfrac12\hat\delta_n^2(\omega))f_0(\omega)\,d\omega\right|\\
&\quad\le\left|\int_{\Omega(\hat f_n)}(\ln(1+\hat\delta_n(\omega))-\hat\delta_n(\omega)+\tfrac12\hat\delta_n^2(\omega))f_0(\omega)\,d\omega\right.\\
&\qquad\left.+\tfrac12\int_{\Omega^c(\hat f_n)}|\hat\delta_n^3(\omega)|\,f_0(\omega)\,d\omega\right|\\
&\quad\le\int_{-\pi}^{\pi}|\hat\delta_n^3(\omega)|\,d\omega=o_P(p_n^{1/2}/n).
\end{aligned}
$$

It follows that $M_{3n}-M_{1n}=o_P(1)$ and $M_{3n}\to^d N(0,1)$. *Q.E.D*

PROOF OF THEOREM 4: Noting $f_n^0(\omega)=f_0(\omega)+a_n g(\omega)$ under H_{an}, we have

$$
Q^2(\hat f_n;f_n^0)=Q^2(\hat f_n;f_0)+2\pi a_n^2\int_{-\pi}^{\pi}g^2(\omega)\,d\omega-4\pi a_n\int_{-\pi}^{\pi}(\hat f_n(\omega)-f_0(\omega))g(\omega)\,d\omega. \tag{A15}
$$

We first show that the last term of (A15) is $o_P(p_n^{1/2}/n)$. Write

$$
\begin{aligned}
&\int_{-\pi}^{\pi}(\hat f_n(\omega)-f_0(\omega))g(\omega)\,d\omega\\
&\quad=\int_{-\pi}^{\pi}(\hat f_n(\omega)-\tilde f_n(\omega))g(\omega)\,d\omega+\int_{-\pi}^{\pi}(\tilde f_n(\omega)-f_0(\omega))g(\omega)\,d\omega.
\end{aligned}
$$

Put $g_j = \int_{-\pi}^{\pi} g(\omega)e^{ij\omega}\,d\omega$. Then the first term is $O_P(n^{-1/2})$ because $|\sum_{j=1}^{n-1} k_{nj}(\hat{R}_j) - \tilde{R}(j))g_j| = O_P(n^{-1/2})$ by the C-S inequality, $\sum_{j=1}^{\infty} g_j^2 < \infty$, and $\sum_{j=1}^{n-1} k_{nj}^2(\hat{R}(j) - \tilde{R}(j))^2 = O_P(n^{-1})$, as shown in the proof of Theorem A.2. Also, the second term is $O_P(n^{-1/2})$ because $\sum_{j=1}^{n-1} k_{nj}\tilde{R}(j)g_j = O_P(n^{-1/2})$ given $E[\sum_{j=1}^{n-1} k_{nj}\tilde{R}(j)g_j]^2 = O(n^{-1})$. It follows that $\int_{-\pi}^{\pi}(\hat{f}_n(\omega) - f_0(\omega))g(\omega)\,d\omega = O_P(n^{-1/2})$. Hence, the last term of (A15) is $O_P(p_n^{1/4}/n)$ given $a_n = p_n^{1/4}/n^{1/2}$. Consequently, $Q^2(\hat{f}_n; f_n^0) = Q^2(\hat{f}_n; f_0) + (p_n^{1/2}/n)2\pi\int_{-\pi}^{\pi} g^2(\omega)\,d\omega + o_P(p_n^{1/2}/n)$. This implies $M_{1n}^a = M_{1n} + \mu(k) + o_P(1)$ and $M_{1n}^a \to^d N(\mu(k), 1)$. The results for M_{2n}^a and M_{3n}^a follow also because it can be shown that $M_{2n}^a = M_{2n} + \mu(k) + o_P(1)$ and $M_{3n}^a = M_{3n} + \mu(k) + o_P(1)$ given $p_n^3/n \to 0$, using an analogous proof for Theorem 4. *Q.E.D.*

PROOF OF THEOREM 5: Given $\mu(k) = 2\pi\int_{-\pi}^{\pi} g^2(\omega)\,d\omega/(2D(k))^{1/2}$, maximal power of the M_{jn}^a is obtained by minimizing $D(k)$. Because k is the Fourier transform of K, k^2 is the Fourier transform of the convolution of K with itself, i.e. $k^2(z) = \int_{-\infty}^{\infty} G(\lambda)e^{i\lambda z}\,d\lambda$, where $G(\lambda) = \int_{-\infty}^{\infty} K(\lambda')K(\lambda - \lambda')\,d\lambda'$. We have $D(k) = \frac{1}{2}\int_{-\infty}^{\infty} k^4(z)\,dz = \pi\int_{-\infty}^{\infty} G^2(\lambda)\,d\lambda$ by Parseval's identity. Hence, minimizing $D(k)$ over $\kappa(\tau)$ is equivalent to minimizing $\int_{-\infty}^{\infty} G^2(\lambda)\,d\lambda$ over $\Lambda(\tau) = \{K: R \to R^+ \mid \int_{-\infty}^{\infty} K(\lambda)\,d\lambda = 1,\ \int_{-\infty}^{\infty} \lambda K(\lambda)\,d\lambda = 0,\ \int_{-\infty}^{\infty} \lambda^2 K(\lambda)\,d\lambda = \tau^2\}$, where $\Lambda(\tau)$ is the dual set of $\kappa(\tau)$. By Ghosh and Huang (1991, Theorem 1), $K^*(\lambda) = (2\sqrt{3}\,\tau)^{-1}1[|\lambda| \le \sqrt{3}\,\tau]$ minimizes $\int G^2(\lambda)\,d\lambda$ over $\Lambda(\tau)$. Because k_D is the Fourier transform of K^*, it follows that k_D minimizes $D(k)$ over $\kappa(\tau)$. *Q.E.D*

PROOF OF THEOREM 6: (a) We first consider M_{1n}. Given $(p_n^{1/2}/n)M_{1n} = \frac{1}{2}Q^2(\hat{f}_n; f_0)/(2D(k))^{1/2}(1 + o(1))$, it suffices to show $Q^2(\hat{f}_n; f_0) \to^p Q^2(f; f_0)$. Because

$$Q^2(\hat{f}_n; f_0) = Q^2(f; f_0) + 2\pi\int_{-\pi}^{\pi} (\hat{f}_n(\omega) - f(\omega))^2\,d\omega + 4\pi\int_{-\pi}^{\pi} (\hat{f}_n(\omega) - f(\omega))(f(\omega) - f_0(\omega))\,d\omega,$$

it suffices to show that the second term is $o_P(1)$ because the last term will be also $o_P(1)$ by the C-S inequality. We shall show (i) $\int_{-\pi}^{\pi}(\hat{f}_n(\omega) - \tilde{f}_n(\omega))^2\,d\omega = o_P(1)$ and (ii) $\int_{-\pi}^{\pi}(\tilde{f}_n(\omega) - f(\omega))^2\,d\omega = o_P(1)$.

It suffices for (i) if $\sum_{j=1}^{n-1} k_{nj}^2(\hat{R}(j) - \tilde{R}(j))^2 = o_P(1)$. In the proof of Theorem A.2, we have obtained $\sum_{j=1}^{n-1} k_{nj}^2(\hat{R}(j) - \tilde{R}(j))^2 \le 4(T_{1n} + T_{2n} + T_{3n})$, where the T_{jn} are as in (A12). Now, $T_{1n} \le (\sum_{j=1}^{n-1} k_{nj}^2)(n^{-1}\sum_{t=1}^{n} \hat{\lambda}_{nt}^2)(n^{-1}\sum_{t=1}^{n} u_t^2) = O_P(p_n/n)$, where $n^{-1}\sum_{t=1}^{n} \hat{\lambda}_{nt}^2 = O_P(n^{-1})$ under the assumed static model and Assumption A.3. Similarly, $T_{2n} = O_P(p_n/n)$ and $T_{3n} = O_P(p_n/n^2)$. Hence, condition (i) holds.

To verify (ii), it suffices to consider

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{j=1}^{n-1} \tilde{R}(j)e^{-ij\omega} - \sum_{j=1}^{\infty} R(j)e^{-ij\omega}\right)^2 d\omega$$

$$= \sum_{j=1}^{n-1} (k_{nj}\tilde{R}(j) - R(j))^2 + \sum_{j=n}^{\infty} R^2(j)$$

$$\le 2\sum_{j=1}^{n-1} (k_{nj} - 1)^2 R^2(j) + 2\sum_{j=1}^{n-1} k_{nj}^2(\tilde{R}(j) - R(j))^2 + \sum_{j=n}^{\infty} R^2(j),$$

where the equality follows by Parseval's identity. The first term is $o(1)$ by Lebesgue's dominated convergence theorem given $k_{nj} - 1 \to 0$ as $p_n \to \infty$ for all $j \ge 1$ and $\sum_{j=0}^{\infty} R^2(j) < \infty$. Next, because $\mathrm{var}(\tilde{R}(j)) = n^{-1}\sum_{i=-n+1}^{n-1}(1 - |i|/n)(R(i+j)R(i-j) - \kappa(j, i, i+j))$ (cf. Hannan (1970, p. 209)), we have $\sup_{j \ge 1}\mathrm{var}(\tilde{R}(j)) = O(n^{-1})$ given Assumption A.4. It follows that $\sum_{j=1}^{n-1} k_{nj}^2(\tilde{R}(j) - R(j))^2 = O_P(p_n/n)$ by Markov's inequality. Finally, $\sum_{j=n}^{\infty} R^2(j) = o_P(1)$ by A.4. Hence, condition (ii) holds.

Consistency of M_{1n} then follows. (b) For M_{2n}, it suffices to show $H^2(\hat{f}_n; f_0) \to^p H^2(f; f_0)$. This follows because by Assumption A.5(a), we have

$$\int_{-\pi}^{\pi} (\hat{f}_n^{1/2}(\omega) - f^{1/2}(\omega))^2 \, d\omega$$

$$\leq \delta^{-1} \int_{-\pi}^{\pi} (\hat{f}_n^{1/2}(\omega) - f^{1/2}(\omega))^2 (\hat{f}_n^{1/2}(\omega) + f^{1/2}(\omega))^2 \, d\omega$$

$$= \delta^{-1} \int_{-\pi}^{\pi} (\hat{f}_n(\omega) - f(\omega))^2 \, d\omega = o_P(1).$$

(c) For M_{3n}, we write

$$\text{(A16)} \qquad I(\hat{f}_n; f_0) = -\int_{\Omega(\hat{f}_n)} \ln(f(\omega)/f_0(\omega)) f_0(\omega) \, d\omega - \int_{\Omega(\hat{f}_n)} \ln(\hat{f}_n(\omega)/f(\omega)) f_0(\omega) \, d\omega.$$

For the first term of (A16), we have

$$\int_{\Omega(\hat{f}_n)} \ln(f(\omega)/f_0(\omega)) f_0(\omega) \, d\omega = -I(f; f_0) - \int_{\Omega^c(\hat{f}_n)} \ln(f(\omega)/f_0(\omega)) f_0(\omega) \, d\omega$$

$$= -I(f; f_0) + o_P(1),$$

where the second term is $o_P(1)$ given $\int_{\Omega^c(\hat{f}_n)} d\omega \leq \delta^{-2} \int_{-\pi}^{\pi} (\hat{f}_n(\omega) - f(\omega))^2 \, d\omega = o_P(1)$ and $I(f; f_0) < \infty$, which ensures absolute continuity of $\mu(\Omega) = \int_\Omega \ln(f(\omega)/f_0(\omega)) f_0(\omega) \, d\omega$.

It remains to show that the second term of (A16) is $o_P(1)$. We first show $\|\hat{f}_n - f\|_\infty = o_P(1)$. As shown in the proof of Theorem 4 (cf. (A14)), we have $\|\hat{f}_n - \tilde{f}_n\|_\infty = O_P(p_n/n^{1/2})$. Also, we have $\|\tilde{f}_n - E\tilde{f}_n\|_\infty = O_P(p_n/n^{1/2})$ by Markov's inequality and $E\sum_{j=1}^{n-1} |k_{nj}(\tilde{R}(j) - E\tilde{R}(j))| \leq \sup_{j \geq 1} \{\text{var}(\tilde{R}(j))\}^{1/2} \sum_{j=1}^{n-1} |k_{nj}| = O(p_n/n^{1/2})$ given Assumptions A.1 and A.4. Furthermore, $\|E\tilde{f}_n - f\|_\infty = o(1)$ given Assumptions A.1 and A.5, and $p_n \to \infty$, as shown in Robinson (1991a, Appendix A). It follows that $\|\hat{f}_n - f\|_\infty = o_P(1)$ by the triangle inequality and $p_n^2/n \to 0$. Hence, we can use the inequality $|\ln(1+z)| \leq 2|z|$ for small z near 0 to obtain

$$\left| \int_{\Omega(\hat{f}_n)} \ln(\hat{f}_n(\omega)/f(\omega)) f_0(\omega) \, d\omega \right| \leq 2 \int_{\Omega(\hat{f}_n)} |\hat{f}_n(\omega)/f(\omega) - 1| \, f_0(\omega) \, d\omega$$

$$\leq 2\delta^{-1} \left(\int_{-\pi}^{\pi} (\hat{f}_n(\omega) - f(\omega))^2 f_0(\omega) \, d\omega \right)^{1/2} = o_P(1).$$

It follows from (A16) that $I(\hat{f}_n; f_0) \to^p I(f; f_0)$. The consistency of M_{3n} then follows. *Q.E.D.*

PROOF OF THEOREM 7: Since $M_{jn} \to^d N(0,1)$ under H_0, the asymptotic significance level of M_{jn} is $1 - \Phi(M_{jn})$, where Φ is the cdf of $N(0,1)$. Define $S_{jn} = -2\ln(1 - \Phi(M_{jn}))$. Because $\ln(1 - \Phi(a)) = -\frac{1}{2}a^2(1 + o(1))$ as $a \to +\infty$ (Bahadur (1960, Section 5)), we have

$$\text{(A17)} \qquad (p_n/n^2) S_{jn} = C_j^2(k) = o_P(1) \qquad (j = 1,2,3),$$

by Theorem 7, where $C_i(k) = \frac{1}{2}Q^2(f; f_0)/(2D(k))^{1/2}$, $C_2(k) = 2H^2(f; f_0)/(2D(k))^{1/2}$, and $C_3(k) = I(f; f_0)/(2D(k))^{1/2}$.

We now consider two sequences of test statistics $\{M_{in}\}$ and $\{M_{jn}\}$ under H_A, $i \neq j$, $i, j = 1,2,3$. Bahadur's asymptotic relative efficiency of M_{in} to M_{jn} is defined as the limit ratio of the sample sizes for both tests to attain the same asymptotic significance level under H_A. Let n_i, n_j be such two required sample sizes for $\{M_{in_i}\}$ and $\{M_{jn_j}\}$ respectively. Then $S_{in_i}/S_{jn_j} \to^p 1$ as $n_i, n_j \to \infty$. Using (A17) and $p_{n_i} = cn_i^{\upsilon}$, we obtain the Bahadur's asymptotic relative efficiency of M_{in} to M_{jn} as $\text{ARE}_B(M_{in}; M_{jn}) = \lim_{n_i \to \infty, n_j \to \infty}(n_j/n_i) = [C_i^2(k)/C_j^2(k)]^{1/(2-\upsilon)}$. The desired results follow immediately. *Q.E.D.*

REFERENCES

ANDERSON, T. W. (1971): *Statistical Analysis of Time Series*. New York: John Wiley.

——— (1993): "Goodness of Fit Tests for Spectral Distributions," *Annals of Statistics*, 21, 830–847.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Testing for Serial Correlation Against an ARMA(1, 1) Process," Yale University, Cowles Foundation Discussion Paper No. 1077.

BAHADUR, R. R. (1960): "Stochastic Comparison of Tests," *Annals of Mathematical Statistics*, 31, 276–295.

BELTRAO, K., AND P. BLOOMFIELD (1987): "Determining the Bandwidth of a Kernel Spectrum Estimate," *Journal of Time Series Analysis*, 8, 21–38.

BERK, K. (1974): "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2, 489–502.

BOX, G., AND D. PIERCE (1970): "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of American Statistical Association*, 65, 1509–1526.

BREUSCH, T. S. (1978): "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17, 334–355.

BREUSCH, T. S., AND A. R. PAGAN (1980): "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47, 239–253.

BROWN, B. M. (1971): "Martingale Central Limit Theorems," *Annals of Mathematical Statistics*, 42, 59–66.

DAVIES, R., AND D. HARTE (1987): "Tests for Hurst Effect," *Biometrika*, 74, 95–101.

DURBIN, J., AND G. S. WATSON (1950): "Testing for Serial Correlation in Least Squares Regression: I," *Biometrika*, 37, 409–428.

——— (1951): "Testing for Serial Correlation in Least Squares Regression: II," *Biometrika*, 38, 159–178.

DURLAUF, S. (1991): "Spectral Based Testing for the Martingale Hypothesis," *Journal of Econometrics*, 50, 1–19.

GEWEKE, J. (1981): "The Approximate Slopes of Econometric Tests," *Econometrica*, 49, 1427–1442.

GHOSH, B. K., AND W. HUANG (1991): "The Power and Optimal Kernel of the Bickel-Rosenblatt Test for Goodness of Fit," *Annals of Statistics*, 19, 999–1009.

GODFREY, L. (1978): "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1293–1302.

HANNAN, E. (1970): *Multiple Time Series*. New York: John Wiley.

HODGES, J. L., AND E. L. LEHMANN (1956): "The Efficiency of Some Nonparametric Competitors of the t Tests," *Annals of Mathematical Statistics*, 27, 324–335.

HONG, Y. (1996): "Cross-Correlation Based Testing for Independence Between Two Stationary Time Series," forthcoming in *Biometrika*.

LJUNG, G. M., AND G. E. P. BOX (1978): "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65, 297–303.

NG, S., AND P. PERRON (1994): "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag," Department de Sciences Economiques et C.R.D.E., Université de Montreal.

PITMAN, E. J. G. (1979): *Some Basic Theory for Statistical Inference*. Chapman Hall: London.

PRIESTLEY, M. B. (1962): "Basic Considerations in the Estimation of Spectra," *Technometrics*, 4, 551–564.

——— (1981): *Spectral Analysis and Time Series*. London: Academic Press.

ROBINSON, P. M. (1991a): "Automatic Frequency Domain Inference on Semiparametric and Non-parametric Models," *Econometrica*, 59, 1329–1363.

——— (1991b): "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," *Journal of Econometrics*, 47, 67–84.

——— (1993): "Highly Insignificant *F*-Ratios," *Econometrica*, 61, 687–696.
——— (1994): "Time Series with Strong Dependence," in *Advances in Econometrics*, Sixth World Congress, Vol. 1. ed. by C. Sims. Cambridge: Cambridge University Press, pp. 47–95.
ULLAH, A. (1993): "Entropy, Divergence and Distance Measures with Econometric Applications," Department of Economics, University of California, Riverside.